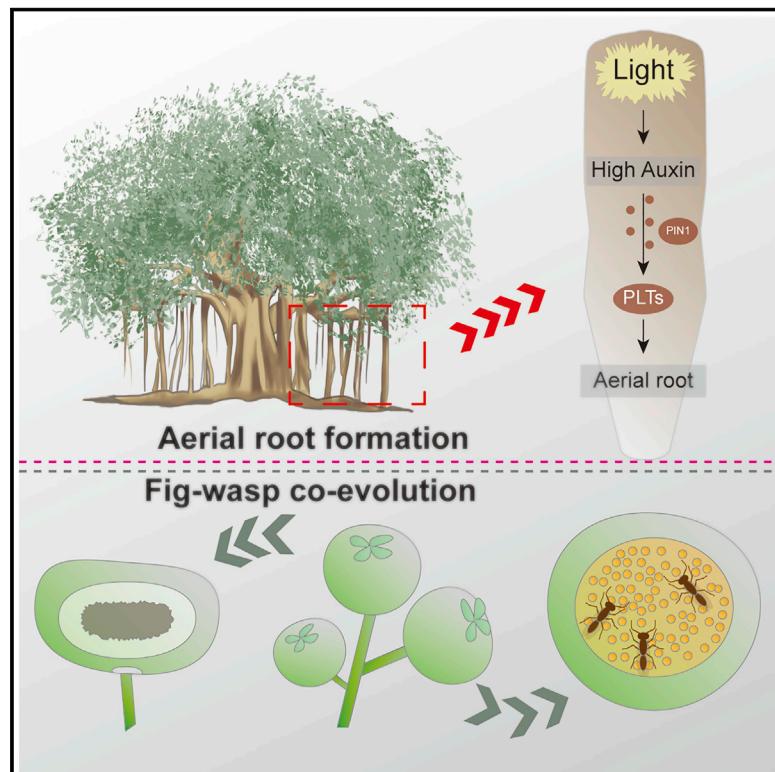


Genomes of the Banyan Tree and Pollinator Wasp Provide Insights into Fig-Wasp Coevolution

Graphical Abstract



Authors

Xingtian Zhang, Gang Wang,
Shengcheng Zhang, ..., Haibao Tang,
Jin Chen, Ray Ming

Correspondence

cj@xtbg.org.cn (J.C.),
rayming@illinois.edu (R.M.)

In Brief

Comparative analyses of two *Ficus* genomes with and without aerial roots reveal insights into the genomic basis of aerial root formation and population genomic analyses of *Ficus* species alongside the genome of a coevolving wasp pollinator enhance our understanding of their species-specific mutualism.

Highlights

- Genomes of two *Ficus* species and one wasp pollinator revealed fig-wasp coevolution
- Aerial root formation is triggered by an auxin-dependent pathway promoted by light
- A male-specific *FhAG2* gene is a candidate gene for sex determination in *F. hispida*
- Genes related to coadaptation contribute to codiversification between fig and wasp



Article

Genomes of the Banyan Tree and Pollinator Wasp Provide Insights into Fig-Wasp Coevolution

Xingtian Zhang,^{1,6} Gang Wang,^{2,6} Shengcheng Zhang,¹ Shuai Chen,¹ Yibin Wang,¹ Ping Wen,² Xiaokai Ma,¹ Yan Shi,¹ Rui Qi,¹ Yang Yang,¹ Zhenyang Liao,¹ Jing Lin,¹ Jishan Lin,¹ Xiuming Xu,¹ Xuequn Chen,¹ Xindan Xu,¹ Fang Deng,¹ Lihua Zhao,¹ Yi-lun Lee,³ Rong Wang,⁴ Xiao-Yong Chen,⁴ Yann-rong Lin,³ Jisen Zhang,¹ Haibao Tang,¹ Jin Chen,^{2,*} and Ray Ming^{5,7,*}

¹Center for Genomics and Biotechnology, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Key Laboratory of Genetics, Breeding and Multiple Utilization of Crops, Ministry of Education, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002

²CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan 666303

³Department of Agronomy, National Taiwan University, Taipei, Taiwan 10617

⁴Zhejiang Tiantong Forest Ecosystem National Observation and Research Station, School of Ecological and Environmental Sciences, East China Normal University, Shanghai 200241

⁵Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

⁶These authors contributed equally

⁷Lead Contact

*Correspondence: cj@xtbg.org.cn (J.C.), rayming@illinois.edu (R.M.)

<https://doi.org/10.1016/j.cell.2020.09.043>

SUMMARY

Banyan trees are distinguished by their extraordinary aerial roots. The *Ficus* genus includes species that have evolved a species-specific mutualism system with wasp pollinators. We sequenced genomes of the Chinese banyan tree, *F. microcarpa*, and a species lacking aerial roots, *F. hispida*, and one wasp genome coevolving with *F. microcarpa*, *Eupristina verticillata*. Comparative analysis of the two *Ficus* genomes revealed dynamic karyotype variation associated with adaptive evolution. Copy number expansion of auxin-related genes from duplications and elevated auxin production are associated with aerial root development in *F. microcarpa*. A male-specific AGAMOUS paralog, *FhAG2*, was identified as a candidate gene for sex determination in *F. hispida*. Population genomic analyses of *Ficus* species revealed genomic signatures of morphological and physiological coadaptation with their pollinators involving terpenoid- and benzenoid-derived compounds. These three genomes offer insights into and genomic resources for investigating the geneses of aerial roots, monoecy and dioecy, and codiversification in a symbiotic system.

INTRODUCTION

Banyan trees are tropical and subtropical plants that are notable for their numerous aerial roots and enormous canopies and are widely used for landscaping. Their aerial roots enable many *Ficus* species to live as hemi-epiphytes, as do the strangler figs often seen in tropical forests. *Ficus* is the largest genus in the Moraceae family, which consists of ~800 species (Berg, 1989), and it is also a relatively ancient genus that is approximately 75 million years old (Craaud et al., 2012).

Ficus species are distributed across a wide variety of ecological niches, implying a high degree of adaptability to environments and their important roles in tropical and subtropical ecosystems (Harrison, 2005). The year-round production of large fruits led to fig species being “keystone resources” in most tropical forests by sustaining a wide range of frugivores, including 1,200 recorded bird and mammal species that, in return, serve as seed dispersers, spreading fig plants across a wide range of habitats (Shanahan

et al., 2001). Individual fig plants are also critical components of tropical forest restoration by attracting and supporting seed dispersers of many different species (Cottee-Jones et al., 2016). Figs are among the earliest domesticated crops and are a recurring theme in religions such as Buddhism and Hinduism (Lev-Yadun et al., 2006). Figs and their obligate pollinator wasp species are a model system for studying fundamental phenomena in ecology and evolution, including cooperation, coevolution, ecological networks, codiversification, and response to climate change (Dunn et al., 2019; Craaud et al., 2012; Herre et al., 2008).

The genus *Ficus* possesses several special characteristics, including an enclosed inflorescence, fig-wasp obligate mutualism, a sex determination system, and some species with a hemi-epiphytic habit. Hemi-epiphytic figs (strangler or banyan figs) are a fascinating form of plant life with seedlings that germinate and grow as epiphytes on the canopy branches of living host trees with extraordinary aerial roots that descend along the trunks of host trees into the soil. *Ficus* plants are easily recognized



Table 1. Statistics for Assembly and Annotation of the Fm and Fh Genomes

ChrID	Fm		Fh	
	No. of Contigs	Length (bp)	No. of Contigs	Length (bp)
Chr1	249	27,719,661	186	24,287,689
Chr2	459	34,781,301	392	24,073,434
Chr3	598	57,014,673	163	23,296,499
Chr4	255	22,703,967	142	20,182,075
Chr5	467	42,015,715	237	32,764,157
Chr6	289	22,471,607	258	23,181,460
Chr7	363	23,317,306	63	20,523,937
Chr8	344	25,036,674	254	21,896,636
Chr9	443	33,480,396	162	22,292,658
Chr10	344	29,290,112	157	22,260,601
Chr11	288	25,352,640	99	17,973,684
Chr12	178	24,581,549	540	36,403,605
Chr13	463	35,153,697	498	35,705,173
Chr14	NA	NA	663	34462318
Total no. of contigs	5,536		4,194	
Total length of contigs (Mb)	426		369	
Total no. of anchored contigs	4,740		3,814	
Total length of chromosome-level assembly (Mb)	403		359	
Anchor rate (%)	94.43		97.17	
No. of protein-coding genes	29,416		27,211	
No. of known miRNAs	125		130	

because they are defined by a unique enclosed inflorescence, the syconium or, commonly, fig. Because of the enclosed urn-shaped inflorescence, fig plants rely on specific insect pollinators and, in turn, provide nourishment and shelter for pollinators to reproduce. Morphological matching between figs and their pollinators is required for successful pollination and oviposition. The body shapes and sizes of pollinators correlate with the size of fig inflorescence (Verkerke, 1989), and specific volatile organic compounds (VOCs) attract these obligate pollinators (Chen et al., 2016; Grison-Pigé et al., 2002; Hossaert-McKey et al., 2010). This obligate mutualism is a fascinating case of extreme plant-insect codiversification and poses questions, including how the mutualism originated and evolved and what genes have contributed to coevolution.

The genus *Ficus* can be classified into six subgenera, including two monoecious subgenera (*Urostigma* and *Pharmacosycea*) and four functionally dioecious subgenera (*Sycomorus*, *Ficus*, *Synoecia*, and *Sycidium*). About half of the *Ficus* species are functionally dioecious. The *RAN1* gene has been reported to be responsible for sex determination in *F. carica* (Mori et al., 2017), the only species distributed mainly in temperate zones. However, the *RAN1* gene did not show clear sex and organ

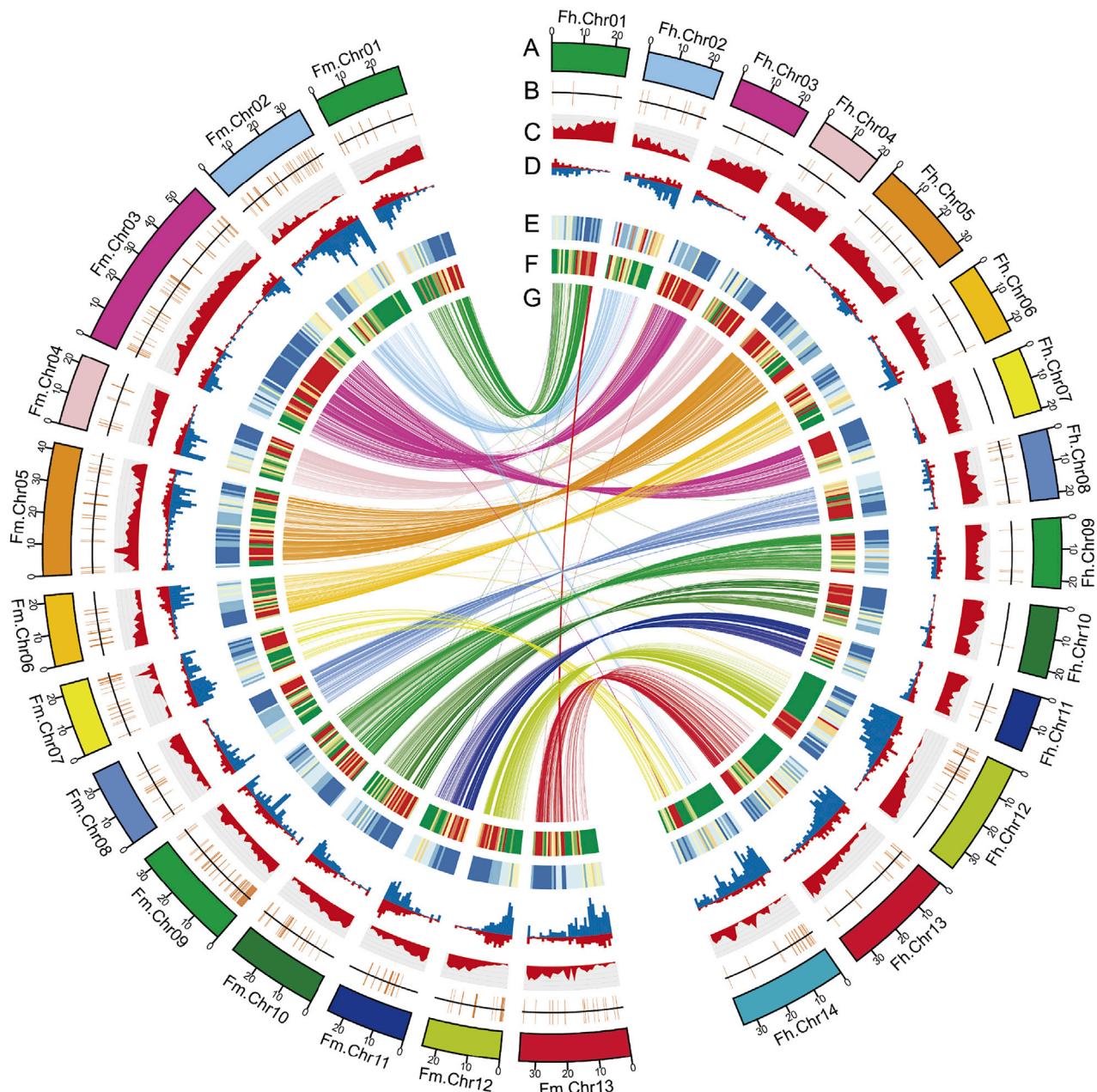
specificity in expression. *F. microcarpa* and *F. hispida* represent two different reproductive systems, monoecy and functional dioecy, and different growth forms, hemi-epiphytic and territorial. Despite the importance of the *Ficus* genus as an ideal model system for studying plant-insect interactions, genetic research of *Ficus* is far from well developed, with no high-quality reference genomes. The high-quality genomes of *F. microcarpa* and *F. hispida* will allow us to uncover the genetic mechanisms behind special features of this large genus, including aerial root development, sex determination, the origin and evolution of this genus, and the codiversification of fig species with their wasp pollinators.

