WILEY-BLACKWELL

Cladistics 37 (2021) 402–422

Cladistics

10.1111/cla.12443

Exploring systematic biases, rooting methods and morphological evidence to unravel the evolutionary history of the genus *Ficus* (Moraceae)

Jean-Yves Rasplus*[a], Lillian Jennifer Rodriguez[b,c], Laure Sauné[a], Yang-Qiong Peng[d], Anthony Bain[e], Finn Kjellberg[f], Rhett D. Harrison[g], Rodrigo A.S. Pereira[h], Rosichon Ubaidillah[i], Christine Tollon-Cordet[j], Mathieu Gautier[a], Jean-Pierre Rossi[a] and Astrid Cruaud[a]

[a]*CBGP, INRAE, CIRAD, IRD, Montpellier SupAgro, Université de Montpellier, Montpellier, 34988, France;* [b]*Institute of Biology, University of the Philippines Diliman, Quezon City, 1101, Philippines;* [c]*Natural Sciences Research Institute, University of the Philippines Diliman, Quezon City, 1101, Philippines;* [d]*CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming, 650223, China;* [e]*Department of Biological Sciences, National Sun Yat-sen University, Kaohsiung, 80424, Taiwan;* [f]*CEFE, CNRS, Université Paul-Valéry Montpellier, EPHE, Université de Montpellier, Montpellier, 34090, France;* [g]*World Agroforestry, Eastern and Southern Africa, Region, 13 Elm Road, Woodlands, Lusaka, 10101, Zambia;* [h]*Departamento de Biologia, FFCLRP, Universidade de São Paulo, Ribeirão Preto, SP, 14040-901, Brazil;* [i]*Museum Zoologicum Bogoriense, LIPI, Gedung Widyasatwaloka, Jln Raya km 46, Cibinong, Bogor, 16911, Indonesia;* [j]*AGAP, INRA, CIRAD, Montpellier SupAgro, Université de Montpellier, Montpellier, 34398, France*

Accepted 11 October 2020

Abstract

Despite many attempts in the Sanger sequencing era, the phylogeny of fig trees remains unresolved, which limits our ability to analyze the evolution of key traits that may have contributed to their evolutionary and ecological success. We used restriction-site-associated DNA sequencing (*c*. 420 kb) and 102 morphological characters to elucidate the relationships between 70 species of *Ficus*. To increase phylogenetic information for higher-level relationships, we targeted conserved regions and assembled paired reads into long loci to enable the retrieval of homologous loci in outgroup genomes. We compared morphological and molecular results to highlight discrepancies and reveal possible inference bias. For the first time, we recovered a monophyletic subgenus *Urostigma* (stranglers) and a clade with all gynodioecious *Ficus*. However, we show, with a new approach based on iterative principal component analysis, that it is not (and will probably never be) possible to homogenize evolutionary rates and GC content for all taxa before phylogenetic inference. Four competing positions for the root of the molecular tree are possible. The placement of section *Pharmacosycea* as sister to other fig trees is not supported by morphological data and considered a result of a long-branch attraction artefact to the outgroups. Regarding morphological features and indirect evidence from the pollinator tree of life, the topology that divides *Ficus* into monoecious *versus* gynodioecious species appears most plausible. It seems most likely that the ancestor of fig trees was a freestanding tree and active pollination is inferred as the ancestral state, contrary to previous hypotheses. However, ambiguity remains on the ancestral breeding system. Despite morphological plasticity, we advocate restoring a central role to morphology in our understanding of the evolution of *Ficus*, as it can help detect systematic errors that appear more pronounced with larger molecular datasets.

Introduction

Ficus (Moraceae) is a pantropical and hyperdiverse genus of plants (*c*. 850 species). As their infructescences (figs) are an important food source for hundreds of frugivorous species (Shanahan et al., 2001), fig trees are key components of tropical ecosystems. They also are known for their intricate relationships with their pollinating wasps (Chalcidoidea: Agaonidae). Indeed, for around 75 Myr, fig trees and agaonids have been obligate mutualists (Cruaud et al.,

*Corresponding author: *E-mail address:* jean-yves.rasplus@inrae.fr

2012). The wasp provides pollination services to the fig tree, the fig tree provides breeding sites for the wasps, and none of the partners is able to reproduce without the other (Galil, 1977; Cook and Rasplus, 2003).

Several studies have attempted to reconstruct the phylogeny of *Ficus* using Sanger sequencing of plastid markers (Herre et al., 1996), external and/or internal transcribed spacers (ETS, ITS; Weiblen, 2000; Jousselin et al., 2003) or nuclear markers (Rønsted et al., 2005; Rønsted et al., 2008; Xu et al., 2011; Cruaud et al., 2012; Pederneiras et al., 2018; Zhang et al., 2018b; Clement et al., 2020). However, the lack of signal in analyzed markers (e.g. approximately 6 kb; 23% of parsimony-informative sites in Cruaud et al., 2012) did not permit resolution of the backbone of the phylogeny. Sequencing of plastid genomes (Bruun-Lund et al., 2017) highlighted a high level of cytonuclear discordance with some subgenera undoubtedly monophyletic (e.g. *Sycidium*) recovered as polyphyletic. Weiblen (2000) proposed the only phylogenetic hypothesis based on morphological data, with a special focus on gynodioecious fig trees. However, his consensus tree was poorly resolved and some groups were recovered as paraphyletic. Consequently, neither molecules nor morphology have enabled the resolution of the *Ficus* tree-of-life yet. There is no consensus on the relationships between major groups of *Ficus* and current classification remains inconsistent with different phylogenetic levels classified under the same taxonomic rank or, contrariwise, identical taxonomic rank appearing at different phylogenetic levels (Table 1).

Therefore, our first objective was to propose a robust phylogenetic hypothesis of fig trees based on nuclear genome-wide markers and morphological characters. Indeed, as they are built independently, molecular and morphological trees should always be compared to highlight discrepancies that may reveal inference bias in either case (Wiens, 2004; Giribet, 2015; Wipfler et al., 2016). We inferred the evolutionary history of 70 species of *Ficus* that represent all subgenera and sections and five outgroups from (i) 102 morphological characters and (ii) Restriction-site-Associated DNA sequencing (RAD-seq). Indeed, RAD-seq has been used successfully to infer ancient evolutionary histories of plants (e.g. Hipp et al., 2020) including a community of 11 strangler figs (sect. *Americanae*; Satler et al., 2019). To increase phylogenetic information for higher-level relationships, we deliberately targeted a low number of highly conserved and shared RAD loci with a rare-cutting restriction enzyme instead of targeting a high number of poorly conserved loci with a frequent-cutter. We assembled paired reads into long loci to enable the retrieval of homologous loci in genomes of outgroups for which missing data were high because of the loss of restriction sites with time. In addition, we performed an analysis of the two main properties known to violate model assumptions (heterogeneity in base composition and evolutionary rates; Brinkmann et al., 2005; Philippe et al., 2017) with existing software and a new approach based on iterative principal component analysis (PCA). We also compared the impact of different rooting strategies on molecular tree topology. Finally, we critically compare morphological and molecular results, and discuss similarities and discrepancies.

The trees were then used to fulfil our second objective, which was to provide an evolutionary analysis of three life-history traits that may have contributed to the evolutionary and ecological success of the genus (life form, breeding system and pollination mode).

Life form

Ficus includes a broad range of habits (trees, hemiepiphytes, lithophytes, shrubs, climbers) with diverse ecologies (Harrison, 2005; Harrison and Shanahan, 2005; Berg and Corner, 2009). Roughly 41% of the fig species are hemi-epiphytes, 10% are climbers and 49% are trees or shrubs (J.Y. Rasplus, unpublished database). Species within each *Ficus* lineage tend to share similar growth forms (Harrison, 2005). However, convergence is thought to be common among fig trees. For example, hemiepiphytism is supposed to have evolved at least four times independently (Jousselin et al., 2003). The only analysis of ancestral character state reconstructions of growth forms in fig trees was equivoqual (Jousselin et al., 2003) and the ancestral growth form of fig trees has yet to be formally identified.

Breeding system

Fifty-two percent of *Ficus* species are monoecious and 48% are gynodioecious (J.Y. Rasplus, unpublished database). In monoecious species, figs contain staminate and pistillate flowers, and produce pollen and seeds. Gynodioecious species are functionally dioecious with male function (pollen) and female function (seed production) segregated on separate individuals. Although dioecy was found to be the ancestral condition of Moraceae (Clement and Weiblen, 2009), the ancestral breeding system of *Ficus* has not yet been reconstructed with confidence (Zhang et al., 2018b). Within *Ficus*, gynodioecy is believed to have evolved at least twice in fig trees with at least three reversals to monoecy (within subgen. *Sycomorus*; Weiblen, 2000; Jousselin et al., 2003).

Pollination mode

Ficus species are pollinated either actively (two-thirds) or passively (one-third; Kjellberg et al., 2001).

Table 1
Classification of the genus *Ficus*

Corner (1958)		Corner (1960a, 1960b, 1960c, 1960d, 1960e, 1965)		Berg and Corner (2005), Rønsted et al. (2008)		Traditional simplified classification adapted from Cruaud et al. (2012)	
Subgenera	Sections	Subgenera	Sections/subsections	Subgenera	Sections/subsections	Subgenera	Sections/subsections
Pharmacosycea Miq.		*Pharmacosycea*	*Pharmacosycea* (Miq.) Benth. & Hook	*Pharmacosycea*	*Pharmacosycea/Bergianae* & *Petenses*	*Pharmacosycea*	*Pharmacosycea*
			Oreosycea (Miq.) Corner		*Oreosycea/Glandulosae* & *Pedunculatae*		*Oreosycea*
				Ficus	*Ficus/Ficus* & *Frutescentiae*	*Ficus*	*Ficus/Ficus* & *Frutescentiae*
Ficus	*Ficus* (Eusyce)	*Ficus*	*Ficus/Ficus* & *Eriosycea*		*Eriosycea* Miq./*Eriosycea* & *Auratae*		*Eriosycea/Eriosycea* & *Auratae*
	Synoecia (Miq.) Benth. & Hook		*Kalosyce* (Miq.) Corner = Synoecia (Miq)	*Synoecia*	*Kissosycea*	*Synoecia*	*Kissosycea*
			Rhizocladus Endl.		*Rhizocladus/Plagiostigma*, *Pogonotrophe*, *Punctulifoliae*, *Trichocarpeae*		*Rhizocladus*
	Sycidium Miq.		*Sycidium* Miq./*Sycidium*, *Varinga*, *Palæomorphe*	*Sycidium*	*Sycidium*	*Sycidium*	*Sycidium*
			Sinosycidium Corner		*Palaeomorphe* King		*Palaeomorphe*
	Adenosperma Corner		*Adenosperma* Corner	*Sycomorus*	*Sycomorus/Sycomorus* & *Neomorphe*	*Sycomorus*	*Sycomorus/Sycomorus* & *Neomorphe*
			Neomorphe King		*Sycocarpus/Sycocarpus* & *Macrostyla*		*Sycocarpus/Sycocarpus* & *Macrostyla*
	Sycocarpus Miq. (Covellia auct.)		*Sycocarpus* Miq./*Auriculisperma*, *Dammaropsis*, *Papuasyce*,		*Adenosperma*		*Adenosperma*
					Hemicardia C.C. Berg		*Hemicardia*
	Sycomorus (Gasp.) Miq.	*Sycomorus*	*Lepidotus*, *Macrostyla*, *Sycocarpus*		*Papuasyce* (Corner) C.C. Berg		*Papuasyce*
					Bosscheria (Teijsm. & de Vriese) C.C. Berg		*Bosscheria*
					Dammaropsis (Warb.) C.C. Berg		*Dammaropsis*
Urostigma (Gasp.) Miq.	*Americana* Miq.	*Urostigma*	*Americana* (Miq.) Corner	*Urostigma*	*Americana*	*Urostigma*	*Americanae**
	Bibracteata Milxdbr. & Burr.		*Conosycea* (Miq.) Corner/*Conosycea*, *Dictyoneuron*, *Benjamina*		*Galoglychia*		*Conosycea*†
	Urostigma		*Galoglychia* (Gasp.) Endl.		*Stilnophyllum/Stilnophyllum*, *Malvanthera*		*Galoglychia*
			Leucogyne Corner		*Urostigma/Conosycea* & *Urostigma*		*Malvanthera**
			Malvanthera Corner				*Urostigma*†
			Stilpnophyllum Endl.				
			Urostigma				

In this paper, we use the simplified classification adapted from Cruaud et al. (2012). NB Subgenus *Urostigma* excluding section *Urostigma* is referred to as *Mixtiflores* by Clement et al. (2020).

*The proper spelling of this section should be *Americanae*, which is the original spelling proposed by Miquel. Indeed, it is a plural adjective.

†*Conosycea*, *Malvanthera* and *Urostigma* are considered as sections instead of subsections and *F. elastica* the only member of the subsect. *Stilnophyllum* sensu Berg and Corner (2009) is included in *Conosycea* as recovered in previous molecular studies.

In passively pollinated figs, emerging wasps are dusted with pollen before leaving their natal fig. In actively pollinated figs, wasps use their legs and pollen pockets to collect and store pollen they will later deposit on flowers of receptive figs, while laying their eggs. Passively pollinated figs have high ratios of anthers to female flowers (Kjellberg et al., 2001) and passive pollination appears to be costly for the fig (Herre et al., 2008). Although relationships among clades of *Ficus* are not resolved, several studies have proposed passive pollination as the ancestral mode for *Ficus*, followed by one shift to active pollination and several reversions to passive pollination (Herre et al., 2008; Machado et al., 2001; Jousselin et al., 2003; Jandér and Herre, 2010). Cruaud et al. (2012) did not validate this hypothesis and showed that the ancestral pollination type was equivocal.