RESULTS

Assembly and Annotation of the Genomes of *F. microcarpa*, *F. hispida*, and the Pollinator Wasp *Eupristina verticillata*

F. microcarpa and *F. hispida* represent two different reproductive systems and carry different numbers of chromosomes: 13 pairs in *F. microcarpa* and 14 pairs in *F. hispida* (Data S2A, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>). The estimated genome sizes are 436 Mb for *F. microcarpa* and 370 Mb for *F. hispida*, based on flow cytometry, which is consistent with results from K-mer counting (Data S2B, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>). A total of 36.87 and 36.12 Gb of subreads were generated on the PacBio RSII platform for sequencing the genomes of *F. microcarpa* and *F. hispida*, respectively (Data S1A, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>). These two genomes were assembled *de novo* by incorporating sequence data from PacBio long reads and physical mapping information from high-throughput chromatin conformation capture (Hi-C; Data S1A and S1B, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>). The analyses yielded an assembly of 426 Mb with a contig N50 value of 908 kb for *F. microcarpa*, covering 97.7% of the estimated 436-Mb genome, and an assembly of 360 Mb with a contig N50 of 492 kb for *F. hispida*, covering 97.3% of the estimated 370-Mb genome (Table S1). The completeness of the assemblies was evaluated using 1,375 conserved embryophyta proteins collected from the Benchmarking Universal Single-Copy Orthologs (BUSCO) program (Simão et al., 2015), showing 95.6% and 97.4% completeness for *F. microcarpa* and *F. hispida*, respectively (Table S2). The assemblies were further assessed using assembled RNA sequencing (RNA-seq) transcripts and Illumina short reads (Table S3; Data S1C, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>). Finally, 13 pseudo-chromosomes comprising 423 Mb in *F. microcarpa* and 14 chromosomes comprising 359 Mb in *F. hispida* were anchored using ALLHiC (Zhang et al., 2019; Table 1; Figure 1; Figures S1A and S1B). Comparative analyses showed that the *F. hispida* Hi-C-based assembly and the high-density genetic map generated from a re-sequenced segregating population were highly consistent (Figures S1C and S1D).

Genome annotation yielded 29,416 and 27,211 protein-coding genes with 94.4% and 95.7% BUSCO completeness in *F. microcarpa* and *F. hispida*, respectively (Table 1; Table S4). Comparison of these two *Ficus* genomes showed that 17,052

**Figure 1. Genome Features of Fm and Fh**

- (A) Chromosome karyotypes.
- (B) SDs.
- (C) Gene density.
- (D) LTR TEs.
- (E) DNA TEs.
- (F) Gene expression analyzed using RNA-seq.
- (G) Synteny between the two genomes.

gene pairs matched in conserved syntenic blocks and could thus be considered orthologous pairs, whereas 13,200 genes (23.31%) appeared to have been rearranged or originated from tandem or dispersed duplications (Table S5). We also annotated

125 and 130 known microRNAs (miRNAs) encoded by the *F. microcarpa* and *F. hispida* genomes, respectively (Table 1). By integrating the orthologous genes from multiple species, we analyzed their phylogenetic relationships and estimated that

Table 2. Statistics for Assembly and Annotation of the Draft Genome of the Wasp *Eupristina verticillata*

Sequencing	<i>E. verticillata</i>
Sequencing platform	PacBio Sequel
Genome sequencing depth (x)	170
Estimated genome size (Mb)	382
Assembly	
Assembly size (Mbp)	387
% of estimated genome size	101
No. of contigs	768
N90 (bp)	702,647
N50 (bp)	3,130,893
Average length (bp)	503,882
No. of gaps	0
BUSCO completeness	97.7
Annotation	
No. of protein-coding genes	14,312
Average gene length (bp)	6,612
Average coding sequence length (bp)	274
Average exon number per gene	4.62
BUSCO completeness (%)	96.7

these two *Ficus* species diverged ~40 million years ago (mya) (Figure S2A).

We sequenced the genome of the wasp *E. verticillata*, the pollinator of *F. microcarpa*, and estimated its genome size to be 382 Mb with heterozygosity of 1.1% based on K-mer counting. We generated a total of ~65 Gb of PacBio long reads with ~170× sequencing coverage (Table 2). Initial assembly using CANU yielded 439 Mb of assembled sequence and accounted for 113% of the estimated genome. Redundant sequences were identified and removed using a whole-genome alignment strategy implemented using the Redundans program (Pryszzcz and Gabaldón, 2016), resulting in an assembled genome of 387 Mb and a contig N50 of 3.1 Mb at 97.7% BUSCO completeness (1.4%; Table 2; Table S2). A combination of *ab initio* gene prediction, homologous protein evidence, and RNA-seq data led to annotation of 14,312 protein-coding genes.

Comparative analyses of the *E. verticillata* genome and the published genome of *Ceratosolen solmsi*, the pollinator of *F. hispida* (Xiao et al., 2013), revealed that these two wasp species diverged ~41.5 mya (Figure S2C), similar to the estimate for the divergence of *F. microcarpa* and *F. hispida* of ~40 mya. We found that the genomes of *E. verticillata* and *C. solmsi* contained 3,934 and 1,080 species-specific genes, respectively (Figure S2B). Analysis of the *E. verticillata* species-specific genes revealed that these genes were enriched in various Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (for instance, neuroactive ligand-receptor interaction and the Hedgehog signaling pathway; Data S2C, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>), suggesting their contribution to species divergence. Synteny analysis revealed a high degree of divergence of these genomes (Data S2D, Mendeley Data:

<https://doi.org/10.17632/34zz94gkch.1>). However, these two species were likely undergoing convergent evolution because a substantially reduced number of chemosensory system genes were detected in both genomes compared with the genomes of other insects (Data S2E–S2I, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>). We also observed several shared events of gene losses, such as clade 3 in the chemosensory protein (CSP) family and clade 2 in the odorant-binding protein (OBP) family, possibly attributable to ancestral divergence from other sister lineages.

Structural Variations and Segmental Duplications in *Ficus* Genomes Are Associated with Adaptive Evolution

Alignment of the genomes of *F. microcarpa* and *F. hispida* revealed many structural variations (SVs). A chromosome fusion or fission event involving *F. microcarpa* chromosome 03 (FmChr03) and its homologs, *F. hispida* chromosomes 03 (FhChr03) and 07 (FhChr07), occurred and was followed by two inversions in each of the two chromosome fragments (Figure 1; Figure S2D). Homologous regions in FmChr03 covered 3.49 Mb with 200 overlapping genes, whereas homologous regions extended 1.34 Mb in FhChr03 and 1.12 Mb in FhChr07, involving a total of 300 genes (Data S1D, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>).

Comparisons of these two *Ficus* genomes with the *Vitis vinifera* genome revealed that most syntenic blocks were unique in *F. microcarpa* (Fm)-*Vitis vinifera* (Vv) or *F. hispida* (Fh)-Vv syntenic pairs, and no whole-genome duplication (WGD) was detected after the triplication event shared by eudicots (Figure S2E). Whole-genome alignment analysis showed that 116.5 Mb (27.5% of the assembled genome) covering 7,636 genes in Fm and 79.4 Mb (22.1%) containing 4,920 genes in Fh were segmental duplications (SDs), with ~38 Mb of SDs shared between the two *Ficus* genomes, indicating that SD is a major contributor to genome size expansion (Data S1E, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>). We annotated transposable elements (TEs) and found that they comprise 46.5% in Fm and 48.9% in Fh (Table S6). An opposite pattern between the distribution of class II TEs and gene-rich regions was observed in the two *Ficus* genomes (Figure 1), with a Spearman's rank correlation coefficient of -0.78 in Fm and -0.66 in Fh. Kimura distance analysis (Kimura, 1980) was used to infer the evolutionary distance of the TEs based on calculation of the pairwise divergence between TE copies and the consensus sequences, showing that a very recent *Ty1/copia* burst event occurred in Fm but not in Fh (Figures S3A and S3B). Further, the long terminal repeats (LTRs) located in SD regions exhibit a similar expansion pattern in both genomes (Figures S3C and S3D).

Gene Ontology (GO) functional enrichment analysis revealed that genes within the rearranged regions of FmChr03 and FhChr03 were enriched in functions mainly related to plant immunity, whereas in the rearranged region of FhChr07, genes involved in plant nutrition were strongly overrepresented (Data S1F, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>). Genes overlapping within the Fm SD regions were significantly enriched in monoterpene/terpene-related biological processes (Data S1G, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>).

34zz94gkch.1; Fisher's exact test, false discovery rate [FDR] < 0.01).

Expansion of Auxin-Related Genes Associated with Aerial Root Development in Fm

The most noticeable feature of banyan *Ficus* species is their extraordinary aerial roots. In contrast, some *Ficus* species (e.g., Fh) do not have aerial roots. To investigate the pathways triggering aerial root formation in Fm, we identified 811 genes that were highly expressed in aerial root tips (Figure S4). These highly expressed genes were significantly enriched for a series of transport-related biological processes (Figure S4D; Fisher's exact test, FDR < 0.05). Notably, *PIN1*, which encodes the auxin efflux carrier (Wisniewska et al., 2006), and the *GNOM* gene, which plays important roles in endosomal recycling of PIN1 proteins to the basal plasma membrane (Richter et al., 2010), were highly expressed in aerial roots (Figure 2A), suggesting highly efficient acropetal auxin transport. Transcripts of the TAR and YUC genes, which are involved in auxin biosynthesis, and those of several transcription factors (TFs), including *IAA14*, *ARF7/19*, *PLETHORA2* (*PLT2*), and *WOX11*, were also more abundant in aerial root tissues (Figure 2A). Substantial tandem duplications and SDs that have occurred in chromosomes 4, 8, and 10 involved auxin-related genes with 21 paralogs in Fm and four in Fh (Figure 2B; Data S2J and S2K, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>). A total of 36 genes that were highly expressed in aerial roots belong to eight gene families, of which 8 (22.2%) were located in SD regions and 11 (30.6%) were tandemly duplicated, exceeding 50% combined. Phylogenetic analysis of the PIN family identified substantial copy number variations in *PIN1* genes (four paralogs in Fm, two copies in Fh, and one in *A. thaliana*) and identified two evolutionary *PIN1* subgroups (Figure 2C). The first group (*PIN1a*), identified in Fm, Fh, and *A. thaliana*, included singular *PIN1* genes that likely represent the ancestral state. The second group (*PIN1b*), found specifically in the two *Ficus* species, included three duplicated copies in Fm and one copy in Fh. RNA-seq analysis showed that *FmPIN1a* was expressed in most of the tissue samples tested; however, *FmPIN1b1*, *FmPIN1b2*, and *FmPIN1b3* were highly expressed in aerial root samples. Expansion of the TAR and YUC gene families has provided fundamental genetic materials for enhanced indole-3-acetic acid (IAA) biosynthesis. The increased copy numbers of the light receptors CRY2 and PHR2 and increased expression of TAR-encoding (e.g., *TAT2a*, *TAT2b*, and *TAT7b*) and YUC-encoding genes (e.g., *YUC6*) further accelerate accumulation of IAA (Figures 2A and 2B).

We quantified endogenous auxin in Fm and Fh, and detected a substantial amount of IAA in Fm aerial roots (15.65 ng/g) compared with leaves of Fm (3.11 ng/g) and Fh (3.38 ng/g), a more than 5-fold increase in aerial roots (Figure 2D), suggesting that enhanced auxin production triggered aerial root development. We propose that an auxin-dependent pathway promoted by light is associated with aerial root initiation, growth, and pattern formation (Figure 2E).