We used the recovered phylogenetic trees to infer the ancestral states for these three life-history traits and discuss their evolutionary lability by addressing two questions: Do we observe multiple instances of reversal to ancestral states? and What is the relative role of phylogenetic conservatism and convergence in explaining similarity in species traits?

Materials and methods

Sampling and classification

Here we use the classification by Berg and Corner (2005) with some modifications used in Cruaud et al. (2012; Table 1). Seventy species of *Ficus* representing all known subgenera and sections as well as four outgroups were included in the analysis (Table S1). The same individual was used for molecular and morphological studies. Plants, twigs, leaves and figs were photographed before sampling of a few leaves that were dried for molecular purposes. Voucher specimens are archived at Centre de Biologie pour la Gestion des Populations (CBGP), Montpellier.

Morphological data

Species were scored for 102 morphological characters (Appendix S1). Seventy-nine were extracted from earlier phylogenetic studies (Weiblen, 2000; Clement and Weiblen, 2009; Chantarasuwan et al., 2015) and sometimes modified (see Appendix S1), whereas 23 characters were used for the first time. Whenever possible we cross-validated our observations and accounted for polymorphism using descriptions available in the literature (Corner, 1938; Corner, 1967; Corner, 1969a; Corner, 1969b; Corner, 1970; Corner, 1978a; Corner, 1978b; Berg and Wiebes, 1992; Berg and Corner, 2005; Berg, 2009; Berg et al., 2011) and two to four conspecific specimens from other localities, when distribution ranges were large. Data were analyzed using maximum parsimony (MP) as implemented in PAUP* v.4.0a (Swofford, 2003). We used an heuristic search with 5000 random addition sequences (RAS) to obtain an initial tree and "tree bisection and reconnection (TBR)" as branch swapping option, with reconnection limit set to 100. One tree was retained at each step. Characters were equally weighted and treated as unordered and nonadditive. Multiple states were interpreted as polymorphism and gaps (characters that were impossible to score because the feature was nonexistent) were treated as missing data. Robustness of the topology was assessed by bootstrap procedures (100 replicates; TBR RAS 100; one tree retained at each step). Character transformations were mapped on the majority-rule consensus tree and the four alternative RAD topologies in PAUP* using the accelerated transformation (ACCTRAN) algorithm.

DNA extraction and library construction

Leaves were either dried with silica gel or sun-dried (without significant effect on the number of RAD loci analyzed). Twenty mg of dried leaves were placed in Eppendorf vials and crushed with ceramic beads in liquid nitrogen. DNA was extracted with the Chemagic DNA Plant Kit (Perkin Elmer Chemagen, Baesweller, Germany, part no. CMG-194), according to the manufacturer's instructions with a modification of the cell lysis. The protocol was adapted to the use of the KingFisher Flex™ (Thermo Fisher Scientific, Waltham, MA, USA) automated DNA purification workstation. The powder was suspended in 400 µL Lysis buffer (200 mM Tris pH = 8.0, 50 mM EDTA, 500 mM NaCl, 1.25% SDS, 0.5% CTAB 1% PVP 40000, 1 g/100 mL Sodium Bisulfite) and incubated for 20 min at 65 °C. Then, 150 µL cold precipitation buffer (sodium acetate 3 M, pH 5.2) was added. Samples were centrifuged for 10 min at 7 500 g and 350 µL of the supernatant were transferred into a 96-deepwell plate. Binding of DNA on magnetic beads, wash buffer use and elution of purified DNA followed Chemagic kit protocol and KingFisher Flex use recommendations.

Library construction followed Baird et al. (2008) and Etter et al. (2011) with modifications detailed in Cruaud et al. (2014) and below. To infer deep phylogenetic relationships, we targeted conserved regions with an infrequent 8-cutter restriction enzyme (*SbfI*). The expected number of cut sites was estimated with the radcounter_v4.xls spread sheet available from the UK RAD Sequencing Wiki (www.wiki.ed.ac.uk/display/RADSequencing/Home). We assumed a 352-Mb haploid genome size (Ohri and Khoshoo, 1987; *c*. 1.44 pg on average for 15 species of *Ficus*) and a 48% GC content (estimated from EST data available on NCBI when we started the project in 2016). Based on those estimates, 9095 cut sites were expected. For each sample 125 ng of input DNA was used. After digestion, 1 µL P1 adapters (100 nM) was added to saturate restriction sites. Samples were then pooled 16-by-16 and DNA of each pool was sheared to a mean size of *c*. 400 bp using the Bioruptor® Pico (Diagenode; 15 s ON/90 s OFF for eight cycles). After shearing, end repair and 3′-end adenylation, DNA of each pool was tagged with a different barcoded P2 adapter. A PCR enrichment step was performed before KAPA quantification. The 2*125 nt paired-end sequencing of the library was performed at MGX-Montpellier GenomiX on one lane of an Illumina HiSeq 2500 flow cell.

Data cleaning and assembly of paired reads into RAD loci

Data cleaning was performed with RADIS (Cruaud et al., 2016), which relies on Stacks (Catchen et al., 2013) for demultiplexing and removal of PCR duplicates. Individual loci were built using *ustacks* [m = 15; M = 2, N = 4; with removal (r) and deleveraging (d) algorithms enabled]. The parameter *n* of *cstacks* (number of mismatches allowed between sample loci when building the catalogue) was set to 20 to cluster enough loci for the outgroups, while ensuring not to cluster paralogues in the ingroup. To target loci with slow or moderate substitution rate, only loci present in 75% of the samples were analyzed. Loci for which samples had three or more sequences were removed from the analysis. Loci were aligned with MAFFT v.7.245 (-linsi option; Katoh and Standley, 2013). The 583 loci obtained in

this step (mergeR1 dataset) were used as a starting point to assemble paired reads into longer RAD loci. The pipeline, scripts and parameters used for the assembly of paired reads are available from https://github.com/acruaud/radseq_ficus_2020. Briefly, for each sample, forward reads used to build the 583 cstacks loci of the mergeR1 dataset as well as corresponding reverse reads were retrieved from original fastq files with custom scripts and assembled with Trinity (Haas et al., 2013). Contigs were aligned to the reference genome of *F. carica* assembly GCA_002002945.1 with LASTZ Release 1.02.00 (Harris, 2007). Homology between cstacks loci and reference genome, and homology between sample contigs within cstacks loci were tested as follows. For each cstacks locus, the genome scaffold with the highest number of alignment hits was considered as likely to contain the RAD locus. When contigs aligned with different parts of the same scaffold, the genome region that showed the highest identity with the sample contigs (as estimated with GENEIOUS 11.1.4; https://www.geneious.com) was considered as the most likely RAD locus. Cstacks loci for which the sample of *F. carica* JRAS06927_0001 included in the RAD library was not properly aligned with the reference genome (hard or soft clipped unaligned ends > 10 bp) or for which the majority of contigs did not align with the same genome region as *F. carica* JRAS06927_0001 were removed. Finally, loci for which at least one sample contig appeared more than once were discarded. When several contigs were retained per sample (e.g. when forward and reverse reads did not overlap; or in case of polyploidy or sequencing mistake) a consensus was built and the IUPAC code was used without considering any threshold, if, for a given position, different nucleotides were present. Of the 583 initial loci, 530 successfully passed quality controls and were retained for phylogenetic analysis.

Retrieval of RAD loci in genome of outgroup species

As expected for a RAD experiment (Rubin et al., 2012; Gautier et al., 2013), outgroups included in the library had a high level of missing data (*c.* 70%), certainly because restriction sites were lost due to mutations. To decrease missing data, we retrieved RAD loci in available outgroup genomes. Aside from *F. carica*, only two genomes of Moraceae were available on NCBI when we performed this study: *Morus notabilis* (assembly GCA_000414095.2) and *Artocarpus camansi* (assembly GCA_002024485.1). Pipeline, scripts and parameters used for the retrieval of RAD loci in genome of outgroups are available from https://github.com/acruaud/radseq_ficus_2020. Briefly, the 530 RAD loci extracted from the genome of *F. carica* in the previous step were aligned with the two outgroup genomes using Lastz. Alignment results were parsed with SAMTOOLS (Li et al., 2009) to select among the genome regions on which a single RAD locus matched. We considered that the genome region with the highest similarity was the most likely to be homologous with the query RAD locus. Putative RAD loci were extracted from the genome with custom scripts and aligned with the contigs of forward and reverse reads produced in the previous step using MAFFT v.7.245 (-linsi option). The final dataset (mergeR1R2) was composed of 530 loci, 71 ingroup species (70 included in the RAD library plus the genome of *F. carica*) and five outgroup species (*Antiaris toxicaria*, *Artocarpus* sp. and *Morus alba* that were included in the RAD library plus the genomes of *A. camansi* and *M. notabilis*). Summary statistics for datasets and samples were calculated using AMAS (Borowiec, 2016). Tests for phylogenetic signal of gaps and missing data were conducted in R (R Core Team, 2018) using the *K* statistic (Blomberg et al., 2003) implemented in the package PHYTOOLS (Revell, 2012) to test whether these properties could generate artificial clustering. The null expectation of *K* under no phylogenetic signal was generated by randomly shuffling the tips of the phylogeny 1000 times.

Dataset cleaning

In order to reduce inference bias due to possible misalignment or default of homology, TREESHRINK (Mai and Mirarab, 2018) was used to detect and remove abnormally long branches in individual gene trees of the mergeR1R2 dataset. As suggested in the manual, the value of b (the percentage increase in the maximum distance between any two leaves above which long terminals should be removed) was determined by a preliminary analysis on a subset of loci and set to 20. Four rounds (two with and two without the outgroups) were performed, to ensure a proper cleaning. Following Tan et al. (2015) we only performed a light filtering of alignment positions that contained gaps to reduce signal loss. Sites with more than 75% gaps were removed from the locus alignments using the program SEQTOOLS implemented in the package PASTA (Mirarab et al., 2014).

Exploration of potential bias

In order to explore potential sources of bias, a correlation analysis between locus properties was performed with R/PERFORMANCE ANALYTICS (Peterson and Carl, 2018). We explored more thoroughly the two main properties known to violate model assumptions: heterogeneity in evolutionary rates and base composition (Brinkmann et al., 2005; Philippe et al., 2017). We used three different methods to test for a possible impact of heterogeneity of evolutionary rates between taxa: (i) the LS^3 approach (Rivera-Rivera and Montoya-Burgos, 2016; Rivera-Rivera and Montoya-Burgos, 2019) (ii) a custom approach based on iterative principal component analysis (PCA) of long branch (LB) heterogeneity scores of taxa (Struck, 2014) in individual gene trees (see below for details on the PCA-based approach) and (iii) different rooting approaches: midpoint rooting, minimal ancestor deviation (MAD) (Tria et al., 2017) and minimum variance rooting (MinVar; Mai et al., 2017). We evaluated a possible impact of base composition heterogeneity among taxa and markers using (i) incremental removal of the most GC-biased loci and (ii) iterative PCA of GC content of taxa in individual gene trees. Indeed, if the study of heterogeneity in base composition among taxa should be performed in all datasets, studying heterogeneity in GC content is especially relevant here as the enzyme used is known to cut in GC-rich regions.

For the LS^3 approach we defined four clades of interest which corresponded to the four highly supported clades recovered in the phylogenetic trees inferred from the mergeR1R2 dataset: Clade1 = sect. *Pharmacosycea*; Clade2 = subg. *Urostigma*, Clade3 = sect. *Oreosycea*, Clade4 = "gynodioecious clade". The LS^3 algorithm then was used to find a subsample of sequences in each locus that evolve at a homogeneous rate across all clades of interests. The minTaxa parameter was set to 1.

The custom iterative PCA approach was developed in R to analyze LB heterogeneity scores and GC content of taxa in RAD loci. The PCA consisted of the eigenanalysis of the matrix of the correlations between loci and yielded a set of principal axes corresponding to linear combinations of these variables (Manly and Alberto, 2017). We used the scores of the taxa along the axes to detect a possible nonrandom distribution of taxa on the reduced space of the PCA. Depending on the studied properties different groups were highlighted and compared (LB scores: sect. *Pharmacosycea* vs. all other fig trees; GC content: *Mixtiflores* vs. all other fig trees and then sect. *Pharmacosycea* vs. other fig trees, *Mixtiflores* excluded). An initial PCA was performed on all loci and the differences among the groups were statistically assessed by means of a Wilcoxon test applied to the score of the loci upon the first PCA axis. Then, the locus showing the highest correlation with the axis was removed and another PCA and Wilcoxon test were performed on the thinned dataset. The locus showing the highest correlation with the first axis

of this new PCA was removed and so on. Only loci for which no structure could be observed were retained for phylogenetic analysis (i.e. loci for which Wilcoxon tests were nonsignificant, indicating no difference between the groups along the first axis of the PCA). Although they are not similar, the PCA approach we developed may be related to multidimensional scaling that has been recently used to select loci that share congruent evolutionary signal (Gori et al., 2016; Duchêne et al., 2018). The PCA approach was implemented to homogenize LB heterogeneity scores and GC content as much as possible before phylogenetic inference. PCA were performed with R/ADE4 (Dray and Dufour, 2007). The R script and a tutorial to perform iterative PCAs are available from https://github.com/acruaud/radseq_ficus_2020.

Phylogenetic inference

Gene trees were inferred with a ML approach as implemented in RAxMLHPC-PTHREADS-AVX v.8.2.4 (Stamatakis, 2014). A rapid bootstrap search (100 replicates) followed by a thorough ML search (-m GTRGAMMA) was performed. Phylogenetic analyses of the concatenated dataset were performed with supermatrix (ML and parsimony) and coalescent-based summary methods. For the ML approach, we used RAxMLHPC-PTHREADS-AVX v.8.2.4 and IQ-TREE v.1.6.7 (Nguyen et al., 2015). Datasets were analyzed without partitioning. A rapid bootstrap search (100 replicates) followed by a thorough ML search (-m GTRGAMMA) was implemented for the RAxML approach. IQ-TREE analysis employed an ML search with the best-fit substitution model automatically selected and branch supports were assessed with ultrafast bootstrap (Minh et al., 2013) and SH-aLRT test (Guindon et al., 2010; 1000 replicates). Parsimony analyses were performed with MPBOOT v.1.1.0 (Hoang et al., 2018; default parameters). Gaps were treated as missing data. Node supports were assessed with a fast approximation of Bootstrap as implemented in MPBOOT (1000 replicates). Finally, ASTRAL-III v.5.6.1 (Zhang et al., 2018a) was used to infer species trees from individual gene trees. To improve accuracy, nodes with BP support < 10 were collapsed in individual gene trees with the perl script AfterPhylo.pl (Zhu, 2014) before species tree inference with ASTRAL. As heterogeneity in locus coverage could mislead phylogenetic analyses (Hosner et al., 2016), analyses were also conducted by removing sequences with less than 50% locus coverage. Trees were annotated with TREEGRAPH 2.13 (Stöver and Müller, 2010) and R/APE (Paradis et al., 2004).

Evolution of life-history traits

Three key traits (life form, breeding system and pollination mode) were studied. Stochastic mapping (Huelsenbeck et al., 2003) as described in Bollback (2006) and implemented in R/ PHYTOOLS was utilized to estimate the ancestral state and the number of transitions for each trait. The transition matrix was first sampled from its posterior probability distribution conditioned on the substitution model (10 000 generations of MCMC, sampling every 100 generations). Then, 100 stochastic character histories were simulated conditioned on each sampled value of the transition matrix. Three Markov models were tested: equal rates model (ER) with a single parameter for all transition rates, symmetric model (SYM) in which forward and reverse transition have the same rate and all rates different model (ARD). AIC scores and Akaike weight for each model were computed.

Computational resources

Analyses were performed on a Dell PowerEdge T630 server with two 10-core Intel(R) Xeon(R) CPUs E5-2687W v3 @ 3.10 GHz and on the Genotoul Cluster (INRA, Toulouse, France, http://bioinfo.genotoul.fr/).

Results

Molecular phylogenetic inference with outgroup rooting

As a reminder, classifications of the genus *Ficus* are provided in Table 1. Datasets are described in Table 2 and features of taxa are reported in Tables S1–S3. An average of 2*416 834 paired reads; 2520 ustacks loci and 2308 cstacks loci were obtained for each sample included in the RAD library (Table S1). In an attempt to solve the tree backbone, we kept only the 583 most conserved loci (shared by 75% of the samples) assembled from forward reads (mergeR1 dataset), 91% of which (530) were retained in the final (mergeR1R2) dataset. Mining of loci in outgroup genomes largely reduced the level of missing data (from *c*. 70–75% in the mergeR1 dataset to *c*. 15–30% in the mergeR1R2 dataset, depending on the outgroup; Table S1). TREESHRINK showed that samples had outlier long branches for 1–55 loci (average 15; Table S4). Outlier sequences were removed and the final dataset used for phylogenetic inference (mergeR1R2) was composed of 70 species of *Ficus* and five outgroups. We did not obtain enough reads for *Sparattosyce dioica* to include it in our analysis. Alignment length was 419 945 bp (Table 2).

RAxML and IQ-TREE produced identical topologies with high statistical support (Fig. 1, S1A-B; Table 3). Neither gaps ($K = 0.418$, $P = 0.342$) nor missing data ($K = 0.384$, $P = 0.555$) were phylogenetically clustered in the ingroup (i.e. taxa with high percentages of missing data/gaps did not cluster together more often than expected by chance). All subgenera except *Ficus* and *Pharmacosycea* were recovered as monophyletic with strong support and all non-monospecific sections of the dataset except *Ficus* were monophyletic with strong support. Section *Pharmacosycea* was sister to all other *Ficus* species with strong support. The remaining species clustered into two highly supported groups: (i) subg. *Urostigma* and sect. *Oreosycea*; (ii) subg. *Ficus*, *Sycomorus*, *Sycidium* and *Synoecia*, hereafter named the "gynodioecious clade" for brevity as it clusters all gynodioecious species of fig trees (although a few monoecious species are present in subg. *Sycomorus*). Relationships within the "gynodioecious clade" were strongly supported with subg. *Sycidium* + *F. carica* sister to subg. *Sycomorus* + other species of the subg. *Ficus* and *Synoecia*. Subsection *Frutescentiae* was sister to a clade grouping sect. *Eriosycea* and subg. *Synoecia*, yet with poor support. Parsimony inferred the same topology as ML with the exception of two unsupported

Table 2
Description of the concatenated datasets analyzed in the study

Dataset name	Description
a-description of the datasets	
mergeR1	Original dataset, clustering of forward reads (R1) into RAD loci with RADIS and stacks, no filtering of loci (m ustacks = 15, M ustacks = 2, N ustacks = 4; n cstacks = 20; matrix completeness = 75%, Npbloci radis = 3 (i.e. RAD loci were removed from the analysis if at least one sample had more than three sequences for this locus in the cstacks catalogue), alignment of individual loci = mafft-linsi, no alignment cleaning)
mergeR1R2	Final dataset obtained with the assembly of paired reads into RAD loci using the mergeR1 dataset as a starting point, filtering of sequences with treeshrink (four rounds), alignment of individual loci with mafft-linsi, alignment cleaning = seqtools 25 (i.e. for each locus, alignment positions with more than 75% of gaps were removed)
mergeR1R2_50%locuscoverage	Starting from the mergeR1R2 dataset, sequences with more than 50% gaps in individual loci were removed
mergeR1R2_PCA	Starting from the mergeR1R2 dataset, loci filtered out whereas Wilcoxon test significant between scattered plots of LB scores for sect. *Pharmacosycea* and other ingroups
mergeR1R2_LS3	Starting from the mergeR1R2 dataset, filtering based on LS^3 (loci with less than one sample per group ($N = 93$) filtered out before analysis as required by the program)
mergeR1R2_GCinfmean	Starting from the mergeR1R2 dataset, loci filtered out when GC content strictly superior to mean GC content (0.406)
mergeR1R2_GCsupmean	Starting from the mergeR1R2 dataset, loci filtered out when GC content inferior of equal to mean GC content (0.406)

Dataset name	mergeR1	mergeR1R2	mergeR1 R2_50%locus coverage	mergeR1 R2_PCA	mergeR1 R2_LS3	mergeR1 R2_GC infmean	merge R1 R2_GC supmean
b-properties of the datasets							
Nb taxa	73	76	76	76	76	76	76
Nb loci	583	530	530	134	377 (rate homogeneity could not be reached for 60 loci)	306	224
Alignment length (bp)	66 197	419 945	419 945	107 499	299 524	239 709	180 236
Variable sites content	0.314	0.453	0.441	0.438	0.432	0.482	0.415
Parsimony-informative sites content	0.156	0.203	0.195	0.192	0.186	0.215	0.188
GC content	0.444	0.405	0.404	0.408	0.408	0.373	0.449
Gap content*	0.004	0.187	0.099	0.183	0.168	0.185	0.189
Missing data content†	0.157	0.193	0.311	0.195	0.267	0.198	0.185
Supplementary figure	NA	S1	S2	S7	S8	S6 (B–H)	S6 (I–O)

*Gap content refers to indels inserted during alignment or missing parts of RAD loci following assembly of forward and reverse reads.
†Missing data refers to missing position in the matrix due to the absence of RAD loci.

changes within the "gynodioecious clade" (Fig. S1C, Table 3). Subgenus *Synoecia* was sister to subsect. *Frutescentiae* (BP = 48), and *F. pedunculosa* did not cluster with subsect. *Frutescentiae* but instead with subg. *Synoecia* (BP = 63). ASTRAL recovered sect. *Oreosycea* sister to the "gynodioecious clade" when the whole mergeR1R2 dataset was considered (Fig. S1D), albeit with low support (PP = 0.2). However, when sequences with <50% locus coverage were removed from each RAD locus, ASTRAL inferred sect. *Oreosycea* as sister to subg. *Urostigma* (Fig. S2, PP = 0.8). Only two unsupported changes were observed between the ASTRAL and ML trees in the shallowest nodes (within sect. *Conosycea* and *Malvanthera*, Fig. S1D).

Identification of potential bias: heterogeneity among RAD loci

Spearman's rank correlation tests showed a significant negative correlation between the proportion of parsimony-informative sites or the average bootstrap support of gene trees and (i) GC content of loci and (ii) LB score heterogeneity of loci (Fig. S3). Furthermore, loci with more homogeneous rates among sites (high alpha) were more informative. Usually, when

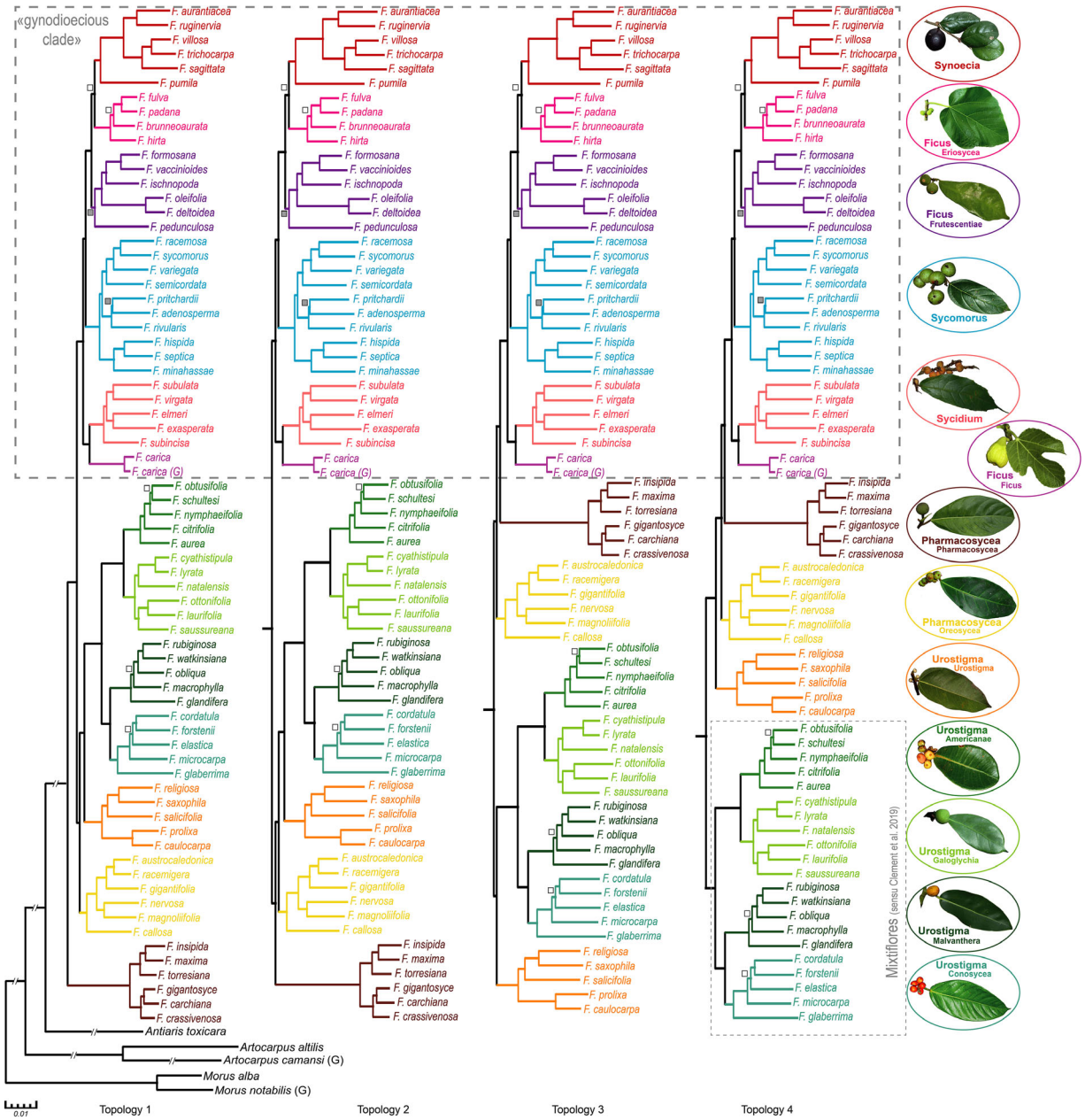


Fig. 1. Maximum-likelihood (ML) trees obtained with the molecular dataset. Genomes are indicated with a (G). Subgenera (larger font size) and (sub)sections (smaller font size) are illustrated with pictograms and highlighted with different colours. Classification follows Cruaud et al. (2012) as reported in Table 1. Only ML trees are represented to illustrate comparison with alternative rooting strategies to outgroup rooting, but parsimony and ASTRAL trees are provided in the Supplementary data (and see text and Table 3). All nodes were supported with RAxML Bootstrap (BP) ≥ 90 and IQ-TREE SH-aLRT ≥ 80/UFBoot ≥ 95 unless specified with a grey square: BP < 90 or SH-aLRT < 80/UFBoot < 95 or a white square: BP < 90 and SH-aLRT < 80/UFBoot < 95. Although a few *Sycomorus* are monoecious, we refer to the clade that groups subgenera *Ficus*, *Sycidium*, *Sycomorus* and *Synoecia* as the "gynodioecious clade" for brevity. *Topology 1*: tree obtained from the mergeR1R2 dataset with outgroup rooting (see Table 2 for a description of the molecular datasets analyzed in this study). This topology is not supported by morphological data (Figs 3 and S9) and supposed to result from an LBA artefact of sect. *Pharmacosycea* to the outgroups. *Topology 2* is obtained for the mergeR1R2_PCA dataset (i.e. when heterogeneity in evolutionary rates is reduced) when midpoint rooting is used (Fig. 2). It is the most-supported by morphological data and evolutionary history of pollinators. *Topology 3* is obtained for the mergeR1R2_PCA dataset when minimum variance rooting (MinVar) rooting is used. It is less supported by morphological data and evolutionary history of pollinators than Topology 2. *Topology 4* is obtained for the mergeR1R2_LS3 dataset and the mergeR1R2_GCinfmean dataset when MinVar rooting is used. It is considered unlikely as the position of sect. *Urostigma* is not supported by morphological data. The position of the root could be driven by the GC-content bias exhibited by *Mixtiflores*. [Colour figure can be viewed at wileyonlinelibrary.com]