Sex Chromosome Evolution and Sex Determination in Fh

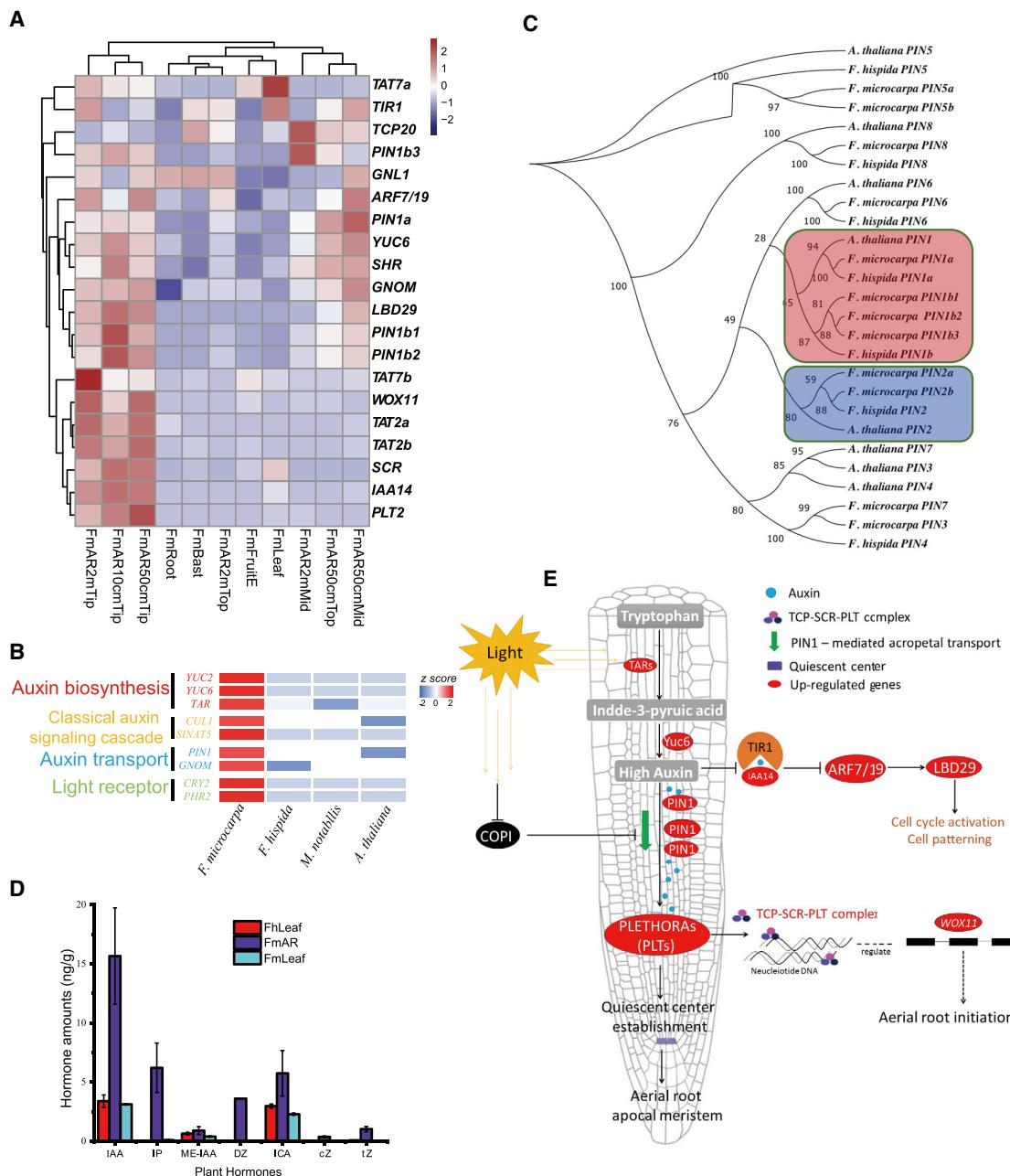
An ultra-high-density F₁ paternal genetic map was constructed based on resequencing of 60 F₁ individuals, 30 each for male

and female (sex ratio 1:1), and their maternal parent. A total of 161,658 high-quality SNP markers were selected as paternal markers with the criteria that the paternal genotype is heterozygous and the maternal genotype is homozygous. These SNPs were merged into 2,725 cosegregating bins. Fourteen linkage groups were constructed, spanning 1,884.99 cM (Figure S1). A non-recombinant region in the first ~2-Mb region in Chr12 was identified as the sex-determining region (SDR) (Figure 3A). However, recombination events were detected in the corresponding region of the X chromosome at 3.4 centiMorgan (cM) at a rate of 590 Kb/cM (Data S1H, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>), which is lower than 350 Kb/cM rate in the pseudo-autosomal region of the Y chromosome or the 330 Kb/cM rate in the rest of the X chromosome, likely because of higher content of repetitive sequences.

We further compared the degrees of heterozygosity of male and female resequencing samples across 14 chromosomes and observed a significant association (t test, p < 0.001) between a 2-Mb genomic region in Chr12 with the SDR, in which male samples show much higher heterozygosity than females (Figure 3B). This region is highly repetitive, containing 93.64% transposons and simple repeats (Table S7), which is a feature of the non-recombinant region in the recently evolved Y chromosome in Fh, as observed in papaya (Wang et al., 2012). In addition to repetitive sequences, a MADS box TF (*Fh.12G0000020*) was annotated in the SDR with high confidence (Figure S5A). Further, we assembled X and Y chromosomes separately *de novo*, based on our newly developed chromosome phasing approach (see STAR Methods and Figure S5B for details). The phased sex chromosomes contained 21.9 Mb and 22.6 Mb in the X and Y chromosomes, respectively, revealing that the Y chromosome had expanded 700 kb (Figure 3C). Comparisons between X and Y chromosomes revealed substantial sequence divergence and allowed us to identify an inversion located between 0.61 and 1.57 Mb on the Y chromosome in Fh that suppresses recombination (Figure 3C).

Comparative analysis of phased X and Y chromosomes revealed a presence or absence variant (PAV) comprising a 42-kb male-specific region that is present only in male individuals. This 42-kb region includes only one protein-coding gene (*Fh.12G0000020*), which is the same gene located in the SDR. RNA-seq analysis revealed substantial expression of *Fh.12G0000020* in male inflorescences at three different stages: stage A, pre-receptive phase, before female flowers mature; stage B, receptive phase, mature female flowers; and stage C, inter-flower phase, during development of pollinated ovules. However, there was no expression of this gene in female inflorescences, strongly suggesting that this gene is a dominant factor affecting sex determination (Figure 3E). Alignment of resequenced genomes to the Fh reference genome showed that the *Fh.12G0000020* gene was present only in male genomes and absent in female genomes (Figure 3D; Figure S5D).

Phylogenetic analysis revealed that the *Fh.12G0000020* gene is a homolog of the AGAMOUS (AG) gene in *Arabidopsis* (Figure S5C), which is a C-class MADS box gene that plays a central role in reproductive organ (stamen and carpel) development (Ito et al., 2007). Three paralogs of AG (*Fh.01G0006690/FhAG1*, *Fh.12G0000020/FhAG2*, and *Fh.v50013370/FhAG3*) were

**Figure 2. Genomic Basis of Aerial Root Development in Fm**

(A) Highly expressed genes in Fm aerial roots compared with other tissues. Fm, *F. microcarpa*; AR, aerial root. 10 cm, 50 cm, and 2 m are the lengths of ARs from top to bottom. Top, Mid (i.e., middle), and Tip represent different positions of ARs. Other tissues include root, bast, fruit at stage E, and leaf. These genes have been implicated in lateral root development (Aida et al., 2004; Richter et al., 2010; Feng et al., 2012; Lavenus et al., 2013; Shimotohno et al., 2018).

(B) Genomic expansion of key genes involved in auxin biosynthesis, signaling, transport, and light reception in Fm compared with other plant species, including Fh, *M. notabilis*, and *A. thaliana*.

(C) Phylogenetic tree of the PIN gene family in Fm, Fh, and *A. thaliana*, illustrating expansion of the *PIN1* and *PIN2* genes in Fm.

(D) Quantitation of plant hormone amounts in Fh leaves (FhLeaf), Fm ARs (FmARs), and Fm leaves (FmLeaf). Plant hormones analyzed include indole-3-acetic acid (IAA), N6-isopentenyladenine (IP), methyl indole-3-acetate (ME-IAA), dihydrozeatin (DZ), indole-3-carboxaldehyde (ICA), cis-zeatin (cZ), and trans-zeatin (tZ).

(E) A proposed model of an auxin-dependent pathway controlling AR patterning, initiation, and formation.

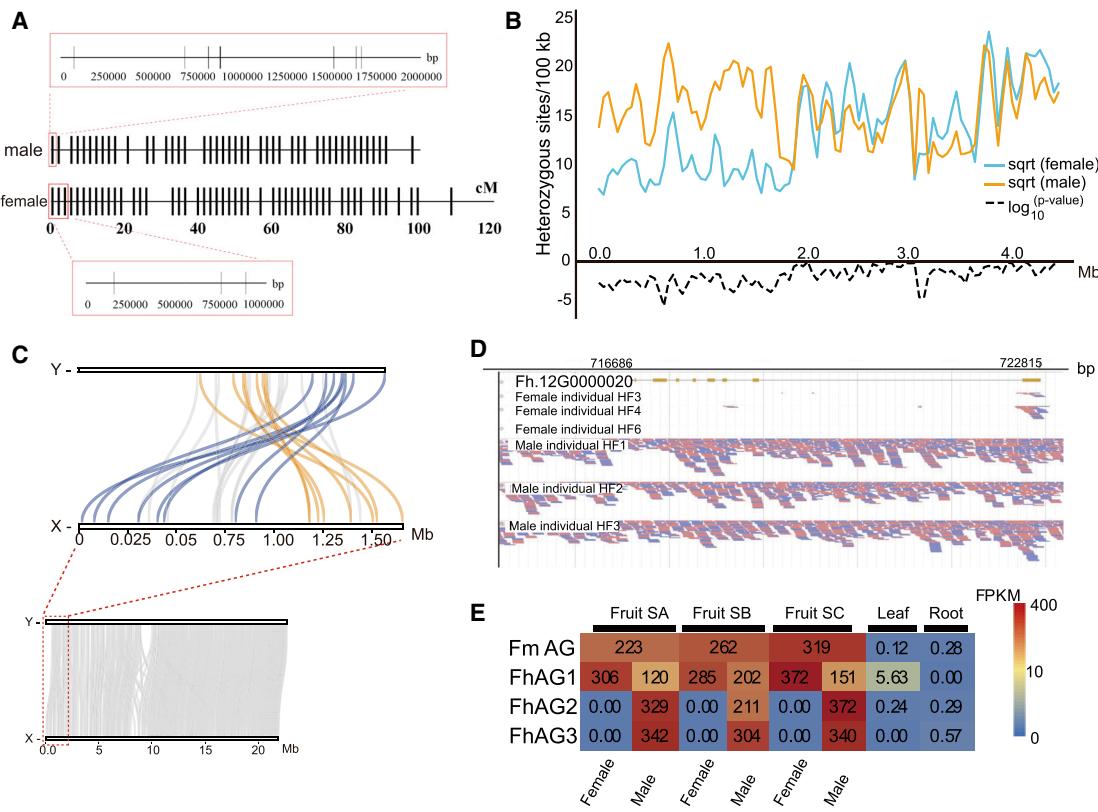


Figure 3. Identification of the Y Chromosome and Sex Determination Candidate Gene in Fh

(A) Comparison of male and female genetic maps in Fh. Male and female genetic maps (LG10), derived from resequencing 30 male and 30 female F₁ individuals, are presented in the center panel. Bin markers located in the sex-determining region (SDR) are highlighted, with their corresponding SNP markers plotted along chromosome positions in red boxes.

(B) Distribution of heterozygous sites along Fh Chr12 regions from 0–5 Mb in male and female plants. Significance was tested by Student's t-test. P < 0.05 was considered as statistical significance.

(C) Comparison of the X and Y chromosomes identified an inversion, illustrated by yellow and blue lines, located between 0.61 and 1.57 Mb on the Y chromosome. The top panel shows a local comparison between X and Y chromosomes within the SDR, and the bottom panel shows a global comparison of the phased sex chromosomes. Background collinearity is indicated by gray lines.

(D) Alignments of sequence reads from female and male individuals to the reference genome revealed a presence or absence variant (PAV) in gene Fh.12G0000020 (FhAG2).

(E) Expression of FmAG, FhAG1, FhAG2, and FhAG3 in samples from various organs and tissues, including stage A (pre-receptive stage, before female flower maturation), stage B (receptive stage, female flower maturation), and stage C (inter-flower stage, pollinated ovule development), leaves, and roots.

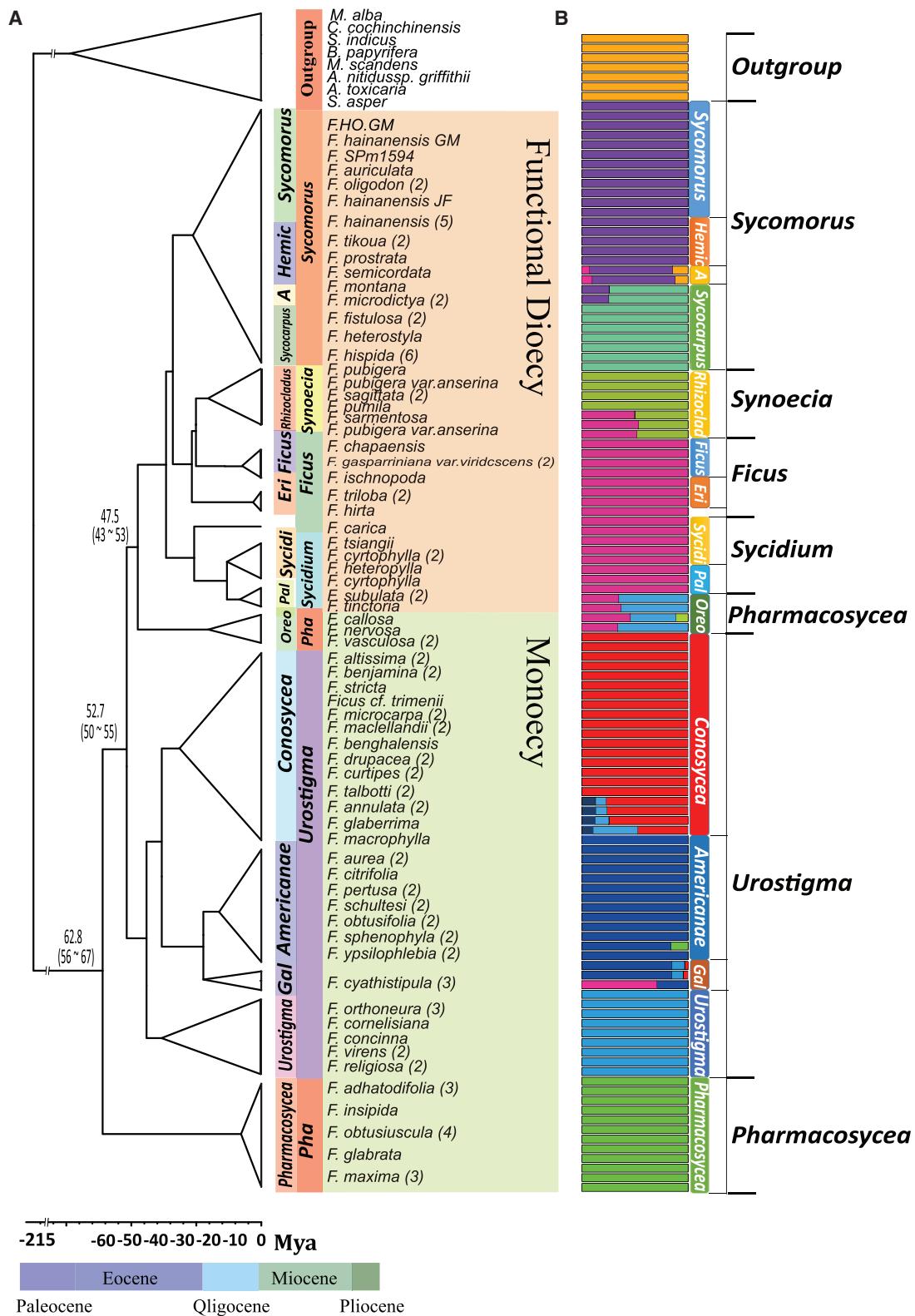
identified in the Fh genome on Chr01, Chr12, and an unanchored contig, respectively; however, only one copy (Fm.01G0007740 or FmAG) was detected in the genome of monoecious Fm (Figure 3E; Figure S5C). These four AG homologs are highly similar in coding sequences, protein sequences, and gene structure (Data S2L, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>). We found that FmAG and FhAG1 showed very similar expression patterns in flowers, leaves, and roots (Figure 3E) and maintained strong collinearity between the homologous genomic regions (Figure S5E), indicating that FhAG1 was likely the ancestral copy and the orthlog of AG in Fh. The expression of paralogs FhAG2 and FhAG3 transcripts was strongly elevated in male inflorescences (Figure 3E) and was present only in male plants (Figure S5D), suggesting that they were functionally redundant and likely duplicated after the divergence of monoecy and dioecy in *Ficus*. The AG paralog FhAG2 was

located in the SDR and originated from gene duplication followed by neofunctionalization. It is a candidate gene for sex determination in Fh. FhAG3 is unlinked, and if it is also located in the SDR, then it could be a candidate gene.

PCR analyses of a male-specific marker derived from the FhAG2 gene sequence were positive in male plants but negative in female plants in a collection of dioecious fig species from other subgenera, including *F. pumila* from *Synoecia* and *F. ischnopoda* and *F. langkokensis* from *Ficus* (Figure S5F).

Origin and Evolution of the Genus *Ficus*

A phylogeny of *Ficus* was established using data from resequenced genomes of 112 *Ficus* accessions comprising 62 *Ficus* species that cover two monoecious subgenera (*Pharmacosycea* and *Urostigma*) and four functionally dioecious subgenera (*Sycomorus*, *Synoecia*, and *Ficus*; Data S1I,



(legend on next page)

Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>), which represent all six subgenera and main clades of *Ficus*. A total of 12.87 million SNPs were retained for further analysis, with 0.42 million SNPs (3.2%) located in single-copy genes.

The maximum likelihood (ML) tree of *Ficus* using the dataset supported the phylogenetic topology of the commonly accepted *Ficus* phylogeny (Figure 4A). The monophyly of most subgenera and sections was supported in this ML tree, which clearly distinguished the six subgenera in *Ficus* (Figure 4A), with the subgenera *Ficus* and *Pharmacosycea* as exceptions. The subgenus *Pharmacosycea* was divided into two groups, with the section *Pharmacosycea* standing as the sister lineage to the remaining *Ficus* species, whereas section *Oreosycea* was recognized as a sister of the functionally dioecious *Ficus* clades. Admixture and genetic relatedness analysis revealed introgression events in section *Oreosycea* that might explain the paraphyly of the subgenus *Pharmacosycea*, which appeared on two branches at different phylogenetic positions (Figure 4; Figure S6). Further, ABBA-BABA analysis supported our hypothesis of introgression because extremely high absolute Z scores were observed between section *Oreosycea* and other sections containing the subgenera *Urostigma*, *Pharmacosycea*, *Americana*, *Palaeomorphe*, and *Sycidium* (Data S2N, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>). In addition, potential extensive interspecific hybridizations across the whole genus were detected during admixture and genetic relatedness analyses (Figure 4B, Figure S6; Data S2M, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>) and were also supported by the occurrence of cyto-nuclear conflicts in the phylogeny (Data S2O).

Ficus species are monoecious or functionally dioecious, depending on the arrangement of unisexual florets within the syconium (Verkerke, 1989; Weiblen, 2000). Our phylogenetic analysis revealed that monoecy represents the ancestral reproductive system across the genus (Figure 4). Molecular dating analysis showed that the divergence between monoecy and functional dioecy occurred in the Eocene epoch (approximately 47.5 mya; Figure 4A), a period coinciding with a high carbon dioxide level along with global warming (Gutjahr et al., 2017).

Genomic Signatures of Codiversification between Fig Species in the Subgenus *Sycomorus* and Their Pollinator Wasps

To investigate the potential molecular mechanism of coevolution between figs and wasps, we identified 2.28 million high-quality SNPs using 14 resequenced species from the subgenus *Sycomorus* and 355,632 SNPs from their corresponding pollinators (*Ceratosolen* spp.). Phylogenetic analysis demonstrated strong conservation of host associations at this taxonomic level, although we

detected several host-shifting events (Figure 5B; Data S1J, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>).

To identify candidate genes that might have undergone reciprocal selection and divergence in these coevolving species, selective sweeps were identified using Tajima's D and SweeD analyses in the fig and wasp genomes. A total of 7.9 Mb of genomic sequences covering 937 protein-coding genes in the Fh genome exhibited negative values of Tajima's D that deviated greatly from zero and overlapped with SweeD results, suggesting that purifying selection drove the evolution of these regions. These selective sweeps were distributed randomly among the 14 chromosomes of Fh, with some chromosomes showing a higher density (Data S2P, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>). In the wasp genome, 6.21 Mb of genomic regions covering 380 protein-coding genes were under strong purifying selection, based on selective sweeps detected by Tajima's D and SweeD analyses.

The extremely strict mutualism system leads to coadaptation and codivergence between figs and wasps, likely resulting in corresponding phenotypic traits, including wasp body size and fig fruit size. To test this hypothesis, we identified a total of 37 genes in the Fh genome, based on homologies to genes in tomato that affect fruit development (Azzi et al., 2015), and found that 18.9% (7 of 37) of the tested genes were under purifying selection, including three auxin signaling-related genes (ARF7, ARF8, and TIR1), two abscisic acid (ABA) biosynthesis-related genes (NCED and FLACCA), one gene involved in primary metabolism (SuSy), and one gene involved in cell cycle control (CDKA) (Figure 5A). Correspondingly, we observed that 14 of 62 (22.5%) wasp genes in the Hippo signaling pathway, which controls organ size in animals through regulation of cell proliferation and apoptosis (Saucedo and Edgar, 2007), were located in genomic regions of selective sweeps (Figure 5A). Global GO and KEGG analyses of the wasp genes under purifying selection also revealed several genes involved in organ development and body size (Data S2Q and S2R, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>).

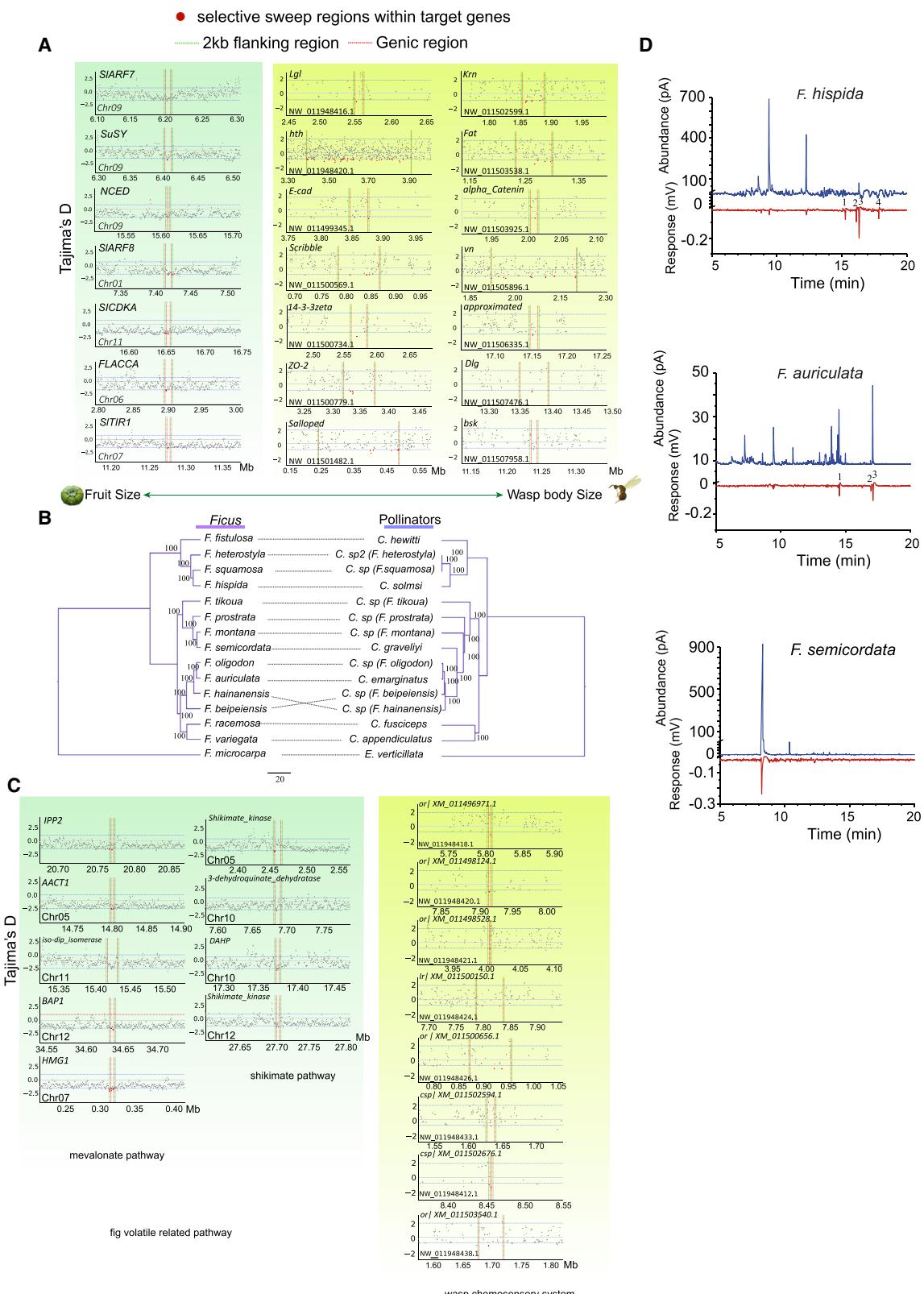
Specific VOCs Attracting Wasp Pollinators in Subgenus *Sycomorus* Figs

A critical hypothesis for the obligate mutualism between figs and wasps was proposed that their interactions were tightly controlled by specific VOC constituents of the floral scents emitted by fig inflorescences (Chen et al., 2016; Grison-Pigé et al., 2002; Hossaert-McKey et al., 2010). We investigated the species-specific signaling compounds that were essential for communication in three fig-wasp pairs—Fh-C. solmsi, *F. auriculata*-C. emarginatus, and *F. semicordata*-C. gravelyi—representing three main fig sections of the subgenus *Sycomorus*

Figure 4. Evolutionary History of the Genus *Ficus*

(A) Phylogenetic tree and molecular dating of sequenced *Ficus* species. Fifteen sections are labeled on the right (Gal, Galoglychia; Oreo, Oreosycea; Sycidium; Eri, Eriosycea; A, Adenosperma; Hemic, Hemicardia), followed by the six subgenera (Pha, Pharmacosycea). Monoecy and functional dioecy are indicated in the rightmost panel. The estimated divergence time with 95% confidence intervals for three important nodes are labeled in the phylogenetic tree, including 62.8 mya for the crown of the *Ficus* genus, 52.7 mya for the diversification from section *Pharmacosycea*, and 47.5 mya for origination of functional dioecious *Ficus* species.

(B) Admixture analysis showing the distribution of K = 9 genetic clusters with the smallest cross-validation error.
See also Figure S6.



(legend on next page)

and their pollinator wasps. The active signaling compounds specific to stage B were the unknown sesquiterpenones FHSTN1, FHSTN2, and FHSTN3 and benzyl benzoate in Fh; the unknown sesquiterpenones FASTN1, FASTN2, and FASTN3 in *F. auriculata*; and 4-methylanisole in *F. semicordata* (Figure 5D; Figure S7). All of these compounds can be categorized into two groups: isoprene-derived compounds synthesized in the mevalonate pathway (all unknown sesquiterpenones) and benzenoid-derived components synthesized in the shikimate pathway (4-methylanisole). Our electrophysiological study supports this hypothesis.

Population genomic analysis of resequenced genomes in the subgenus *Sycomorus* validated evolutionary constraint of genes involved in the mevalonate and shikimate pathways. A total of nine genes under strong negative selection were enriched for functions in terpenoid metabolite-related pathways, including flavonoid biosynthesis, terpenoid backbone biosynthesis, diterpenoid biosynthesis, and ubiquinone and other terpenoid-quinone biosynthesis (Data S1K, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>). Another nine genes that have undergone strong purifying selection were enriched for functions in benzenoid-related pathways, including phenylalanine metabolism; phenylpropanoid biosynthesis; and stilbenoid, diarylheptanoid, and gingerol biosynthesis (Data S1K, Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>). We further tested genes encoding 10 important enzymes involved in the mevalonate pathway (Caspi et al., 2016) and seven enzymes involved in the shikimate pathway and detected selective sweeps in five genes in the mevalonate pathway (*IPP2*, *AACT1*, *BAP1*, *HMG1*, and *iso-dip isomerase*) and four in the shikimate pathway (*DAHP*, *3-dehydroquinate_dehydratase*, and two shikimate kinases) (Figure 5C). These results strongly suggest that selective pressure in genomic regions containing genes in the mevalonate and shikimate pathways has contributed to production of specific VOCs and species diversification in subgenus *Sycomorus* figs.