Table 3
Summary of the phylogenetic relationships inferred from the different datasets when topologies were rooted with outgroups

	mergeR1R2				mergeR1R2_PCA				mergeR1R2_LS3				mergeR1R2_GCinfmean				mergeR1R2_GCsupmean			
Relationships	RAxML	IQ-TREE	MP Boot	Astral	RAxML	IQ-TREE	MP Boot	Astral	RAxML	IQ-TREE	MP Boot	Astral	RAxML	IQ-TREE	MP Boot	Astral	RAxML	IQ-TREE	MP Boot	Astral
Subg. *Pharmacosycea* monophyletic	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Sect. *Pharmacosycea* sister to other *Ficus*	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Subg. *Urostigma* monophyletic (sect. *Urostigma* (*Americanae* + *Galoglychia*) + (*Conosycea* + *Malvanthera*))	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Gynodioecious clade* monophyletic	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sect. *Oreosycea* + Subg. *Urostigma*	Y	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y
Subg. *Ficus* monophyletic	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
F. carica + sg. *Sycidium*	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Subg. *Sycidium* sister to other gynodioecious *Ficus*	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	Y
Subg. *Sycomorus* sister to other gynodioecious *Ficus*	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	Y	Y	N
Ssect. *Frutescenciae* incl. *F. pedunculosa*	Y	Y	N	Y	N	N	N	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	Y
Subg. *Synoecia* + *F. pedunculosa*	N	N	Y	N	Y	Y	Y	Y	N	N	Y	N	N	N	N	N	N	N	Y	N
Sect. *Eriosycea* + ssect. *Frutescentiae*	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Sect. *Eriosycea* + subg. *Synoecia*	Y	Y	N	Y	N	N	N	Y	Y	Y	N	N	Y	Y	Y	Y	N	N	N	Y
Ssect. *Frutescentiae* + subg. *Synoecia*	N	N	Y	N	Y	Y	Y	N	N	N	Y	Y	N	N	N	N	Y	Y	Y	N

Datasets are described in Table 2 and topologies are available as supplementary data (Figs S1, S1–S8).

*Although a few *Sycomorus* are monoecious, we refer to the clade that groups subgenera *Ficus*, *Sycidium*, *Sycomorus* and *Synoecia* as the "gynodioecious clade" for brevity. N =- Not recovered; Y = recovered. Colour coding is as follows: white = not recovered; light grey = RAxML Bootstrap < 90, IQ-TREE SH-aLRT < 80/UFBoot < 95; MPBoot BP < 90 and Astral posterior probability < 0.9; dark grey = RAxML Bootstrap ≥ 90, IQ-TREE SH-aLRT ≥ 80/UFBoot ≥ 95; MPBoot BP ≥ 90 and Astral posterior probability ≥ 0.9. Note for the Astral analysis: for each locus, sequences with more than 50% missing data were removed before gene tree inference as it seemed to distort species tree inference and nodes with BP < 10 were collapsed in each gene tree before species tree inference (see text).

they want to test for the impact of locus properties on phylogenetic inference, authors built data subsets by (incrementally) removing the most biased or the most heterogeneous loci (e.g. Romiguier et al., 2016; Bossert et al., 2017; Cruaud et al.,). Nevertheless, it may be difficult to estimate whether topological changes result from the actual removal of bias or only from the decrease in phylogenetic signal induced by locus removal. From what we observed (Table 3), loci with high GC content might have misled parsimony, which is the only method to infer a sister taxa relationship between subgenus *Synoecia* + *F. pedunculosa* + subsect. *Frutescentiae* from the complete dataset. Indeed, parsimony recovers the same relationships as the ML and ASTRAL approaches (i.e. subgenus *Synoecia* + sect. *Eriosycea* sister to *F. pedunculosa* + subsect. *Frutescentiae*) when only loci with low GC content are analyzed. In addition, ASTRAL is the only method that recovers subgenus *Synoecia* + sect. *Eriosycea* sister to *F. pedunculosa* + subsect. *Frutescentiae* when loci with the highest GC content are analyzed. Incomplete lineage sorting is expected to be more important in GC-rich than GC-poor regions (Romiguier et al., 2016). ASTRAL was developed to better handle incomplete lineage sorting (Zhang et al., 2018a), which may explain this latter result. From these results, it seems that heterogeneity of GC content among loci in the complete dataset did not bias ML and ASTRAL inferences.

Identification of potential bias: heterogeneity among taxa

Usually when heterogeneity in base composition or evolutionary rates among taxa is highlighted, biased taxa are removed to detect possible topological changes. This is typically the case with long-branched taxa (Bergsten, 2005). However, with this approach, biased taxa cannot be included in phylogenies. Here, we tried to develop a new approach to detect whether we could keep at least some loci for biased taxa to include them in the phylogeny.

As differences in coverage of RAD loci across samples prevented a proper calculation of GC content in the mergeR1R2 dataset (i.e. loci were only partially sequenced in some samples), we focused on the mergeR1 dataset to explore GC-content bias among samples with PCA (Fig. S4). The first principal component (PC1) discriminated between (i) all sections of the subg. *Urostigma* except sect. *Urostigma* (i.e. the *Mixtiflores* group sensu Clement et al. (2020)) and (ii) all other species of *Ficus* (eigenvalues = 9.50% for PC1 and 8.57% for PC2; Fig. S4Ab). Within the remaining species of *Ficus*, PC1 discriminated between (i) sect. *Pharmacosycea* and (ii) other species of *Ficus* (eigenvalue = 12.55% for PC1 and 6.64% for PC2; Fig. S4Ac). Wilcoxon tests showed that the distance separating (i) *Mixtiflores* and other fig trees on one side and (ii) sect. *Pharmacosycea* and other fig trees (*Mixtiflores* excluded) on the other side was significant for all iterations of the PCA (Fig. S4B,C). This means that it was not possible to homogenize GC content of taxa before phylogenetic inference to fit with model assumptions. The GC content of *Mixtiflores* was significantly higher and that of sect. *Pharmacosycea* significantly lower than the GC content of other fig trees (Fig. S4D,E).

In addition to heterogeneity of GC content among *Ficus* lineages, we highlighted heterogeneity in evolutionary rates. Two groups were highlighted on the PCA of LB scores across all RAD loci: sect. *Pharmacosycea* and all other fig trees (Fig. S5, eigenvalues = 18.76% for PC1 and 4.34% for PC2). Moreover, on average, 29.0% of the loci were flagged for sect. *Pharmacosycea* by LS^3, whereas only 19.4% were flagged for other fig trees (Table S5). Attempts to reduce heterogeneity in evolutionary rates with LS^3 and the custom PCA approach failed. Whatever the concatenated dataset analyzed (supposedly cleaned or not of bias), the branch leading to sect. *Pharmacosycea* was the longest (Fig. 2, S1F, S6F, S6M, S7F, S8E) and sect. *Pharmacosycea* always had significantly higher LB heterogeneity scores than all other taxa (about 7.5 points more for the mergeR1R2_PCA and _LS3 datasets and about 10 points more for the mergeR1R2, _GCinfmean, _GCsupmean datasets; Table 4).

Impact of rooting strategies on the molecular tree

Fast-evolving or compositionally biased ingroup taxa can be drawn towards the outgroups, especially when the outgroup is distantly related to the ingroup (Long Branch Attraction (LBA) artefact (Bergsten, 2005)), which is the case here (Fig. S1). For that reason, we tested alternative rooting methods to outgroup rooting (only on the ML trees as branch lengths are required). Although outgroup rooting always recovered the long-branched sect. *Pharmacosycea* as sister to the remaining fig trees (Topology 1, Fig. 2), other rooting methods suggested three alternative positions for the root: (i) on the branch separating the "gynodioecious clade" from other *Ficus* species (Topology 2); (ii) on the branch separating subgenus *Urostigma* from other fig trees (Topology 3); or (iii) on the branch separating sect. *Conosycea*, *Malvanthera*, *Americanae* and *Galoglychia* (*Mixtiflores*) from the remaining fig trees (Topology 4). The root ambiguity index calculated by MAD was high (0.788–0.999; average = 0.938) which indicates that root inference was problematic for all datasets.

Fig. 2. Impact of rooting methods on tree estimation. Three alternative methods to outgroup rooting were tested: midpoint rooting, minimal ancestor deviation (MAD) (Tria et al., 2017) and minimum variance rooting (MinVar; Mai et al., 2017). Outgroups were removed from the datasets described in Table 2 before analysis with IQ-TREE (best-fit substitution model automatically selected). Full trees are available as supplementary data. Top figure: Unrooted tree obtained from the mergeR1R2 dataset with branch colours corresponding to their ancestor relative deviation value AD; alternative positions for the root are indicated with arrows. Middle figure: Illustration of the four topologies obtained with different rooting strategies. Bottom figure: Summary table of the topologies obtained from different datasets and rooting strategies. In all cases, the MAD approach resulted in very high ambiguity scores for the root (0.788–0.999; average = 0.938). When alternative positions for the root were identified by MAD, the preferred topology is listed first in the table; then, topologies within 0.01 units of ancestral deviation scores are listed by decreasing order of ancestral deviation. Statistical support of nodes varied with analyzed datasets. Black squares indicate strong support: IQ-TREE SH-aLRT ≥ 80/UFBoot ≥ 95; white squares indicate low support: IQ-TREE SH-aLRT < 80/UFBoot < 95. *Although a few *Sycomorus* are monoecious, we refer to the clade that groups subgenera *Ficus*, *Sycidium*, *Sycomorus* and *Synoecia* as the "gynodioecious clade" for brevity. [Colour figure can be viewed at wileyonlinelibrary.com]

Morphological study

The morphological matrix is provided in Appendix S2. The nexus file that contains the majority-rule consensus tree obtained from the morphological data and the four conflicting RAD topologies on which reconstruction of ancestral character states was performed with Mesquite v.3.31 (Maddison and Maddison, 2018) can be opened with PAUP* or Mesquite. Among the 102 morphological characters used, 100 were parsimony-informative. The heuristic search yielded 213 equally parsimonious trees of 736 steps long (CI = 0.443, RI = 0.710). The majority-rule consensus (MRC) tree and the strict consensus trees are depicted in Fig. 3. The consistency index of each character is provided in Appendix S1.

As expected, statistical support was generally low. Only a few nodes, all of them also supported in the molecular tree, received bootstrap supports > 80: sections *Eriosycea* (BP = 100), *Malvanthera* (BP = 96); *Pharmacosycea* (BP = 87) and *Urostigma* (BP = 92). Three main clades were recovered: (i) a monophyletic subg. *Pharmacosycea*; (ii) subg. *Urostigma*; (iii) the "gynodioecious clade". Interestingly, the MRC tree differed slightly from the outgroup rooted molecular hypothesis (Topology 1; Fig. 1). The main differences were: (i) the branching order of the most basal nodes with sect. *Oreosycea* + sect. *Pharmacosycea* (BP = 54) sister to all other *Ficus*, whereas only sect. *Pharmacosycea* was recovered as sister to the remaining fig trees in the molecular topology; (ii) the position of *F. carica* that is sister to section *Eriosycea* in the morphological tree vs. sister to subg. *Sycidium* in the RAD trees; (iii) sect. *Urostigma* sister to sect. *Conosycea* vs. sister to all other sections of *Urostigma*; and (iv) species of subg. *Synoecia* forming a grade within subg. *Ficus*, although it is monophyletic and nested within subg. *Ficus* in the molecular trees. Character transformations inferred with PAUP* on the four competing molecular topologies are illustrated in Fig. S9 (ACCTRAN optimization). Topologies 1 and 4 were the less compatible with morphological data, whereas Topology 2 was supported by the highest number of unambiguous transformations followed by Topology 3.