77 genes that mediated detection of chemical cues were identified in the *C. solmsi* wasp genome, the pollinator of Fh. The proteins encoded by these genes allow wasps to process olfactory (Or) and gustatory (Gr) cues, including OBPs, CSPs, and ionotropic receptors (Xiao et al., 2013). Among these genes, five encoding ORs, one encoding CSPs, and one ionotropic receptor (IR) exhibited selective sweep signals (Figure 5C). Selective pressure in genomic regions with genes for recognition of chemical signals contributed to the species divergence of the pollinator wasp of *Sycomorus* figs. Selective

signals in genes affecting organ size and specific VOC production in figs and chemosensory receptors in obligate pollinator wasps support a critical role of morphological and physiological coadaptation between figs and pollinator wasps in their codiversification.

DISCUSSION

The ratio of auxin to cytokinin in plant meristematic tissues affects shoot or root development, and higher ratios of auxin to cytokinin lead to root development. Duplication of genes controlling auxin biosynthesis, signaling, and auxin transport in Fm may have led to increased auxin production, as indicated by 5-fold higher concentrations of endogenous auxin in Fm aerial root tips compared with leaves of Fh, which does not have aerial roots. Further, duplication of the *CRY2* and *PHR2* genes, which encode light receptors, has occurred in Fm but not in Fh, leading to our conclusion that an auxin-dependent pathway promoted by light is associated with aerial root initiation and development (Figure 2E). A high degree of IAA accumulation might initially have resulted from expansion of the TAR and YUC gene families (Mashiguchi et al., 2011). Further, exposure to light promotes expression of *TAT2a*, *TAT2b*, *TAT7b*, and *YUC6* in aerial root tips of Fm (Halliday et al., 2009) and accelerates auxin accumulation. High auxin concentrations upregulate expression of *ARF7/19* via suppression of *IAA14* activity (Lavenus et al., 2013) and further promote expression of *LBD29*, which activates the cell cycle and cell patterning (Feng et al., 2012). In addition, PIN1-mediated auxin acropetal transport regulates expression of the *PLETHORA* gene (*PLT2*), an AP2-class TF (Aida et al., 2004). The PLT protein forms a complex with the TCP (plant-specific teosinte-branched cycloidea PCNA) and SCR (SHORT-ROOT) proteins and further upregulates the WOX gene (*WOX11*) (Aida et al., 2004; Breuninger et al., 2008), which might contribute to aerial root initiation. The PLT protein along with the TFs SCR and SHR/SCARECROW maintains quiescent center identity (Aida et al., 2004), which might play an important role in the aerial root apical meristem.

Identification of a non-recombining sex co-segregating region in Chr12 indicated that a pair of nascent sex chromosomes has evolved in dioecious Fh. An inversion was detected in this non-recombining SDR, revealing the mechanism of recombination suppression. The 2-Mb SDR contains 93.64% repetitive sequences, unusually high even for Y chromosomes, which appeared to be in an rapid expansion phase after recombination suppression. The flower development C-class gene *AG* was

Figure 5. Coevolution and Genes that Have Undergone Selective Sweeps Related to Morphological and Physiological Codivergence between *Sycomorus* Species and Their Pollinators

- (A) Tajima's D analysis of fruit growth-related genes in Fh and Hippo signaling pathway-related genes in wasps that have undergone selective sweeps. Each dot represents a sliding window with a 1-kb window size and 500-bp step size. Gray dots are genomic background, and red ones are windows under selective sweeps. Genic regions and 2-kb flanking regions for target genes are highlighted with red and green dashed lines, respectively.
- (B) Phylogenetic analysis of fig plants from section *Sycomorus* and their pollinators. Bootstrap values (>90) were retained in the ML tree.
- (C) Tajima's D analysis of fig VOC-related genes and wasp chemosensory system-related genes that have undergone selective sweeps.
- (D) Electrophysiologically active compounds of three *Ficus* species and responses of their obligate pollinators. Simultaneous recordings of gas chromatography (flame ionization detector [FID]) are shown as blue lines, and electroantennographic detector (EAD) signals released from antennae of female *Ceratosolen* wasps are shown as red lines. Compounds from *Ficus* resulting in the strongest EAD activity peaks in wasps were identified as follows: unknown sesquiterpenone compounds and benzyl benzoate (1–4, shown in black) from Fh, unknown sesquiterpenones (1–3, shown in black) from *F. auriculata*, and 4-methylanisole from *F. semicordata*.

duplicated twice with three paralogs in dioecious Fh, whereas there was no duplication in monoecious Fm. *FhAG2* was located in the SDR and was male-specific with no X allele, a strong candidate for sex determination. This AG paralog was present in male genomes but absent in female genomes of three other dioecious species. Two linked sex determination genes have been proposed to control dioecy in figs and to initiate sex chromosome evolution in plants (Storey 1975; Charlesworth and Charlesworth 1978), which is widely accepted, although there are exceptions, such as the Y chromosome encoding small RNA as a sex determinant in persimmons (Akagi et al., 2014).

The previously identified sex-determining factor *RAN1* of *F. carica* (Mori et al., 2017) was universally expressed in both male and female with no significant difference (*t* test, $p > 0.05$) in male and female inflorescences at early stages (stages A and B) in Fh (Figure S5H). *FhRAN1* was upregulated in female inflorescences at stage C, a developmental phase coincident with seed maturation. *FhRAN1* was not controlling sex determination but likely involved in seed development in Fh.

75 million years of obligate coevolution between figs and wasps have shaped morphological and physiological adaptations in this system of coevolution (Craud et al., 2012; Herre et al., 2008; Verkerke, 1989). *Ficus* has extremely high species diversity globally and locally. The major factor contributing to reproductive isolation among sympatric closely related fig species is pollinator specificity (Grison-Pigé et al., 2002; Wang et al., 2016). An interesting question in the coevolution of figs and their wasp pollinators concerns the factors (genes/compounds) that determine coadaptation and codiversification between figs and fig wasps. Species-specific fig floral volatiles emitted during the receptive phase, usually comprised of a species-specific mixture of common terpenoid, benzenoid, and aliphatic compounds, are essential for maintaining pollinator specificity (Grison-Pigé et al., 2002; Hossaert-McKey et al., 2010). Population genomic analysis of subgenus *Sycomorus* figs and their obligate pollinator wasps and electrophysiological testing of pollinators responding to floral scents emitted from three different *Ficus* species support the important roles of the mevalonate and shikimate pathways in attracting species-specific pollinators (Figure 5) and reveal potential molecular mechanisms of codiversification in this obligate mutualism. Simultaneous selection for phenotypic coadaptation of figs and wasps, including chemical recognition of plants by pollinators, has left footprints in their genomes, and the evidence of this presented here supports previous studies showing that coevolution based on morphological and chemical phenotypic traits is critical for codiversification of fig-wasp mutualism (Wang et al., 2013, 2016; Jousselin et al., 2003; Hossaert-McKey et al., 2010; Souto-Vilarós et al., 2018).

We also observed that many SDs possibly driven by TEs contributed to genome size expansion in Fm, which was associated with several aspects of adaptive evolution. Terpenoids constitute the largest class of chemical constituents of volatiles in plants and play various important roles; for example, in plant defense and insect-plant interactions. Genes overlapping in SDs in Fm are significantly enriched for monoterpene/terpene-related pathways that synthesize fundamental materials involved in fig defense and fig-wasp interactions,

and these genes provide variations upon which evolution forces act.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead Contact
 - Materials Availability
 - Data and Code Availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Plants
 - Animals
- **METHOD DETAILS**
 - Genome sequencing and assembly
 - Genetic map and validation of genome assemblies
 - Genome annotation of *F. microcarpa* and *F. hispida*
 - Identification of segmental duplications
 - Identification of structural variations (SVs) and copy number variations (CNVs)
 - Phylogenetic tree reconstruction
 - Phasing of X and Y chromosomes in *F. hispida*
 - Identification of sex-determining regions in *F. hispida*
 - Identification of differentially expressed genes between male and female fruits
 - Population genomics
 - ABBA-BABA analysis
 - Dating the divergence time of the genus *Ficus*
 - Cophylogenetic and codivergence analysis between figs and wasps
 - Identification of active volatiles compounds for pollinator attraction
 - Identification of Hormone levels in plants
- **QUANTIFICATION AND STATISTICAL ANALYSES**

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.cell.2020.09.043>.

ACKNOWLEDGMENTS

We would like to thank Prof. Rodrigo A.S. Pereira, Prof. Wenshan Wu, Dr. Simon Segar, and Dr. Souto-Vilarós Daniel for collecting plant samples; Prof. Qingyi Yu, Dr. Ching Man Wai, and Prof. Liwu Zhang for discussions regarding sex determination; and Xiu-Lian Wen and Li-Sha Li for assistance with VOC-related experiments. We thank the Xishuangbanna National Nature Reserve, the Xishuangbanna Station for Tropical Rainforest Ecosystem Studies, and the Gardening Horticulture Department and Public Technology Service Center of XTBG, CAS for permission and assistance with field sampling and identification of active VOCs. We also thank Wuttichai Mhuantong, Wirulda Pootakham, Jeremy Shearman, and Sithichoke Tangphatsornruang for assistance with obtaining banyan tree images. This project was supported by a startup fund from Fujian Agriculture and Forestry University, a National Science Foundation (NSF) Plant Genome Research Program Award (DBI-1546890 to R.M.), National Natural Science Foundation of China (NSFC) grants 31870359 (to G.W.) and U1402264 (to J.C.), and the Youth Innovation Promotion Association of CAS (2020392 to G.W.).

AUTHOR CONTRIBUTIONS

R.M. conceived this project and coordinated research activities. R.M., J.C., G.W., and X.Z. designed the experiments. G.W. and X.Z. collected plant and wasp materials. F.D. and Xindan Xu isolated DNA and RNA and performed genome sequencing. L.Z. performed the karyotype analysis. X.Z., S.Z., S.C., X.C., Xiuming Xu, Jing Lin, R.W., and X.-Y.C. assembled, annotated, and analyzed the genomes. R.Q., Y.Y., and Z.L. performed RNA-seq analysis and analyzed gene families. X.Z., Jishan Lin, S.Z., Z.L., S.C., J.Z., and Y.W. performed aerial root-related analyses. X.Z., G.W., S.Z., and H.T. developed a novel algorithm to phase sex chromosomes, identified the sex determination region, and analyzed sex evolution. X.Z., G.W., Y.S., X.M., Y.W., Y.-L.L., and Y.-r.L. performed population genetics analyses for *Ficus* evolution, cophylogeny, and codiversification selection analysis. G.W. and P.W. conducted the electrophysiological experiments and identified active compounds. X.Z., R.M., G.W., H.T., and Y.-r.L. wrote and revised the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 17, 2019

Revised: May 4, 2020

Accepted: September 17, 2020

Published: October 8, 2020

REFERENCES

- Abrusán, G., Grundmann, N., DeMester, L., and Makalowski, W. (2009). TEclass—a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* 25, 1329–1330.
- Aida, M., Beis, D., Heidstra, R., Willemsen, V., Blilou, I., Galinha, C., Nus-saume, L., Noh, Y.-S., Amasino, R., and Scheres, B. (2004). The PLETHORA genes mediate patterning of the *Arabidopsis* root stem cell niche. *Cell* 119, 109–120.
- Akagi, T., Henry, I.M., Tao, R., and Comai, L. (2014). Plant genetics. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* 346, 646–650.
- Alexander, D.H., and Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* 12, 246.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Azzi, L., Deluche, C., Gévaudant, F., Frangne, N., Delmas, F., Hernould, M., and Chevalier, C. (2015). Fruit growth-related genes in tomato. *J. Exp. Bot.* 66, 1075–1086.
- Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., and Eichler, E.E. (2002). Recent segmental duplications in the human genome. *Science* 297, 1003–1007.
- Bao, Z., and Eddy, S.R. (2002). Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res.* 12, 1269–1276.
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580.
- Berg, C.C. (1989). Classification and distribution of *Ficus*. *Experientia* 45, 605–611.
- Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29, 2607–2618.
- Bickhart, D.M., Hou, Y., Schroeder, S.G., Alkan, C., Cardone, M.F., Matukumalli, L.K., Song, J., Schnabel, R.D., Ventura, M., Taylor, J.F., et al. (2012). Copy number variation of individual cattle genomes using next-generation sequencing. *Genome Res.* 22, 778–790.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Breuninger, H., Rikirsch, E., Hermann, M., Ueda, M., and Laux, T. (2008). Differential expression of WOX genes mediates apical-basal axis formation in the *Arabidopsis* embryo. *Dev. Cell* 14, 867–876.
- Buels, R., Yao, E., Diesh, C.M., Hayes, R.D., Munoz-Torres, M., Holt, G., Goodstein, D.M., Elsik, C.G., Lewis, S.E., Stein, L., and Holmes, I.H. (2016). JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* 17, 66.
- Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., and Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18, 188–196.
- Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C.A., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., et al. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 44 (D1), D471–D480.
- Chakraborty, M., Baldwin-Brown, J.G., Long, A.D., and Emerson, J.J. (2016). Contiguous and accurate *de novo* assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res.* 44, e147.
- Charlesworth, B., and Charlesworth, D. (1978). A model for the evolution of dioecy and gynodioecy. *Am. Nat.* 112, 975–997.
- Chen, C., Song, Q., Proffit, M., Bessière, J.-M., Li, Z., and Hossaert-McKey, M. (2009). Private channel: a single unusual compound assures specific pollinator attraction in *Ficus semicordata*. *Funct. Ecol.* 23, 941–950.
- Chen, Y., Huang, M., Wu, W., Wang, A., Bao, T., Zheng, C., Chou, L., Tzeng, H., and Tu, S. (2016). The floral scent of *Ficus pumila* var. *pumila* and its effect on the choosing behavior of pollinating wasps of *Wiebeisia pumilae*. *Acta Ecol. Sin.* 36, 321–326.
- Chin, C.S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueiroa-Balderas, R., Morales-Cruz, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* 13, 1050–1054.
- Conow, C., Fielder, D., Ovadia, Y., and Libeskind-Hadas, R. (2010). Jane: a new tool for the cophylogeny reconstruction problem. *Algorithms Mol. Biol.* 5, 16.
- Cottee-Jones, H.E.W., Bajpai, O., Chaudhary, L.B., and Whittaker, R.J. (2016). The importance of *Ficus* (Moraceae) trees for tropical forest restoration. *Biotropica* 48, 413–419.
- Cruaud, A., Rønsted, N., Chantarasuwan, B., Chou, L.S., Clement, W.L., Couloix, A., Cousins, B., Genson, G., Harrison, R.D., Hanson, P.E., et al. (2012). An extreme case of plant-insect codiversification: figs and fig-pollinating wasps. *Syst. Biol.* 61, 1029–1047.
- Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., and Aiden, E.L. (2017). *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356, 92–95.
- Dunn, D.W., Zhang, X., Wen, X., Sun, B., and Wang, R. (2019). Differential deployment of sanctioning mechanisms in male and female host trees in a gynodioecious fig-wasp mutualism. *Bull. Ecol. Soc. Am.* 100, e01515.
- Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28, 2239–2252.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Feng, Z., Sun, X., Wang, G., Liu, H., and Zhu, J. (2012). LBD29 regulates the cell cycle progression in response to auxin during lateral root formation in *Arabidopsis thaliana*. *Ann. Bot.* 110, 1–10.
- Gardner, E.M., Sarraf, P., Williams, E.W., and Zerega, N.J.C. (2017). Phylogeny and biogeography of *Macfura* (Moraceae) and the origin of an anachronistic fruit. *Mol. Phylogenetic Evol.* 117, 49–59.
- Grison-Pigé, L., Bessière, J.M., and Hossaert-McKey, M. (2002). Specific attraction of fig-pollinating wasps: role of volatile compounds released by tropical figs. *J. Chem. Ecol.* 28, 283–295.

- Gutjahr, M., Ridgwell, A., Sexton, P.F., Anagnostou, E., Pearson, P.N., Pälike, H., Norris, R.D., Thomas, E., and Foster, G.L. (2017). Very large release of mostly volcanic carbon during the Palaeocene-Eocene Thermal Maximum. *Nature* **548**, 573–577.
- Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., et al. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666.
- 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.
- Halliday, K.J., Martínez-García, J.F., and Josse, E.-M. (2009). Integration of light and auxin signaling. *Cold Spring Harb. Perspect. Biol.* **1**, a001586.
- Harrison, R.D. (2005). Figs and the diversity of tropical rainforests. *Bioscience* **55**, 1053–1064.
- Herre, E.A., Jander, 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.
- Hossaert-McKey, M., Soler, C., Schatz, B., and Proffit, M. (2010). Floral scents: their roles in nursery pollination mutualisms. *Chemoecology* **20**, 75–88.
- Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T., et al. (2009). High-throughput genotyping by whole-genome resequencing. *Genome Res.* **19**, 1068–1076.
- Ito, T., Ng, K.-H., Lim, T.-S., Yu, H., and Meyerowitz, E.M. (2007). The homeotic protein AGAMOUS controls late stamen development by regulating a jasmonate biosynthetic gene in *Arabidopsis*. *Plant Cell* **19**, 3516–3529.
- Jousselin, E., Rasplus, J.Y., and Kjellberg, F. (2003). Convergence and coevolution in a mutualism: evidence from a molecular phylogeny of *Ficus*. *Evolution* **57**, 1255–1269.
- Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120.
- Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736.
- Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59.
- Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.
- Lavenus, J., Goh, T., Roberts, I., Guyomarc'h, S., Lucas, M., De Smet, I., Fukaki, H., Beeckman, T., Bennett, M., and Laplaze, L. (2013). Lateral root development in *Arabidopsis*: fifty shades of auxin. *Trends Plant Sci.* **18**, 450–458.
- Lev-Yadun, S., Ne'eman, G., Abbo, S., and Flaishman, M.A. (2006). Comment on "Early domesticated fig in the Jordan Valley". *Science* **314**, 1683.
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100.
- Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
- Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
- Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944.
- Mashiguchi, K., Tanaka, K., Sakai, T., Sugawara, S., Kawaide, H., Natsume, M., Hanada, A., Yaeno, T., Shirasu, K., Yao, H., et al. (2011). The main auxin biosynthesis pathway in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **108**, 18512–18517.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.
- Morgulis, A., Gertz, E.M., Schäffer, A.A., and Agarwala, R. (2006). Window-Masker: window-based masker for sequenced genomes. *Bioinformatics* **22**, 134–141.
- Mori, K., Shirasawa, K., Nogata, H., Hirata, C., Tashiro, K., Habu, T., Kim, S., Himeno, S., Kuhara, S., and Ikegami, H. (2017). Identification of *RAN1* orthologue associated with sex determination through whole genome sequencing analysis in fig (*Ficus carica* L.). *Sci. Rep.* **7**, 41124.
- Nattestad, M., and Schatz, M.C. (2016). Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics* **32**, 3021–3023.
- 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.
- Patterson, M., Marschall, T., Pisanti, N., van Iersel, L., Stougie, L., Klau, G.W., and Schönhuth, A. (2015). WhatsHap: Weighted Haplotype Assembly for Future-Generation Sequencing Reads. *J. Comput. Biol.* **22**, 498–509.
- Pavlidis, P., Živkovic, D., Stamatakis, A., and Alachiotis, N. (2013). SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234.
- Perteal, M., Kim, D., Perteal, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667.
- Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). *De novo* identification of repeat families in large genomes. *Bioinformatics* **21** (Suppl 1), i351–i358.
- Pryszzcz, L.P., and Gabaldón, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* **44**, e113.
- Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A.J.S.B. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904.
- Richter, S., Anders, N., Wolters, H., Beckmann, H., Thomann, A., Heinrich, R., Schrader, J., Singh, M.K., Geldner, N., Mayer, U., and Jürgens, G. (2010). Role of the *GNOM* gene in *Arabidopsis* apical-basal patterning—From mutant phenotype to cellular mechanism of protein action. *Eur. J. Cell Biol.* **89**, 138–144.
- Saucedo, L.J., and Edgar, B.A. (2007). Filling out the Hippo pathway. *Nat. Rev. Mol. Cell Biol.* **8**, 613–621.
- Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259.
- Shanahan, M., So, S., Compton, S.G., and Corlett, R. (2001). Fig-eating by vertebrate frugivores: a global review. *Biol. Rev. Camb. Philos. Soc.* **76**, 529–572.
- Shimotohno, A., Heidstra, R., Blilou, I., and Scheres, B. (2018). Root stem cell niche organizer specification by molecular convergence of PLETHORA and SCARECROW transcription factor modules. *Genes Dev.* **32**, 1085–1100.

- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212.
- Souto-Vilarós, D., Proffit, M., Buatois, B., Rindos, M., Sisol, M., Kuyaiva, T., Isua, B., Michalek, J., Darwell, C.T., Hossaert-McKey, M., et al. (2018). Pollination along an elevational gradient mediated both by floral scent and pollinator compatibility in the fig and fig-wasp mutualism. *J. Ecol.* 106, 2256–2273.
- Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: Join Map. *Plant J.* 3, 739–744.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7, 62.
- Storey, W.B. (1975). Figs, J. Janick and J.N. Moore, eds. (1975). *Advances in Fruit Breeding* (Purdue University Press), pp. 568–589.
- Verkerke, W. (1989). Structure and function of the fig. *Experientia* 45, 612–622.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9, e112963.
- Wang, J., Na, J.-K., Yu, Q., Gschwend, A.R., Han, J., Zeng, F., Aryal, R., Van-Buren, R., Murray, J.E., Zhang, W., et al. (2012). Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc. Natl. Acad. Sci. USA* 109, 13710–13715.
- Wang, G., Compton, S.G., and Chen, J. (2013). The mechanism of pollinator specificity between two sympatric fig varieties: a combination of olfactory signals and contact cues. *Ann. Bot.* 111, 173–181.
- Wang, G., Cannon, C.H., and Chen, J. (2016). Pollinator sharing and gene flow among closely related sympatric dioecious fig taxa. *Proc. Biol. Sci.* 283, 20152963.
- Weiblen, G.D. (2000). Phylogenetic relationships of functionally dioecious *FICUS* (Moraceae) based on ribosomal DNA sequences and morphology. *Am. J. Bot.* 87, 1342–1357.
- Wisniewska, J., Xu, J., Seifertová, D., Brewer, P.B., Ruzicka, K., Blilou, I., Rouquié, D., Benková, E., Scheres, B., and Friml, J. (2006). Polar PIN localization directs auxin flow in plants. *Science* 312, 883.
- Wu, G.A., Terol, J., Ibanez, V., López-García, A., Pérez-Román, E., Borredá, C., Domingo, C., Tadeo, F.R., Carbonell-Caballero, J., Alonso, R., et al. (2018). Genomics of the origin and evolution of *Citrus*. *Nature* 554, 311–316.
- Xiao, J.-H., Yue, Z., Jia, L.-Y., Yang, X.-H., Niu, L.-H., Wang, Z., Zhang, P., Sun, B.-F., He, S.-M., Li, Z., et al. (2013). Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. *Genome Biol.* 14, R141.
- Xie, T., Zheng, J.-F., Liu, S., Peng, C., Zhou, Y.-M., Yang, Q.-Y., and Zhang, H.-Y. (2015). *De novo* plant genome assembly based on chromatin interactions: a case study of *Arabidopsis thaliana*. *Mol. Plant* 8, 489–492.
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591.
- Yang, X., and Li, L. (2011). miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. *Bioinformatics* 27, 2614–2615.
- Zhang, X., Nie, M., Zhao, Q., Wu, Y., Wang, G., and Xia, Q. (2015). Genome-wide patterns of genetic variation among silkworms. *Mol. Genet. Genomics* 290, 1575–1587.
- Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., Zhu, F., Jones, T., Zhu, X., Bowers, J., et al. (2018). Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* 50, 1565–1573.
- Zhang, X., Zhang, S., Zhao, Q., Ming, R., and Tang, H. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat. Plants* 5, 833–845.
- Zimin, A.V., Marçais, G., Pujić, D., Roberts, M., Salzberg, S.L., and Yorke, J.A. (2013). The MaSuRCA genome assembler. *Bioinformatics* 29, 2669–2677.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
<i>Ficus microcarpa</i> genome assembly	This study	GWH: SA:PRJCA002187
<i>Ficus hispida</i> genome assembly	This study	GWH: PRJCA002187
<i>Eupristina verticillata</i> genome assembly	This study	GWH: PRJCA002187
<i>Ficus</i> re-sequencing data	This study	GWH: PRJCA002187
<i>Ficus</i> RNA-seq data	This study	GWH: PRJCA002187
<i>Ceratosolen solmsi</i> genome assembly	Xiao et al., 2013	https://www.ncbi.nlm.nih.gov/genome/23331?genome_assembly_id=48818
Data S1 and S2 for additional tables and figures supporting the manuscript	This study	Mendeley Data: https://doi.org/10.17632/34zz94gkch.1
Experimental Models: Organisms/Strains		
<i>Ficus microcarpa</i>	Guanyin Lake in Fujian Agricultural and Forestry University (Fuzhou, China). This study	N/A
<i>Ficus hispida</i>	Xishuangbanna Tropical Botanical Garden (Mengla, China). This study	N/A
<i>Eupristina verticillata</i>	Jinshan College in Fujian Agricultural and Forestry University (Fuzhou, China). This study	N/A
Oligonucleotides		
Male specific primers in <i>F. hispida</i> (Forward: GACTCATATAGAAGATGTGCGAAGA; Reverse: TCCGAGAAAGGAATATAACGCCT)	This paper	N/A
Primers for house keeping gene alpha tubulin (Forward: TTGCACCCCTGGTAGCGTAG; Reverse: TCTCCGGTCTCGAGTCACT)	This paper	N/A
Software and Algorithms		
Pilon	Walker et al., 2014	https://github.com/broadinstitute/pilon
CANU	Koren et al., 2017	https://github.com/marbl/canu
HiC-Pro	Servant et al., 2015	https://github.com/nservant/HiC-Pro
FALCON	Chin et al., 2016	http://pb-falcon.readthedocs.io/en/latest
SMARTdenovo	N/A	https://github.com/ruanjue/smardenovo
BUSCO	Simão et al., 2015	https://busco.ezlab.org/
Quickmerge	Chakraborty et al., 2016	https://github.com/mahulchak/quickmerge
3D-DNA	Dudchenko et al., 2017	https://github.com/theaidenlab/3d-dna
ALLHiC	Zhang et al., 2019	https://github.com/tangerzhang/ALLHiC
JoinMap	Stam, 1993	https://www.kyazma.nl/index.php/JoinMap/
PASA	Haas et al., 2003	https://github.com/PASApipeline/PASApipeline/
RepeatModeler	Bao and Eddy, 2002	http://www.repeatmasker.org/RepeatModeler/
RepeatMasker v4.05	N/A	http://www.repeatmasker.org
TEclass	Abrusán et al., 2009	http://www.compgen.uni-muenster.de/tools/teclass//index.hbi?
TRF v4.07	Benson, 1999	https://tandem.bu.edu/trf/trf.html