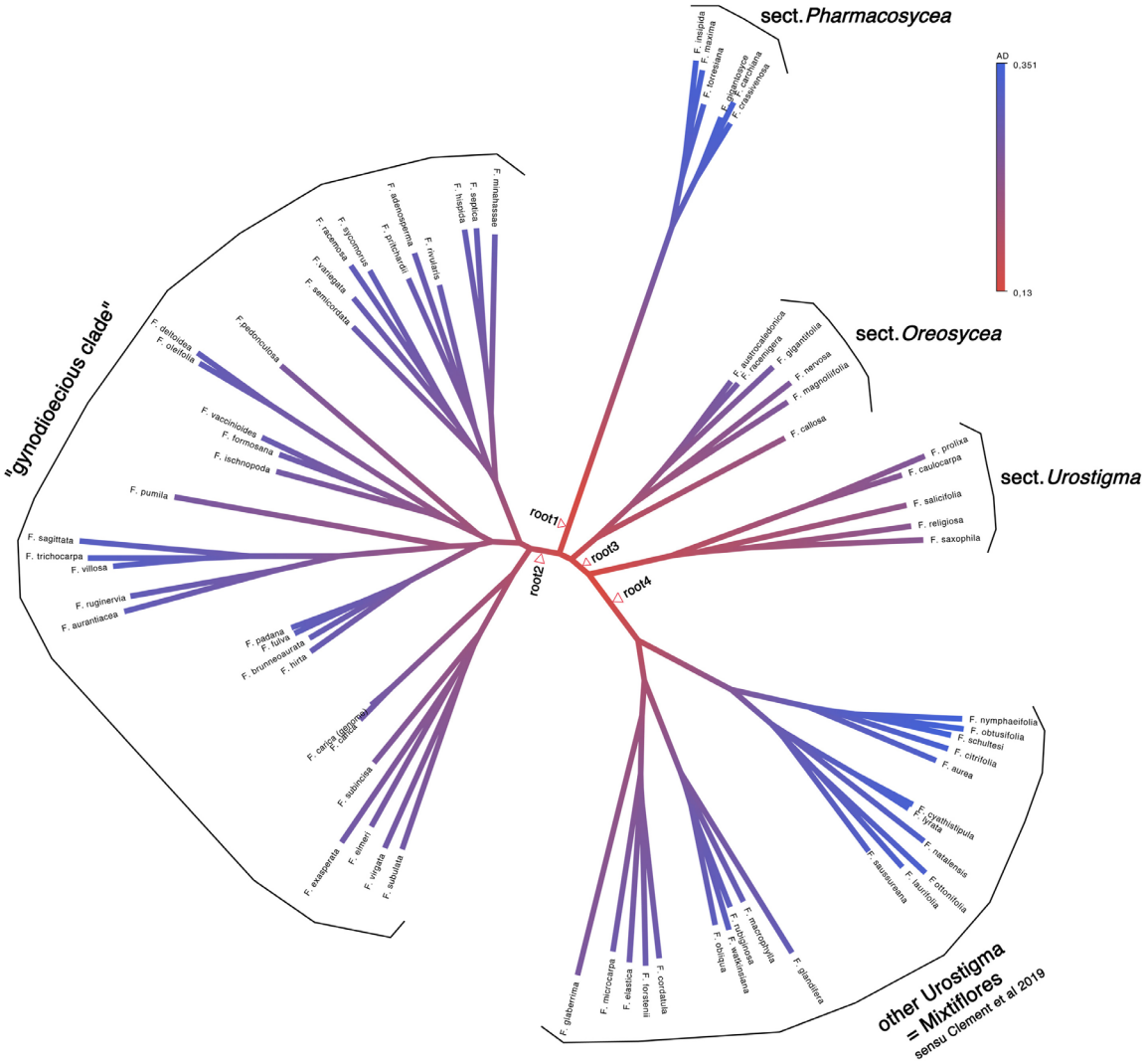
Reconstruction of traits evolution under stochastic mapping

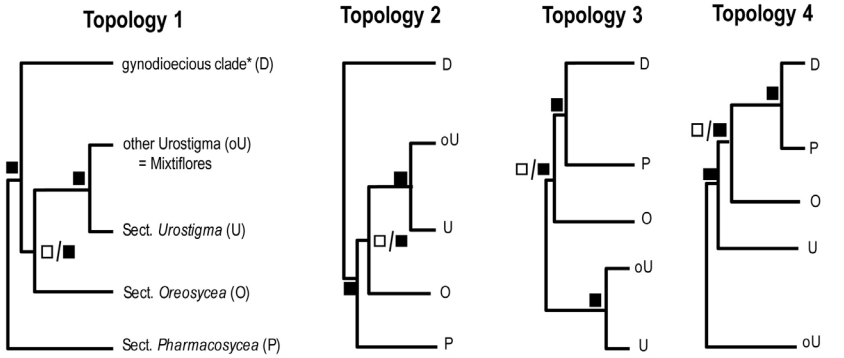
Because of the low resolution of the morphological tree, reconstructions were performed on the molecular trees only (ML). For all topologies and traits, the ER model had the lowest AIC and highest Akaike weight (Table S6). Therefore, the ER model subsequently was chosen to trace the evolution of breeding system, pollination mode and life form in *Ficus* (Fig. S10). Results are given in Table 5, and revealed that the ancestor of all extant *Ficus* was most likely an actively pollinated, monoecious tree from which hemi-epiphytes/hemi-epilithes and root climbers evolved. Gynodioecy appeared once in the genus and monoecy re-appeared at least twice in the "gynodioecious clade". Active pollination was lost several times independently.

Discussion

A first phylogenomic hypothesis for fig trees and its morphological counterpart

Here, we propose a new phylogenetic hypothesis for the genus *Ficus* based on genome-wide nuclear markers (Fig. 1). Instead of analyzing a high number of short (built from forward reads only) and variable loci, we analyzed a low number of longer and more conserved loci. The reasons for this choice were two-fold: (i) to decrease potential bias resulting from missing data, saturation or heterogeneity of evolutionary rates among ingroup taxa; and (ii) to enable the retrieval of homologous loci in published genomes of outgroups for which RAD loci were not sequenced because of mutations accumulated in restriction sites. In that sense, our approach could be compared to hybrid capture of Ultra-Conserved Elements but without requiring probe design from reference genomes.





	Rooting method			
Data sets	outgroup	midpoint	MAD	MinVar
mergeR1R2	1	1	4,3,1	1
mergeR1R2_PCA	1	2	1,4,3,2	3
mergeR1R2_LS3	1	1	4,3,1,2	4
mergeR1R2_GCinfmean	1	1	4,3	4
mergeR1R2_GCsupmean	1	1	1,4,3	1

Table 4
Comparison of average Long Branch heterogeneity scores (LB scores) between sect. *Pharmacosycea* and all other fig trees in concatenated datasets (maximum-likelihood, ML)

	mergeR1R2	mergeR1R2_PCA	mergeR1R2_LS3	mergeR1R2_GCinfmean	mergeR1R2_GCsupmean
Sect. *Pharmacosycea*	−8.39	−10.58	−11.40	−9.58	−17.75
Other fig trees	−18.44	−18.00	−18.93	−19.17	−6.97
Wilcoxon test (*P* value)	6.9e-05	0.00018	0.00012	6.3e-05	7.5e-05

Datasets are described in Table 2. Taxa properties are in Table S1. Topologies are available as supplementary data (Figs S1, S6–S8).

Our analyses highlight heterogeneity in both evolutionary rates and GC content among *Ficus* lineages. We show that sect. *Pharmacosycea* has significantly higher LB heterogeneity scores than all other taxa and, regardless of all the attempts made to reduce this bias (custom PCA approach and LS[3]), the branch leading to sect. *Pharmacosycea* is still—by far—the longest (i.e. branch length/evolutionary rates cannot be properly homogenized; Table 4, Figs S1, S6–S8). In addition, the custom PCA approach shows that it was not possible to homogenize GC content of taxa before phylogenetic inference to fit with model assumptions. The GC content of *Mixtiflores* (sect. *Americanae*, *Galoglychia*, *Conosycea*, *Malvanthera*) is significantly higher and GC content of sect. *Pharmacosycea* is significantly lower. As heterogeneity in evolutionary rates and base composition are considered important sources of systematic bias (Brinkmann et al., 2005; Philippe et al., 2017), we built a phylogenetic hypothesis of the same taxa from morphological features. The recovered trees were structured enough (Fig. 3) to allow comparison with the molecular trees.

Agreement between morphological and molecular hypotheses

In agreement with previous molecular studies based on nuclear data, we confirm the monophyly of the subgenera *Sycidium* and *Sycomorus*. The monophyly of *Sycomorus* was not supported in the morphological study by Weiblen (2000) but is confirmed by our morphological analysis. This suggests that the polyphyly of these two subgenera observed by Bruun-Lund et al. (2017) in their phylogenetic hypothesis based on plastid genomes could be due to heteroplasmy.

Although its monophyly has never been questioned by former botanists (Corner, 1958; Berg, 1989), the subgenus *Urostigma* (sacred banyan trees and giant stranglers) had never been recovered as monophyletic in molecular studies so far and *Galoglychia* was missing from Weiblen (2000) to formally test its monophyly. Here, we highlight a strongly supported subgenus *Urostigma* both with molecular and morphological data. This highly diversified and widespread monoecious subgenus presents a relatively uniform morphology over its range and is well-characterized by nonambiguous apomorphies (Fig. 3): all species have aerial roots and they have only one waxy gland located at the base of the midrib.

We highlight a strongly supported clade that groups all gynodioecious fig trees (Fig. 1), which corresponds to a previous circumscription of subgenera within *Ficus* (Table 1; Corner, 1958). This monophyly was already highlighted by Weiblen (2000) and is confirmed here (Fig. 3). Aside from the breeding system, this clade is well-defined by several nonambiguous synapomorphies (species have generally < 10 lateral veins; figs are frequently stipitate; bears more than three ostiolar bracts; the stigma in short-styled pistillate flowers is mostly cylindrical; and the fruits are compressed).

Finally, as observed in previous molecular works, the subgenus *Ficus* appeared polyphyletic in our molecular and morphological trees, even though results differ between the two approaches as discussed in the next section.

Discrepancies between molecular and morphological evidence

On the morphological tree, *F. carica*, the type species of the subg. *Ficus*, is recovered as sister to sect. *Eriosycea*. In the molecular tree, *F. carica* is recovered as sister to subg. *Sycidium* with strong support. This relationship has been observed already with molecular data, albeit with low support (Cruaud et al., 2012). This surprising result nevertheless is supported by four homoplastic synapomorphies: deciduousness; asymmetrical lamina; margin of perianth hairy and the presence of pistillodes in male flowers, although several species of subg. *Pharmacosycea* and *Sycomorus* also have pistillodes. More species are needed to confirm this result, especially species of the subseries *Albipilae* Corner. However, if the molecular position of *F. carica* is confirmed, then subsect. *Frutescentiae* and sect. *Eriosycea* should not be considered as belonging to subg. *Ficus* any more, if we expand the circumscription of subg. *Ficus* to all gynodioecious clades.

The evolutionary history of subg. *Synoecia* seems to be linked to the evolutionary history of subg. *Ficus* in