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MAKER	Cantarel et al., 2008	https://www.yandell-lab.org/software/maker.html
Trinity	Haas et al., 2013	https://github.com/trinityrnaseq/trinityrnaseq/
SNAP	Korf, 2004	https://github.com/trinityrnaseq/trinityrnaseq/
GENEMARK	Besemer et al., 2001	http://exon.gatech.edu/GeneMark/
AUGUSTUS	Stanke et al., 2006	http://bioinf.uni-greifswald.de/augustus/
HiSAT2	Kim et al., 2015	http://daehwankimlab.github.io/hisat2/
StringTie	Pertea et al., 2016	https://ccb.jhu.edu/software/stringtie/
RNAfold	Lorenz et al., 2011	https://www.tbi.univie.ac.at/RNA/RNAfold.1.html
miRDeep-P	Yang and Li, 2011	https://sourceforge.net/projects/mirdp/?stars=4
bowtie	Langmead et al., 2009	http://bowtie-bio.sourceforge.net/index.shtml
Trimmomatic	Bolger et al., 2014	http://www.usadellab.org/cms/?page=trimmomatic
DESeq2	Love et al., 2014	http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html
OmicShare tools	N/A	https://www.omicshare.com/tools/Home/Soft/pathwaysea
BWA	Li and Durbin, 2009	http://bio-bwa.sourceforge.net/
GATK pipeline	McKenna et al., 2010	https://gatk.broadinstitute.org/hc
samtools	Li et al., 2009	http://www.htslib.org/
bcftools	Li, 2011	https://samtools.github.io/bcftools/bcftools.html
IQ-Tree	Nguyen et al., 2015	http://www.iqtree.org/
MUSCLE	Edgar, 2004	http://www.drive5.com/muscle/muscle.html
RAxML	Stamatakis, 2014	https://cme.h-its.org/exelixis/web/software/raxml/
Admixture	Alexander and Lange, 2011	http://software.genetics.ucla.edu/admixture
ANGSD	Korneliussen et al., 2014	http://www.popgen.dk/angsd
LASTZ	N/A	http://www.bx.psu.edu/~rsharris/lastz/
BLAST	Altschul et al., 1990	https://blast.ncbi.nlm.nih.gov/Blast.cgi
MaSuRCA	Zimin et al., 2013	http://www.genome.umd.edu/masurca.html
MUMmer 4	Marçais et al., 2018	https://github.com/mummer4/mummer
Assemblytics	Nattestad and Schatz, 2016	http://assemblytics.com/
Jbrowse	Buels et al., 2016	https://jbrowse.org/
minimap2	Li, 2018	https://github.com/lh3/minimap2
WhatsHap	Patterson et al., 2015	https://whatshap.readthedocs.io/en/latest/
Jane 4	Conow et al., 2010	https://www.cs.hmc.edu/~hadas/jane/
VCF-kit	N/A	https://vcf-kit.readthedocs.io/en/latest/
SweeD	Pavlidis et al., 2013	https://github.com/alachins/sweed
PAML	Yang, 2007	http://abacus.gene.ucl.ac.uk/software/paml.html
tracer	Rambaut et al., 2018	https://www.beast2.org/tracer-2/
popCNV	N/A	https://github.com/tangerzhang/popCNV
estimate_genome_size.pl	N/A	https://github.com/josephryan/estimate_genome_size.pl
sexPhase	N/A	https://github.com/tangerzhang/sexPhase
ApproximateCNV	N/A	https://github.com/sc-zhang/bioscripts/blob/master/approximate_cnv.py

RESOURCE AVAILABILITY**Lead Contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Ray Ming (rayming@illinois.edu).

Materials Availability

This study did not generate new unique reagents

Data and Code Availability

The genome assemblies, gene annotation and Illumina re-sequencing short reads have been deposited to Genome Sequence Archive (GSA) and Genome Warehouse (GWH) database in BIG Data Center (<https://bigd.big.ac.cn/gsa/>) under BioProject Accession number GSA: PRJCA002187. Codes that were used to phase sex chromosomes (sexPhase) in *F. hispida* could be found in GitHub (<https://github.com/tangerzhang/sexPhase>). Copy number variations of AG genes were identified using our custom code ApproximateCNV (https://github.com/sc-zhang/bioscripts/blob/master/approximate_cnv.py). Additional figures and tables supporting the manuscript are archived in Mendeley Data: <https://doi.org/10.17632/34zz94gkch.1>.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plants

Tissues of *F. microcarpa* were collected from a mature plant located in the middle of Guanyin Lake under the maintenance of Fujian Agriculture and Forestry University (Fuzhou, China) with institutional permission. A single mature male tree of *F. hispida* maintained by Xishuangbanna Tropical Botanical Garden was selected with institutional permission for genome sequencing and assembly.

Animals

The mature female wasp pollinators (*Eupristina verticillata*) were collected from fruit samples of *F. microcarpa*.

METHOD DETAILS

Genome sequencing and assembly

Illumina short reads sequencing

DNA was extracted using the QIAGEN DNaesy Plant Mini Kit and sequenced in Illumina X10 platform with 150-bp reads length and 300–500 insert size.

Pacbio library construction and Sequencing

More than 5 µg of sheared and concentrated DNA was applied to size-selection by BluePippin system. ~20 kb SMRTbell™ libraries were prepared according to the released protocol from Pacbio Company. A total of 63 Single-Molecule Real-Time (SMRT) cells (30 cells for *F. microcarpa* and 33 cells for *F. hispida*) were sequenced on Pacbio RSII system with P6-C4 chemistry. 36.87 and 36.12 Gb subreads were generated for *F. microcarpa* and *F. hispida*, respectively (Data S1A).

Hi-C library construction and sequencing

Hi-C libraries were created from young leaves of *Ficus* plants in BiorMarker Company according to previously published method (Xie et al., 2015). Briefly, the leaves were fixed with formaldehyde, lysed, and then the cross-linked DNA digested with HindIII over-night. Sticky ends were biotinylated and proximity-ligated to form chimeric junctions that were enriched for, and then physically sheared to a size of 500–700 bp. Chimeric fragments representing the original cross-linked long-distance physical interactions were then processed into paired-end sequencing libraries. A total of 147 and 201 million of 150-bp paired-end reads were produced on Illumina HiSeq X ten system. Hi-C sequencing was assessed using HiC-Pro program (Servant et al., 2015) and the results showed a decent proportion of validate reads (no less than 80% in the two *Ficus* Hi-C libraries), suggesting high quality of the Hi-C data (Data S1B).

Estimation of *Ficus* genome size

To estimate the genome size of the two *Ficus* plants, Illumina short reads were recruited to determine the distribution of K-mer values using published PERL scripts (estimate_genome_size.pl). Genome size could be determined by the total number of K-mers divided by the peak value of the K-mer distribution. According to the K-mer based estimation theory, the size of the haploid genome of *F. hispida* was estimated to be 366 Mb, which is similar with the genome size of *F. carica* (Mori et al., 2017). However, the estimated genome size based on K-mer of *F. microcarpa* is 430 Mb, 60 Mb larger than *F. hispida*. To validate the genome size estimation, we further performed flow cytometry analysis to estimate the nuclear DNA content. The estimated genome sizes are consistent with our K-mer based estimation with an average 436 Mb for *F. microcarpa* and 370 Mb for *F. hispida*.

Contig assembly

The full PacBio subreads of the two *Ficus* genomes were self-corrected using CANU (Koren et al., 2017) version 1.7 with parameter corOutCoverage = 80 and corrected reads were subjected to three widely-used Pacbio assembler, including CANU (Koren et al., 2017), FALCON (Chin et al., 2016) and SMARTdenovo. Each round of assemblies was inspected through evaluation of N50s, assembled genome size as well as complete/duplicated BUSCO ratio (Table S1). To improve the contiguity of contig assemblies, we used Quickmerge program (Chakraborty et al., 2016) to reconcile FALCON and SMARTdenovo assemblies for *F. microcarpa*, while integrate CANU and SMARTdenovo assemblies for *F. hispida* (Table S1). The total length of the final contig assembly for *F. microcarpa* was 426 Mb with a contig N50 size of 908 kb, and the assembled size of *F. hispida* is 360 Mb with a contig N50 size of 492 kb. The contig-level assembly of the wasp genome was achieved using a similar strategy with *Ficus* genome assemblies by combining 170 × PacBio long reads and 84 × Illumina short reads. A total of 65 Gb subreads generated from the PacBio Sequel platform

were assembled using CANU assembler (Koren et al., 2017) with default parameters. The redundant sequences were identified and removed based on a whole-genome alignment strategy implemented by Redundans program (Pryszcz and Gabaldón, 2016) with at least 90% identity and 60% overlap between contigs. The resulting contigs was further polished by illumina short reads using Pilon program (Walker et al., 2014) with parameters:–diploid –threads 6 –changes –tracks –fix bases –verbose –minddepth 4. The genome assemblies were assessed using a series of approaches, including BUSCO (Simão et al., 2015), Illumina reads mapping, alignment of RNA-seq assembled transcripts.

Hi-C scaffolding and chromosome assembly

Young leaves collected from a single tree (*F. microcarpa* or *F. hispida*) were delivered to BioMarker Company for Hi-C libraries construction using the previously published approach (Xie et al., 2015). Chimeric fragments representing the original cross-linked fragments were then processed into paired-end sequencing libraries and sequenced on the Illumina HiSeq X Ten platform. Hi-C reads were uniquely mapped to the contig assemblies and reads within 500 bp regions of HindIII restriction sites were retained for further analysis. Mis-joined contigs were corrected by detecting abrupt long-range contact patterns using the 3D-DNA pipeline (Dudchenko et al., 2017). The Hi-C corrected contigs were further linked into 13 pseudo-chromosomes in *F. microcarpa* and 14 pseudo-chromosomes in *F. hispida* using the ALLHiC pipeline (Zhang et al., 2019). The accuracy of Hi-C based chromosome construction was evaluated by chromatin contact matrix (Figure S1).

Genetic map and validation of genome assemblies

Genetic map of *F. hispida*

Seeds were collected from a single mother tree of *F. hispida* located in Xishuangbanna Tropical Botanical Garden and 2-month seedlings were collected for whole genome re-sequencing. Sex was identified for each F₁ individual using a male-specific primer designed in this study (see Key Resources Table). The maternal and paternal ultra-high density genetic maps were constructed based on re-sequencing of 30 male F₁ individuals, 30 female F₁ individuals from a segregating population as well as their maternal tree. Meanwhile, the reference genome sequencing data was served as paternal samples. Reads were then aligned to *F. hispida* reference genome assembly and SNP calling was performed using the Genome Analysis Toolkit (McKenna et al., 2010) following the best practices workflow for variant discovery. A total of 4,450,765 high-quality SNPs were identified and these SNPs were further merged into bin-markers based on a modified version of sliding-window approach (Huang et al., 2009). The genetic map assembled into 14 linkage groups using JoinMap 5.0 (Stam, 1993), spanning a total of genetic distance of 1,641.86 cM in maternal genetic map and 1,884.99 cM in paternal genetic map (Figure S1C).

Validation of genome assembly

The genome completeness was assessed based on 1,440 conserved plant genes in BUSCO program (Simão et al., 2015) with default parameters. In addition, PASA assembled transcripts based on RNA-seq reads were aligned to the two *Ficus* genome assemblies and the results showed that more than 99.9% of transcripts were alignable to the assemblies, with 99.88% and 99.93% of single base accuracy for *F. microcarpa* and *F. hispida*, respectively. We also mapped the illumina reads from short insert size libraries back to genome assemblies using BWA (Li and Durbin, 2009) (version 0.7.8). Our results revealed that 97.98% of *F. microcarpa* reads and 96.28% of *F. hispida* reads were mappable to the assemblies, suggesting the high completeness of the assemblies (Table S3). Moreover, 99.23% of *F. microcarpa* genome sequences and 98.92% of *F. hispida* sequences were covered by the illumina reads, indicating assembly accuracy in our assemblies (Table S3). Genome wide analysis of chromatin interactions further validated the Hi-C scaffolding of the two *Ficus* genomes. Presence of prominent blocks along the diagonal indicates high quality assembly, as contacts between nearby regions are generally more frequent than contacts between remote regions. Otherwise, incorrectly assembled genome usually exhibits incongruities such as contact-enriched blocks located away from the diagonal. The Hi-C heatmap revealed a well-organized interaction contact pattern along the diagonals within each pseudo-chromosome (Figure S1). In addition, comparison of genetic map of *F. hispida* and the Hi-C based assembly reveals high consistency, indicating the accuracy of our assembly (Figure S1D).

Genome annotation of *F. microcarpa* and *F. hispida*

RNA extraction and sequencing

Total RNA was extracted from selected samples in this study using RNAPrep Pure plant Kit and genomic DNA contamination was removed using RNase-Free DNase I (Takara Company, China) following the manufacturer's instructions. RNA integrity was evaluated on a 1.0% agarose gel stained with ethidium bromide (EB). The quality and quantity were further assessed using NanoDrop 2000C and Agilent 2100 platforms. For illumina sequencing, 3 µg RNA from each sample were used to cDNA library construction using the NEBNext Ultra RNA library Prep Kit. The PCR products obtained were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system. Library preparations were sequenced on illumina HiSeq 2500 or HiSeq X ten systems according to experimental designs.