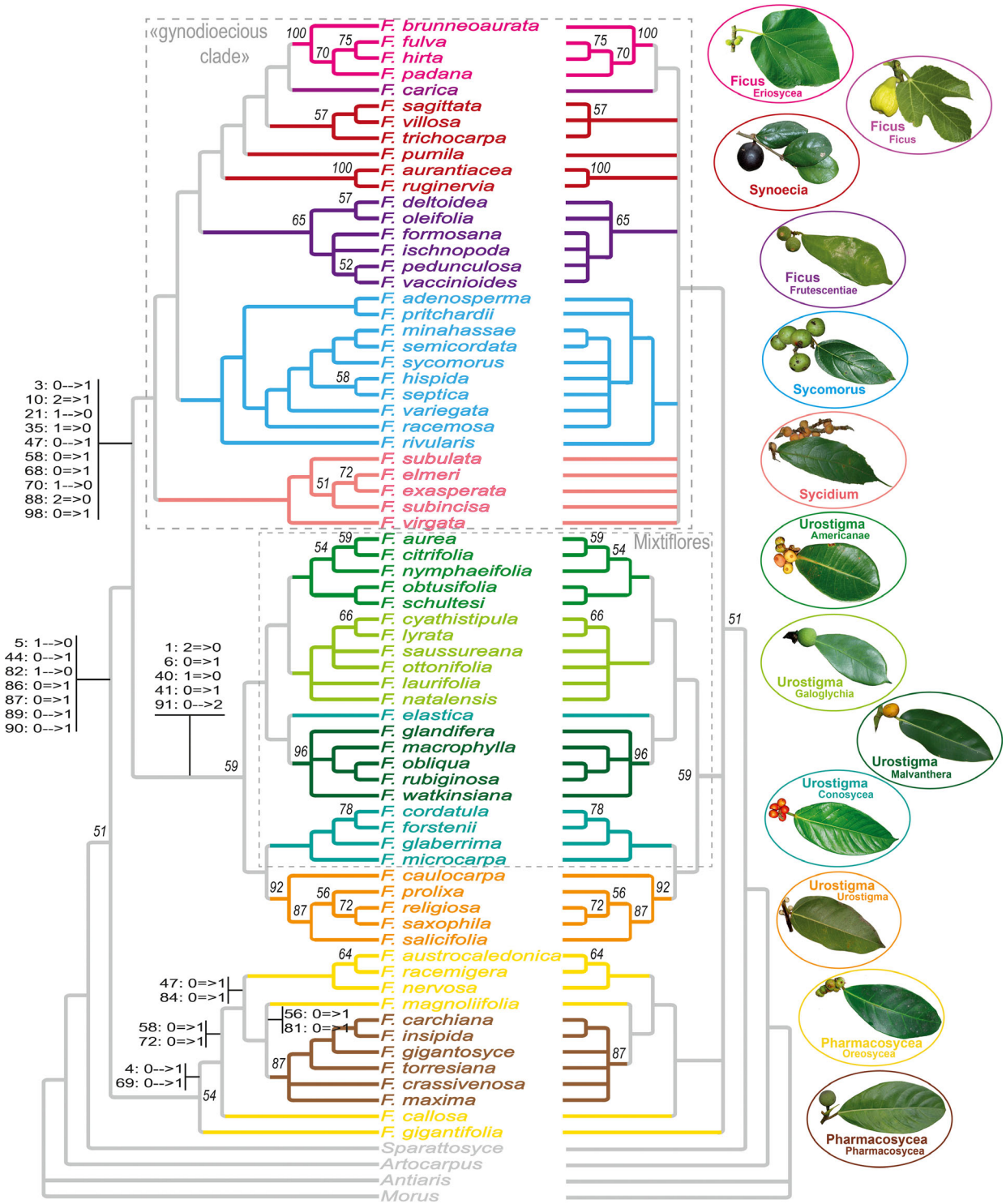


Fig. 3. Phylogenetic trees obtained with the morphological dataset. Left: majority-rule consensus tree; Right: strict consensus tree. Bootstrap (100 replicates) at nodes. Ambiguous (single arrow) and nonambiguous (double arrow) transformations inferred by PAUP* (ACCTRAN optimization) are listed for key nodes as follows: character: ancestral state →/⇒ derived state (see list of character/states in Appendix S1). Subgenera (larger font size) and (sub)sections (smaller font size) are illustrated with pictograms and highlighted with different colours as for Fig. 1. Classification follows Cruaud et al. (2012) as reported in Table 1. Although a few *Sycomorus* are monoecious, we refer to the clade that groups subgenera *Ficus*, *Sycidium*, *Sycomorus* and *Synoecia* as the "gynodioecious clade" for brevity. [Colour figure can be viewed at wileyonlinelibrary.com]

Table 5
Summary of the reconstruction of traits evolution on the molecular trees

	Topology 1	Topology 2	Topology 3	Topology 4
Breeding system	Monoecious	Ambiguous (gynodioecious/monoecious)	Monoecious	Monoecious
Pollination mode	Active	Active	Active	Active
Life form	Tree	Tree	Ambiguous (tree/hemi-epiphyte)	Hemi-epiphyte

The character state of the most recent common ancestor to all fig trees is mentioned for the three analyzed traits. The full reconstructions can be found in Fig. S10. NB The word hemi-epiphytes is used for all species with aerial roots (i.e. hemi-epiphytes s.s. and the few hemi-epilithes).

its current circumscription. Indeed, in all molecular studies that were representative enough of the biodiversity of the genus, *Synoecia* always clustered with sect. *Eriosycea* and subsect. *Frutescentiae*. In the morphological tree of Weiblen (2000), subg. *Synoecia* appeared monophyletic and sister to sect. *Eriosycea* (no representative of the subsect. *Frutescentiae* was included). The same sister taxa relationship is observed in our ML and ASTRAL trees, although this is the only part of the tree that received low support. The results of the parsimony analysis (subg. *Synoecia* + subsect. *Frutescentiae*) may be a consequence of difficulties in handling heterogeneity of GC content among loci (see results). *Synoecia* is not recovered as monophyletic in our morphological tree. However, this may be the result of a high level of homoplasy in the analyzed characters. Again, further studies are needed, but *Synoecia* may simply constitute a lineage that has evolved as root climbers as originally suggested by Corner (1965).

The second discrepancy between our morphological and molecular results concerns the relationships between sect. *Malvanthera*, *Conosycea* and *Urostigma* of subg. *Urostigma*. *Malvanthera* and *Conosycea* are recovered sister in all molecular analyses based on nuclear data including ours. However, all *Conosycea* except *F. elastica* are sister to *Urostigma* in our morphological tree. The morphological results may be due to a lack of signal as all species have aerial roots and show a relatively uniform morphology—at least for a set of characters that are meaningful across the entire genus. These observations lead us to consider that the molecular hypotheses better reflect the history of the subgenus *Urostigma*.

The last discrepancy between morphological and molecular data concerns the subgenus *Pharmacosycea*. Sections *Pharmacosycea* and *Oreosycea* form one of the few supported clades of our morphological tree (BP > 50; Fig. 3), whereas they form a grade in the molecular tree (Fig. 1). The monophyly of the subgenus *Pharmacosycea* has never been questioned by former botanists (Corner, 1958; Berg, 1989) but has been challenged by all molecular analyses published so far. It is noteworthy that the morphological strict consensus tree presented by Weiblen (2000) was modified to show a monophyletic sect. *Pharmacosycea* that was present in the majority-rule consensus of the bootstrap trees but not in the most parsimonious tree in which *F. albipila* (sect. *Oreosycea*) was sister to *F. insipida* (sect. *Pharmacosycea*) (see legend of Fig. 5 and text in Weiblen (2000)). As for other groups of fig trees, apomorphies shared by sect. *Pharmacosycea* and *Oreosycea* are difficult to find because there are always a few species that differ from the original ground plan. However, if we consider unambiguous apomorphies that are shared between sect. *Pharmacosycea* and at least one species of sect. *Oreosycea*, six homoplastic characters can be retained (Fig. 3): epidermis of petiole flaking off; absence of coloured spot on figs; fig stipe present; staminate flowers scattered among pistillate flowers; pistillode present; pistillate perianth partially connate with tepals fused basally.

It could be argued that morphological results are due to convergence and this argument cannot be definitely ruled out. However, the morphological tree shows a high level of congruence with the molecular tree for other *Ficus* groups and it is difficult to understand why morphology would be misleading only for this subgenus. The ecology of the two sections is close but so is the ecology of all species from subg. *Urostigma* for instance. In contrast, exploration of GC content and evolutionary rates clearly shows that sect. *Pharmacosycea* does not exhibit the same properties as all other fig trees, which could mislead molecular inferences. Denser sampling of subg. *Pharmacosycea* in future molecular works, including species of sect. *Oreosycea* subseries *Albipilae* Corner, may help to better resolve relationships between these two groups. Although sect. *Pharmacosycea* may not render sect. *Oreosycea* paraphyletic as observed in the morphological tree, they could be at least closely related.

Rooting issues, higher-level relationships and clues from the pollinator tree of life

Deeper phylogenetic relationships remain the most problematic issue. The unrooted molecular tree highlights the problem we face to root the tree-of-life of fig trees (Fig. 2): (i) a long branch leading to a recent

diversification of sect. *Pharmacosycea*, and (ii) short surrounding branches supporting the "gynodioecious clade", sect. *Oreosycea* and subg. *Urostigma*, suggesting fast diversification of the ancestors of the present lineages. This pattern is predicted to favour artefactual rooting of trees when distant outgroups are used. Further, the recently developed minimal ancestor deviation (MAD) approach that is robust to variation in evolutionary rates among lineages (Tria et al., 2017) shows that root inference is problematic in the original dataset (mergeR1R2) and in all other datasets built to test for potential bias (Fig. 2).

Four competing topologies are suggested. Given the long branch leading to sect. *Pharmacosycea* and the impossibility of homogenizing its evolutionary rate (and GC content) with those of other fig trees, it seems reasonable to suspect that Topology 1 results from an LBA artefact. Indeed, long-branched taxa can cluster with outgroups with high statistical support irrespective of their true phylogenetic relationships (convergent changes along the two long branches are interpreted as false synapomorphies because current models do not reflect evolutionary reality; Phillips et al., 2004). It is known that LBA tends to be reinforced as more and more data are considered (e.g. Boussau et al., 2014), which is probably the case here, even though we used an outgroup belonging to Castillae—the closest relative of *Ficus* (Clement and Weiblen, 2009). So far, in all molecular studies, sect. *Pharmacosycea* was recovered as sister to all other fig trees (with low or high support, but see fig. 1 of Clement *et al*. (2020), which depicts the evolutionary history of Involucraoideae, where section *Albipilae* was sister to all other fig trees with low support; a result that is not recovered from the dataset centered on *Ficus* spp. on their fig. 2]. Topology 1 contradicts morphological evidence (Fig. 3) and previous classification (Table 1). We must note that the imbalance between overall short internal branches and long branches leading to sect. *Pharmacosycea* on the one hand and outgroups on the other is observed recurrently in all molecular-based analyses of the *Ficus* phylogeny. Hence, all molecular analyses might have been trapped by an LBA artefact with a misplacement of the root on the long branch leading to sect. *Pharmacosycea*.

There is a last line of evidence that should be considered to critically interpret the phylogenies presented here: the evolutionary history of the pollinators. Interestingly, the position of the pollinators of sect. *Pharmacosycea* (genus *Tetrapus*) as sister to all other species of Agaonidae recovered in early molecular studies (Machado et al., 2001; Lopez-Vaamonde et al., 2009; Cruaud et al., 2010) has been shown to result from an LBA artefact (Cruaud et al., 2012). In addition, pollinators of sect. *Oreosycea* (*Dolichoris*) and sect. *Pharmacosycea* (*Tetrapus*) also are grouped by morphology (see supplementary data of Cruaud et al., 2012). Notably *Dolichoris* and *Tetrapus* share a unique metanotal structure in males of agaonids (fig. S14A,B in Cruaud et al., 2012), which advocates close relationships instead of convergence between their host figs. Although they cannot be used as direct evidence, these results are circumstantial evidence suggesting that Topology 1 does not accurately reflect early divergence events within fig trees.

There are three known techniques to reduce LBA (Bergsten, 2005). The first possibility is reducing the branch length of the ingroup taxa that are drawn towards the outgroups. Here, we show that the branch leading to sect. *Pharmacosycea* was still significantly longest regardless of the attempt made to reduce this bias. The second possibility is outgroup removal (Bergsten, 2005). However, as sect. *Pharmacosycea* is among the first lineages to diverge, this method is not helpful. The third possibility is increasing sampling. This should definitely be attempted in the future but without guarantee as LBA may be too strong to be broken (Boussau et al., 2014). More generally, given strong bias highlighted here, increasing species sampling appears at least as relevant (if not more) as increasing the number of sequenced regions for each species analyzed.

Topology 4 is the least supported by morphological data (Fig. S9) and appears unlikely given the highly supported monophyly of subg. *Urostigma* in morphological analyses (Fig. 3). A possible explanation for its recovery would be the high GC content of *Mixtiflores* that could be linked to a high number of single nucleotide polymorphisms supporting the group and an increased branch length, which distorts the calculation of ancestral deviation. In addition, given the life forms observed in Moraceae it appears most likely that the ancestor of fig trees was a freestanding tree, which contradicts the scenario of trait evolution based on Topology 4 (Table 5, Fig. S10).

Therefore, two topologies still remain likely (topologies 2 & 3; Fig. 1). They differ only by the position of the grade composed by sections *Oreosycea* and *Pharmacosycea*. In Topology 2, these sections cluster with subg. *Urostigma* and the genus *Ficus* is divided into two groups: monoecious and gynodioecious species. In Topology 3, the sections cluster with the "gynodioecious clade" and the genus is split into species with aerial roots on one side and other fig trees on the other side. Topology 2 is the most supported by morphological data on fig trees (Fig. S9). Importantly, unambiguous transformations that support the monoecious/gynodioecious split are not only linked to breeding system (or pollination mode). Indeed, characters supporting the split are related to tree height; number of lateral veins; fig stipitate or not; pistillate flowers sessile or not; shape of stigma in short-styled pistillate

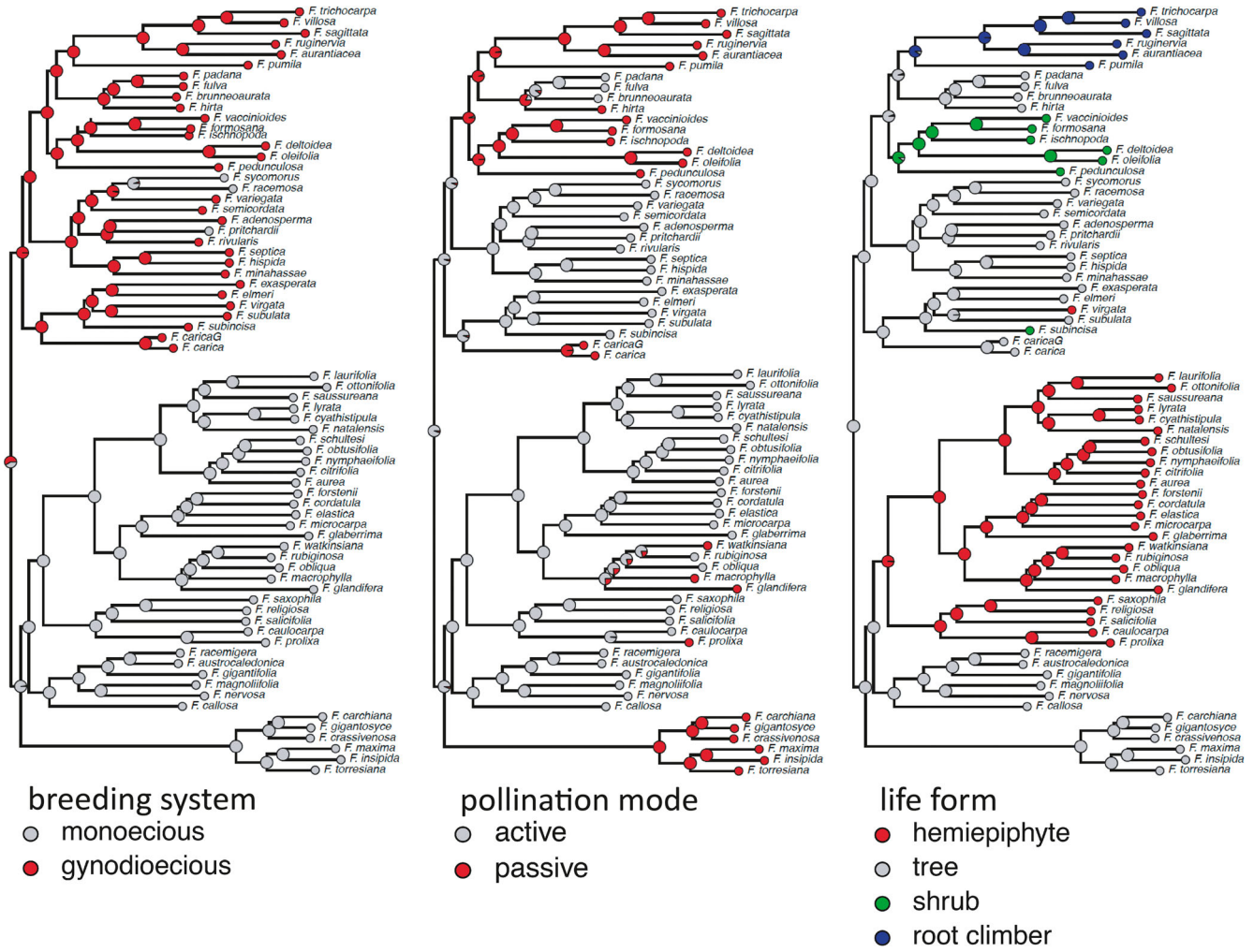


Fig. 4. Reconstruction of traits evolution on the favored topology (Topology 2). All reconstructions are available in Fig. S10. NB the word hemi-epiphytes is used for all species with aerial roots (i.e. hemi-epiphytes s.s. and the few hemi-epilithes). [Colour figure can be viewed at wiley onlinelibrary.com]

flowers; lamina thickness; lamina margin; furcation of lateral veins; tertiary venation; presence of interfloral bracts. In contrast, Topology 3 and, more specifically, the sister taxa relationships between sect. *Oreosycea*, sect. *Pharmacosycea* and the "gynodioecious clade", is not supported by any unambiguous morphological transformation (Fig. S9).

Finally, the current molecular phylogenetic hypothesis of pollinators of fig trees provides more support for Topology 2. Indeed, reconciliation of the agaonid tree-of-life with Topology 3 would require a partial reversal of the pollinator tree that contradicts transformation series for the structure of male mesosoma and female basal flagellomeres (anelli) (see Cruaud et al., 2012, fig. S14). The current pollinator tree supports a progressive fusion of male mesosoma and female anelli, whereas Topology 3 would favour a subdivision of mesosoma and basal flagellomere as the most derived character state, a trend that has never been observed in Hymenoptera.

Therefore, after considering all of the evidence (bias, morphology, pollinators) we consider that the most likely topology for the *Ficus* tree-of-life is Topology 2. Interestingly, this topology agrees with one of the first statements of Corner (1958): "the first division of *Ficus* is into the monoecious and dioecious species, but it is more convenient to recognize three subgenera, namely the monoecious banyans (*Urostigma*), the monoecious trees (*Pharmacosycea*) and the dioecious *Ficus*." Nevertheless, increasing sampling is required to resolve the root of the *Ficus* tree-of-life. Until now, no investigation has been undertaken to identify another root for *Ficus* than the one identified by outgroup rooting. Resolving this issue is important not only to understand the evolution of *Ficus*, but also to our understanding of key questions such as how species' traits

have changed over time and how fig trees co-diversified with their pollinating wasps.

On the way of revisiting the evolution of life-history traits

Although a conclusive topology still evades us, we can go one step forward in our understanding of the evolution of key life-history traits. It appears most likely that the ancestor of fig trees was a freestanding tree. In addition, most growth forms have evolved from tree ancestors. It seems that phylogenetic or biogeographical constraints likely played an important role in the evolution of growth forms and convergence may be less common than thought previously (monophyly of subgen. *Urostigma*; monophyly of subg. *Synoecia* + *F. pumila*).

There is a consensus concerning the ancestral pollination mode that is inferred as active for all topologies. This result contradicts the widely held hypothesis that passive pollination was the ancestral pollination mode in *Ficus*. Instead, it suggests that from an active pollination mode fig trees became passive multiple time independently overtime, with no reversion possible to an active mode. In our analysis, the only case where this assumption appears contradicted is the apparent reversion to an active pollination within *Eriosycea*. This reversion may just be artefactual, due to the incomplete sampling analyzed. Indeed, *F. laevis*, a species sister to all other *Eriosycea* (Li et al., 2012), is an active fig tree.

Finally, ambiguity remains for the ancestral breeding system (including for our favoured topology; Fig. 4). However, from our results, gynodioecy appears to have evolved only once in *Ficus*.

Conclusion

We present a phylogenetic hypothesis for the genus *Ficus* based on genome-wide nuclear markers and a sampling effort representative of all subgenera and sections. For the first time, we recover a monophyletic and strongly supported subgenus *Urostigma* and a strongly supported clade grouping all gynodioecious fig trees. Our analysis of biases, general pattern of rooting preferences and morphological data completed with indirect evidence from the pollinator tree-of-life, highlights that previous molecular studies might have been trapped by LBA, which resulted in an artefactual placement of sect. *Pharmacosycea* as sister to all other fig trees. The next key step will be to increase sample size. Indeed, confidence in phylogenetic inference and trait evolution should increase with increasing datasets encompassing as much as possible the overall diversity in the studied group. Increasing species sampling appears at least as relevant as increasing the number of sequenced regions for each species analyzed. This work is a first step towards a clarification of the classification and evolutionary history of fig trees. Taxonomic changes foreseen by previous molecular works and new ones will need to be undertaken. But taxonomists will need to be cautious and humble as invalidating current and widely used names without due care and diligence will generate more confusion than clarity.

Acknowledgements

We thank S. Santoni (INRAE) for technical advices, MGX for sequencing, the Genotoul bioinformatics platform Toulouse Midi-Pyrenees for computing resources, two anonymous reviewers for their careful reading of our manuscript and C.C. Berg and E.J.H. Corner for their eternal contributions. This work was partially funded by UPD-OVCRD (171715 PhDIA) and UPD-NSRI (BIO-18-1-02). AC thanks C.J. Rivera-Rivera for his help with running LS^3.

Conflict of interest

None declared.

Author contributions

J.Y.R. and A.C. conceived the study, analyzed data and drafted the manuscript; J.Y.R., L.J.R., Y.Q.P, D.R.Y, F.K., A.B, R.D.H., R.U., R.A.S.P, S.F and A.C. collected samples; J.Y.R. identified or verified identification of all samples; L.J.R., L.S. C.T.C. and A.C. performed lab work; and M.G and J.P.R. contributed scripts. All authors gave final approval for publication.

Data availability statement

Raw paired reads are available as NCBI Sequence Read Archives (BioProject IDs PRJNA474284 and PRJNA668043). Pipeline and scripts are available from: https://github.com/acruaud/radseq_ficus_2020. Datasets and trees are available from: http://doi.org/10.5281/zenodo.4077409.

References

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. and Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3, e3376.

Berg, C.C., 1989. Classification and distribution of *Ficus*. Experientia 45, 605–611.
Berg, C.C., 2009. Moraceae (Ficus). Department of Plant and Environmental Sciences, University of Gothenburg. Flora of Ecuador. 27C. Göteborg.
Berg, C.C. and Corner, E.J.H., 2005. Moraceae – *Ficus*. Flora Malesiana, Ser. I, 17/2, Leiden.
Berg, C.C., Pattharahirantricin, N. and Chantarasuwan, B. (Eds.), 2011. Moraceae. The Forest Herbarium, Department of National Parks, Wildlife and Plant Conservation, Bangkok.
Berg, C.C. and Wiebes, J.T., 1992. African Fig Trees and Fig Wasps. Verhandelingen der Koninklijke Nederlandse Akademie van Wetenschappen, Amsterdam.
Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163–193.
Blomberg, S.P., Garland, J.T. and Ives, A.R., 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57, 717–745.
Bollback, J.P., 2006. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7, 88.
Borowiec, M.L., 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4, e1660.
Bossert, S., Murray, E.A., Blaimer, B.B. and Danforth, B.N., 2017. The impact of GC bias on phylogenetic accuracy using targeted enrichment phylogenomic data. Mol. Phylogenet. Evol. 111, 149–157.
Boussau, B., Walton, Z., Delgado, J.A., Collantes, F., Beani, L., Stewart, I.J., Cameron, S.A., Whitfield, J.B., Johnston, J.S., Holland, P.W.H. *et al*., 2014. Strepsiptera, phylogenomics and the long branch attraction problem. PLoS One 9, e107709.
Brinkmann, H., van der Giezen, M., Zhou, Y., Poncelin de Raucourt, G. and Philippe, H., 2005. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. Syst. Biol. 54, 743–757.
Bruun-Lund, S., Clement, W.L., Kjellberg, F. and Rønsted, N., 2017. First plastid phylogenomic study reveals potential cyto-nuclear discordance in the evolutionary history of *Ficus* L. (Moraceae). Mol. Phylogenet. Evol. 109, 93–104.
Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. and Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124–3140.
Chantarasuwan, B., Berg, C.C., Kjellberg, F., Ronsted, N., Garcia, M., Baider, C. and van Welzen, P.C., 2015. A new classification of *Ficus* subsection *Urostigma* (Moraceae) based on four nuclear DNA markers (ITS, ETS, G3pdh, and ncpGS), morphology and leaf anatomy. PLoS One 10, e0128289.
Clement, W.L., Bruun Lund, S., Cohen, A., Kjellberg, F., Weiblen, G.D. and Rønsted, N., 2020. Evolution and classification of figs (*Ficus*) and their close relatives (Castilleae) united by involucral bracts. Bot. J. Linn. Soc. (193), 316–339.
Clement, W.L. and Weiblen, G.D., 2009. Morphological evolution in the mulberry family (Moraceae). Syst. Bot. 34, 530–552.
Cook, J.M. and Rasplus, J.-Y., 2003. Mutualists with attitude: coevolving fig wasps and figs. Trends Ecol. Evol. 18, 241–248.
Corner, E.J.H., 1938. A revision of *Ficus*, subgenus *Synoecia* Gardens'. Bull. Singapore 10, 82–161.
Corner, E.J.H., 1958. An introduction to the distribution of *Ficus*. Reinwardtia 4, 325–355.
Corner, E.J.H., 1960a. Taxonomic notes on *Ficus* Linn., Asia and Australasia I. Subgen. *Urostigma* (Gasp.) Miq. Gard. Bull. Singapore 17, 368–485.
Corner, E.J.H., 1960b. Taxonomic notes on *Ficus* Linn., Asia and Australasia II. Subgen. *Pharmacosycea* Miq. Gard. Bull. Singapore 17, 405–415.
Corner, E.J.H., 1960c. Taxonomic notes on *Ficus* Linn., Asia and Australasia III. Subgen. *Ficus* and sect. *Ficus*. Gard. Bull. Singapore 17, 416–441.
Corner, E.J.H., 1960d. Taxonomic notes on *Ficus* Linn., Asia and Australasia IV. Subgen. *Ficus* sect. *Sycidium* Miq. Gard. Bull. Singapore 17, 442–485.
Corner, E.J.H., 1960e. Taxonomic notes on *Ficus* Linn., Asia and Australasia V. Subgen. *Ficus* sect. *Rhizocladus, Kalosyce, Sinosycidium, Adenospermaand Neomorphe*. Gard. Bull. Singapore 18, 1–35.
Corner, E.J.H., 1965. Check-list of *Ficus* in Asia and Australasia with a key to identification. Gard. Bull. Singapore 21, 1–186.
Corner, E.J.H., 1967. *Ficus* in the Solomon Islands and its bearing on the post-Jurassic history of Melanesia. Phil. Trans. R. Soc. Lond. B 253, 23–159.
Corner, E.J.H., 1969a. The complex of *Ficus deltoidea*; a recent invasion of the Sunda Shelf. Phil. Trans. R. Soc. Lond. B 256, 281–317.
Corner, E.J.H., 1969b. *Ficus* sect. *Adenosperma*. Phil. Trans. R. Soc. Lond. B 256, 319–355.
Corner, E.J.H., 1970. *Ficus* subgen. *Pharmacosycea* with reference to the species of New Caledonia. Phil. Trans. R. Soc. Lond. B 259, 383–433.
Corner, E.J.H., 1978a. *Ficus dammaropsis* and the multibracteate species of *Ficus* sect. *Sycocarpus*. Phil. Trans. R. Soc. Lond. B 281, 373–406.
Corner, E.J.H., 1978b. *Ficus glaberrima* Bl. and the pedunculate species of *Ficus* subgen. *Urostigma* in Asia and Australasia. Phil. Trans. R. Soc. Lond. B 281, 347–371.
Cruaud, A., Delvare, G., Nidelet, S., Sauné, L., Ratnasingham, S., Chartois, M., Blaimer, B.B., Gates, M., Brady, S.G., Faure, S. *et al*., Ultra-Conserved Elements and morphology reciprocally illuminate conflicting phylogenetic hypotheses in Chalcididae (Hymenoptera, Chalcidoidea). Cladistics, 1–35. https://doi.org/10.1101/761874
Cruaud, A., Gautier, M., Galan, M., Foucaud, J., Sauné, L., Genson, G., Dubois, E., Nidelet, S., Deuve, T. and Rasplus, J.-Y., 2014. Empirical assessment of RAD sequencing for interspecific phylogeny. Mol. Biol. Evol. 31, 1272–1274.
Cruaud, A., Gautier, M., Rossi, J.-P., Rasplus, J.-Y. and Gouzy, J., 2016. RADIS: analysis of RAD-seq data for InterSpecific phylogeny. Bioinformatics 32, 3027–3028.
Cruaud, A., Jabbour-Zahab, R., Genson, G., Cruaud, C., Couloux, A., Kjellberg, F., van Noort, S. and Rasplus, J.-Y., 2010. Laying the foundations for a new classification of Agaonidae (Hymenoptera: Chalcidoidea), a multilocus phylogenetic approach. Cladistics 26, 359–387.
Cruaud, A., Rønsted, N., Chantarasuwan, B., Chou, L.S., Clement, W., Couloux, A., Cousins, B., Forest, F., Genson, G., Harrison, R.D. *et al*., 2012. An extreme case of plant-insect co-diversification: figs and fig-pollinating wasps. Syst. Biol. 61, 1029–1047.
Dray, S. and Dufour, A., 2007. The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22, 1–20.
Duchêne, D.A., Bragg, J.G., Duchêne, S., Neaves, L.A., Potter, S., Moritz, C., Johnson, R.N., Ho, S.Y.W. and Eldridge, M.D.B., 2018. Analysis of phylogenomic tree space resolves relationships among marsupial families. Syst. Biol. 67, 400–412.
Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A. and Cresko, W.A., 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In: Orgogozo, V. and Rockman, M.V. (Eds.), Molecular Methods for Evolutionary Genetics. Humana Press, New York, NY, pp. 157–178.
Galil, J., 1977. Fig biology. Endeavour 1, 52–56.
Gautier, M., Gharbi, K., Cézard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., Cornuet, J.-M. and Estoup, A., 2013. The effect of RAD allele drop-out on the estimation of genetic variation within and between populations. Mol. Ecol. 22, 3165–3178.
Giribet, G., 2015. Morphology should not be forgotten in the era of genomics – a phylogenetic perspective. Zool. Anzeig. 256, 96–103.
Gori, K., Suchan, T., Alvarez, N., Goldman, N. and Dessimoz, C., 2016. Clustering genes of common evolutionary history. Mol. Biol. Evol. 33, 1590–1605.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M. *et al*., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512.

Harris, R.S., 2007. Improved pairwise alignment of genomic DNA. Ph.D. Thesis. The Pennsylvania State University.

Harrison, R.D., 2005. Figs and the diversity of tropical rainforests. Bioscience 55, 1053–1064.

Harrison, R.D. and Shanahan, M., 2005. Seventy-seven ways to be a fig: an overview of a diverse assemblage of figs in Borneo. In: Roubik, D.W., Sakai, S. and Hamid, A.A. (Eds.) Pollination Ecology and the Rain Forest Canopy: Sarawak Studies. Springer Verlag, New York, NY, Vol. 111–127, pp. 246–249.

Herre, E.A., Jandér, K.C. and Machado, C.A., 2008. Evolutionary ecology of figs and their associates: recent progress and outstanding puzzles. Annu. Rev. Ecol. Evol. Syst. 39, 439–458.

Herre, E.A., Machado, C.A., Bermingham, E., Nason, J.D., Windsor, D.M., McCafferty, S., Van Houten, W. and Bachmann, K., 1996. Molecular phylogenies of figs and their pollinator wasps. J. Biogeogr. 23, 521–530.

Hipp, A.L., Manos, P.S., Hahn, M., Avishai, M., Bodénès, C., Cavender-Bares, J., Crowl, A.A., Deng, M., Denk, T., Fitz-Gibbon, S. *et al*., 2020. Genomic landscape of the global oak phylogeny. New Phytol. 226, 1198–1212.

Hoang, D.T., Vinh, L.S., Flouri, T., Stamatakis, A., von Haeseler, A. and Minh, B.Q., 2018. MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. BMC Evol. Biol. 18, 11.

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L. and Kimball, R.T., 2016. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). Mol. Biol. Evol. 33, 1110–1125.

Huelsenbeck, J.P., Nielsen, R. and Bollback, J.P., 2003. Stochastic mapping of morphological characters. Syst. Biol. 52, 131–158.

Jandér, K.C. and Herre, E.A., 2010. Host sanctions and pollinator cheating in the fig tree–fig wasp mutualism. Proc. R. Soc. Lond. B 277, 1481–1488.

Jousselin, E., Rasplus, J.Y. and Kjellberg, F., 2003. Convergence and coevolution in a mutualism evidence from a molecular phylogeny of *Ficus*. Evolution 57, 1255–1272.

Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kjellberg, F., Jousselin, E., Bronstein, J.L., Patel, A., Yokoyama, J. and Rasplus, J.Y., 2001. Pollination mode in fig wasps: the predictive power of correlated traits. Proc. R. Soc. Lond. B 268, 1113–1121.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 26, 2078–2079.

Li, H.Q., Wang, S., Chen, J.Y. and Gui, P., 2012. Molecular phylogeny of *Ficus* section *Ficus* in China based on four DNA regions. J. Syst. Evol. 50, 422–432.

Lopez-Vaamonde, C., Cook, J.M., Rasplus, J.-Y., Machado, C.A. and Weiblen, G., 2009. Molecular dating and biogeography of fig-pollinating wasps. Mol. Phylogenet. Evol. 52, 715–726.

Machado, C.A., Jousselin, E., Kjellberg, F., Compton, S. and Herre, E.A., 2001. Phylogenetic relationships, historical biogeography and character evolution of fig-pollinating wasps. Proc. R. Soc. Lond. B 268, 685–694.

Maddison, W.P. and Maddison, D.R., 2018. Mesquite: a modular system for evolutionary analysis. Version 3.51. http://www.mesquiteproject.org

Mai, U. and Mirarab, S., 2018. TreeShrink: fast and accurate detection of outlier long branches in collections of phylogenetic trees. BMC Genom. 19, 272.

Mai, U., Sayyari, E. and Mirarab, S., 2017. Minimum variance rooting of phylogenetic trees and implications for species tree reconstruction. PLoS One 12, e0182238.

Manly, B.F.J. and Alberto, J.A.N., 2017. Multivariate Statistical Methods: A Primer. Boca Raton: Chapman and Hall/CRC.

Minh, B.Q., Nguyen, M.A.T. and von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30, 1188–1195.

Mirarab, S., Nguyen, N. and Warnow, T., 2014. PASTA: ultra-large multiple sequence alignment. Res. Comput. Mol. Biol. 22, 177–191.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A. and Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.

Ohri, D. and Khoshoo, T.N., 1987. Nuclear DNA contents in the genus *Ficus* (Moraceae). Plant Syst. Evol. 156, 1–4.

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

Pederneiras, L.C., Gaglioti, A.L., Romaniuc-Neto, S. and de Freitas Mansano, V., 2018. The role of biogeographical barriers and bridges in determining divergent lineages in *Ficus* (Moraceae). Bot. J. Linn. Soc. 187, 594–613.

Peterson, B.G. and Carl, P., 2018. PerformanceAnalytics: econometric tools for performance and risk analysis. R package version 1.5.2. https://CRAN.R-project.org/package=PerformanceAnalytics

Philippe, H., de Vienne, D.M., Ranwez, V., Roure, B., Baurain, D. and Delsuc, F., 2017. Pitfalls in supermatrix phylogenomics. Eur. J. Taxon. 283, 1–25.

Phillips, M.J., Delsuc, F. and Penny, D., 2004. Genome-scale phylogeny and the detection of systematic biases. Mol. Biol. Evol. 21, 1455–1458.

R Core Team, 2018. R Version 3.5.1 (Feather Spray): A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Revell, L.J., 2012. Phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3, 217–223.

Rivera-Rivera, C.J. and Montoya-Burgos, J.I., 2016. LS3: a method for improving phylogenomic inferences when evolutionary rates are heterogeneous among taxa. Mol. Biol. Evol. 33, 1625–1634.

Rivera-Rivera, C.J. and Montoya-Burgos, J.I., 2019. LSX: automated reduction of gene-specific lineage evolutionary rate heterogeneity for multi-gene phylogeny inference. BMC Bioinformatics 20, 420.

Romiguier, J., Cameron, S.A., Woodard, S.H., Fischman, B.J., Keller, L. and Praz, C.J., 2016. Phylogenomics controlling for base compositional bias reveals a single origin of eusociality in corbiculate bees. Mol. Biol. Evol. 33, 670–678.

Rønsted, N., Weiblen, G., Clement, W.L., Zerega, N.J.C. and Savolainen, V., 2008. Reconstructing the phylogeny of figs (*Ficus*, Moraceae) to reveal the history of the fig pollination mutualism. Symbiosis 45, 45–55.

Rønsted, N., Weiblen, G.D., Cook, J.M., Salamin, N., Machado, C.A. and Savolainen, V., 2005. 60 million years of co-divergence in the fig-wasp symbiosis. Proc. R. Soc. Lond. B 272, 2593–2599.

Rubin, B.E.R., Ree, R.H. and Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. PLoS One 7, e33394.

Satler, J.D., Herre, E.A., Jander, K.C., Eaton, D.A.R., Machado, C.A., Heath, T.A. and Nason, J.D., 2019. Inferring processes of coevolutionary diversification in a community of Panamanian strangler figs and associated pollinating wasps. Evolution 73, 2295–2311.

Shanahan, M., So, S., Compton, S. and Corlett, R., 2001. Fig-eating by vertebrate frugivores: a global review. Biol. Rev. Camb. Philos. Soc. 76, 529–570.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Stöver, B.C. and Müller, K.F., 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics 11, 7.
Struck, T.H., 2014. TreSpEx – detection of misleading signal in phylogenetic reconstructions based on tree information. Evol. Bioinform. 10, 51–67.
Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA.
Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M. and Dessimoz, C., 2015. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. Syst. Biol. 64, 778–791.
Tria, F.D.K., Landan, G. and Dagan, T., 2017. Phylogenetic rooting using minimal ancestor deviation. Nat. Ecol. Evol. 1, 193.
Weiblen, G.D., 2000. Phylogenetic relationships of functionally dioecious *Ficus* (Moraceae) based on ribosomal DNA sequences and morphology. Am. J. Bot. 87, 1342–1357.
Wiens, J.J., 2004. The role of morphological data in phylogeny reconstruction. Syst. Biol. 53, 653–661.
Wipfler, B., Pohl, H., Yavorskaya, M.I. and Beutel, R.G., 2016. A review of methods for analysing insect structures – the role of morphology in the age of phylogenomics. Curr. Opin. Insect Sci., 18, 60–68.
Xu, L., Harrison, R.D., Yang, P. and Yang, D.-R., 2011. New insight into the phylogenetic and biogeographic history of genus *Ficus*: vicariance played a relatively minor role compared with ecological opportunity and dispersal. J. Syst. Evol. 49, 546–557.
Zhang, C., Rabiee, M., Sayyari, E. and Mirarab, S., 2018a. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics, 19, 153.
Zhang, Q., Onstein, R.E., Little, S.A. and Sauquet, H., 2018b. Estimating divergence times and ancestral breeding systems in *Ficus* and Moraceae. Ann. Bot. 123, 191–204.
Zhu, Q., 2014. AfterPhylo. A Perl script for manipulating trees after phylogenetic reconstruction. Available from https://github.com/qiyunzhu/AfterPhylo/

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix S1. List of morphological characters and character states.

Appendix S2. List of morphological characters and character states.

Figure S1. Trees_mergeR1R2_seqtools25
Figure S2. Trees_mergeR1R2_seqtools25-50
Figure S3. Correlation_lociproperties
Figure S4. PCA_GCcontent_mergeR1
Figure S5. PCA_LB_mergeR1R2
Figure S6. Trees_mergeR1R2_impactGCcontent
Figure S7. Trees_mergeR1R2_impactLBscores_PCA
Figure S8. Trees_mergeR1R2_impactLBscores_LS3
Figure S9. Transfo_morpho
Figure S9. Traits
Table S1. Sampling_and_stats
Table S2. Taxa_properties_in_loci_mergeR1_DEF
Table S3. Taxa_properties_in_loci_mergeR1R2_DEF
Table S4. Results_treeshrink_DEF
Table S5. Loci_filtered_LS3_DEF
Table S6. Traits_mapping