Repeat annotation

We first customized a *de novo* repeat library of the genome using RepeatModeler, which can automatically execute two *de novo* repeat finding programs, including RECON (version 1.08) (Bao and Eddy, 2002) and RepeatScout (Price et al., 2005). The consensus TE sequences generated above were imported into RepeatMasker v4.05 to identify and cluster repetitive elements. Unknown TEs were further classified using TEclass (Abrusán et al., 2009). To identify tandem repeats within the genomes, the Tandem Repeat

Finder (TRF) package (version 4.07, citation) ([Benson, 1999](#)) was used with the modified parameters of “1 1 2 80 5 200 2000 –d -h” to find high order repeats.

Gene annotation

Annotation of protein-coding genes was based on *ab initio* gene predictions, transcript evidence and homologous protein evidence, all of which could be implemented in the MAKER computational pipeline ([Cantarel et al., 2008](#)). Briefly, RNA-seq samples were assembled by Trinity ([Haas et al., 2013](#)) with two different strategies using default parameters, *de novo* assembly and genome-guided assembly. Assembled transcripts were inspected by calculation of FPKM expression value and removed if FPKM < 1 and iso percentage < 3%. The filtered transcripts were imported to PASA program ([Haas et al., 2003](#)) for construction of comprehensive transcripts as PASA is able to take advantage of the high sensitivity of reference-based assembly while leveraging the ability of *de novo* to detect novel transcripts. The nearly “full-length” transcripts selected from PASA assembled transcripts were imported to data training programs, including SNAP ([Korf, 2004](#)), GENEMARK ([Besemer et al., 2001](#)) and AUGUSTUS ([Stanke et al., 2006](#)). Afterward, MAKER pipeline was used to integrate multiple tiers of coding evidence and generate a comprehensive set of protein-coding genes. The second round of MAKER was run to improve gene annotation. The predicated gene models with AED scores less than 0.2 were extracted for re-training using SNAP ([Korf, 2004](#)), GENEMARK ([Besemer et al., 2001](#)) and AUGUSTUS ([Stanke et al., 2006](#)). In addition, the RNA-seq reads were mapped to genomes using HiSAT2 ([Kim et al., 2015](#)) and re-assembled using StringTie ([Pertea et al., 2016](#)). The assembled RNA-seq transcripts along with homologous proteins from *Arabidopsis thaliana*, *Morus notabilis*, papaya, grape, tomato and rice were imported to MAKER pipeline. After filtering putative transposon-derived gene models, a total of 29,146 and 27,211 gene models were annotated in *F. microcarpa* and *F. hispida*, respectively. BUSCO evaluation of the protein-coding annotation showed 94.4% completeness in *F. microcarpa* and 95.7% in *F. hispida* ([Table S4](#)).

miRNA annotation

Publicly available plant miRNAs were downloaded from miRBase (latest access in July 5, 2018) and mapped to the two genomes using bowtie ([Langmead et al., 2009](#)) with at most 3 mismatches allowed. Alignments were filtered using a PERL script (filter_alignments.pl) implemented in miRDP1.3 package ([Yang and Li, 2011](#)). The RNA secondary structures were predicted using RNAfold program ([Lorenz et al., 2011](#)). We further applied miRDeep-P program to identify and filter miRNAs, which used an improved plant scoring system and specifically predict miRNAs for plants.

Identification of segmental duplications

Identification of segmental duplications in the two *Ficus* genomes were based on previously reported method ([Bailey et al., 2002](#)) with minor revision. Briefly, the genome assemblies were fragmented into trackable 400-kb segments and transposable elements within each segment were removed leaving the unique genome for further whole genome alignments using LASTZ program. Pairwise alignments identified duplicated segments if they have > 88% identity and > 500 bp aligned length. The second round of whole genome alignments are performed using the sequences with TE reinserted and the final set of segmental duplications were identified if the global alignments reported > 1-kb bases matched and > 90% identity. The source code of our modified WGAC based SD detection method were archived in GitHub. The shared SDs between the two *Ficus* genomes were determined based on BLAST ([Altschul et al., 1990](#)) alignments with evalue 10^{-3} . Redundant aligned regions were merged into a comprehensive alignment block using a home-made PERL script. The total length of alignment block within each BLAST hit were calculated and considered as the size of shared SDs.

Identification of structural variations (SVs) and copy number variations (CNVs)

Identification of SVs

De novo genome assembly of re-sequenced *Ficus* samples were performed using MaSuRCA ([Zimin et al., 2013](#)) program with parameters designed for large genomes. Assembled contigs were mapped against *F. hispida* reference genome using MUMmer 4 whole genome alignment program ([Marçais et al., 2018](#)). Further, structure variations were identified using Assemblytics ([Nattestad and Schatz, 2016](#)), which is a Web-based SV analytics tool.

Identification of CNVs

The CNV and deletion calling method from Bickhart ([Bickhart et al., 2012](#)) was applied with some revisions similar as we did before ([Zhang et al., 2015](#)). Briefly, the whole genome was thoroughly masked using RepeatMasker with parameter “-s,” Tandem Repeats Finder ([Benson, 1999](#)) and WindowMasker ([Morgulis et al., 2006](#)) and further extended 36 bp in both directions to remove edge effects for mapping short reads. Reads were counted in overlapping, sliding 5-kb windows of non-masked, non-gapped sequences. To avoid sequencing biases in GC-rich and GC-poor regions with high-throughput sequencing technologies, the GC bias was corrected using LOESS smoothing toward a pattern of uniform coverage at all GC percentage bin as previously described ([Bailey et al., 2002](#)). CNV candidate windows were initially defined as having six out of seven or more sequential overlapping windows with read depth values that differed significantly from the whole-genome average depth (> mean+2 × STDEV for duplications and < mean – STDEV for deletions). The breakpoints were refined using 1-kb sliding window if they lay on the edges of the long windows and met the cutoff value of CNVs. The absolute copy number of each gene was calculated based on their coordinates in annotated GFF. The CNV and deletion regions were visualized using JBrowse ([Buels et al., 2016](#)).

Phylogenetic tree reconstruction

Protein sequences of single-copy genes from five plant species including *F. microcarpa*, *F. hispida*, *Arabidopsis thaliana*, *Morus notabilis*, and *Oryza sativa* were extracted from self-blast results using a home-made PERL script. MUSCLE (Edgar, 2004) was used to generate multiple sequence alignments for protein sequences in each single-copy family with default parameters. The alignments for each family were then concatenated into an alignment supermatrix for prediction of the best substitution models and tree reconstruction using the IQ-TREE program (Nguyen et al., 2015).

Phasing of X and Y chromosomes in *F. hispida*

Identification of X and Y specific SNPs

To study the sex chromosome evolution in *F. hispida*, we developed a pipeline to *de novo* assemble the X and Y chromosome from the mixed assembly of FhChr12 (Figure S5B). First, whole genome re-sequencing reads from 3 male and 3 female individuals with ~40 x coverage for each sample were mapped to *F. hispida* reference genome, respectively, and further subject to SNP calling using the genome analysis toolkit (GATK; V 3.5-0-g36282e4) (McKenna et al., 2010) following the best practices workflow for variant discovery. SNPs were further clustered into X specific locus if they were only homozygous in female individuals and Y specific locus if they were heterozygous and only existed in male samples.

Pacbio reads phasing

The corrected Pacbio reads from *F. hispida* genome were mapped against the reference genome using minimap2 program (Li, 2018) and the resulted alignment file was used for variants calling in GATK 4 pipeline (McKenna et al., 2010). Subsequently, SNPs with genotype “0/1” were selected for phasing. We used WhatsHap (Patterson et al., 2015), a fixed parameter tractable (FPT) approach to solve wMEC (weighted minimum error correction) program, to generate the linked SNPs which were present in two different haplotypes for the diploid genome. The linked SNPs were considered as phased SNP blocks. A total of 580 phased SNP blocks were identified in *F. hispida* chromosome 12 and these blocks were further used for identification of X or Y specific SNP blocks. The X-specific and Y-specific SNP locus identified above were recruited and traced back to the phased SNP blocks. The phased SNP blocks were considered as origination from X chromosome (X-Phased SNP blocks) if they contained a dominant proportion (> 70%) of X-specific SNPs than Y-specific SNP. Y-Phased SNP blocks were identified using the same method. We totally phased 163 sex specific blocks, containing 56,294 heterozygous SNPs (genotype 0/1). Pacbio reads were further traced based on the heteryzgous SNPs for each phased block. Pacbio reads could be clustered into reads from X chromosome if they contained variants located in X-Phased SNP blocks, and vice versa for Y originated reads. Since Y-specific regions are only present in Y chromosome and absent in X chromosome, the SNP based phasing method will not include the pacbio reads located in these Y-specific regions. Therefore, we manually added these regions into a list of Y-specific reads. Totally, 78,235 X-specific and 77,926 Y-specific reads were identified based on our method.

De novo assembly of X and Y chromosomes

De novo contig level assemblies for X and Y specific reads were performed separately using CANU (Koren et al., 2017) with default parameter. Hi-C reads originated from *F. hispida* chromosome 12 were re-mapped to the X and Y contigs, and scaffolding of X chromosome and Y chromosome were constructed using ALLHiC pipeline (Zhang et al., 2019) with diploid assembly model. Finally, a X chromosome comprising of 21.9 Mb sequences, and a Y chromosome containing 22.6 Mb sequences were constructed.

Identification of sex-determining regions in *F. hispida*

A total of 15 male and 11 female individual *F. hispida* plants were collected from Xishuangbanna. Genomic DNA was extracted from young leaves using the QIAGEN DNeasy Plant Mini Kit and libraries with 300–500-bp inserts were sequenced on the Illumina X Ten platform to obtain 150-bp reads. SNPs and InDels were called using a method described previously (Zhang et al., 2018). Briefly, Illumina short reads were mapped against the *F. hispida* reference genome using BWA (Li and Durbin, 2009) with default parameters. Variant detection was performed using the Genome Analysis Toolkit (McKenna et al., 2010) following the best practices workflow for variant discovery. High-quality SNPs were retained after filtering out SNPs located within repetitive regions and SNPs with extremely low or high depth. Heterozygous sites were calculated for 100-kb sliding windows in 50-kb steps for each individual. Potential SDRs were identified when heterozygosity of genomic regions in male individuals differed significantly from that in genomic regions in female individuals (*t* test, *p* < 0.01). Selected windows were joined together if they contained overlapping coordinates. Finally, a 2-Mb region located on chromosome 12 was identified as sex-determining region.

Identification of differentially expressed genes between male and female fruits

Female and male flowers were collected under three stages A, B and C. Total RNA was extracted from selected samples in this study using RNAPrep Pure plant Kit and genomic DNA contamination was removed using RNase-Free DNase I (Takara Company, China) following the manufacturer's instructions. The extracted RNA from male and female fruits with three biological replicates were sequenced on illumina HiSeq X10 platform, yielding 4 Gb data for each sample. RNA-seq reads were trimmed using Trimmomatic (Bolger et al., 2014) program and mapped against 27,211 *F. hispida* annotated gene models using bowtie with only best alignment retained. FPKM (fragments per kilobase of exon per million fragments mapped) were calculated using RSEM (Li and Dewey, 2011) program, which was implanted in Trinity package (Haas et al., 2013). Further, the DEGs (differentially expressed genes) were identified using DESeq2 (Love et al., 2014). GO enrichment and KEGG pathway analysis was performed using the OmicShare tools, a free online platform for data analysis.

Population genomics**Variants calling**

A total of 2.5 billion 150-bp paired-end Illumina reads yielded an average coverage of $16 \times$ per accession. Raw reads were quality filtered to remove adapters and low quality bases ($Q < 30$). Quality filtered reads were aligned against the *F. microcarpa* reference genome using bwa with default parameters. Variant detection was performed using the genome analysis toolkit (GATK; V 3.5-0-g36282e4) (McKenna et al., 2010) following the best practices workflow for variant discovery. Resulting BAM files were locally realigned using the IndelRealigner to remove erroneous mismatches around small-scale insertions and deletions. Variants were called in each accession separately using the HaplotypeCaller and individual gVCF files were merged using GenotypeGVCFs. This two-step approach includes quality recalibration and re-genotyping in the merged vcf. file, ensuring variant accuracy. Meanwhile, samtools/bcftools (Li, 2011) were also applied to call SNPs using the same data with default parameters. SNPs were filtered based on the following criteria: 1) SNPs were removed if they were only present in one of the two pipelines (GATK and SAMtools/BCFtools); 2) SNPs in repeat regions; 3) SNPs with read depth $> 1,000$ or < 5 ; 4) SNPs with missing rate $> 40\%$; 5) SNPs with < 5 bp distance with nearby variant sites; 6) non- biallelic SNPs were removed.

Maximum-likelihood tree inference and ADMIXTURE analysis

Phylogenetic trees of nuclear and chloroplast DNA sequences were constructed based on SNPs within regions of single-copy genes. The filtered SNPs were converted to phylip and aligned fasta format. The maximum likelihood (ML) trees were constructed using two popular programs: IQ-Tree (Nguyen et al., 2015) with self-estimated best substitution models and RAxML (Stamatakis, 2014) with GTRCAT model. We observed similar topology structures in the two programs. Ancestral population stratification among the whole *Ficus* genus was inferred using ADMIXTURE software (Alexander and Lange, 2011). The optimal ancestral population structure was estimated from the same variants set with STRUCTURE using ancestral population sizes $K = 1-20$ and choosing the population with the smallest cross-validation error ($K = 9$ in this case). The parameter standard errors were estimated using bootstrapping (bootstrap = 200) when doing the ADMIXTURE analyses.

Coefficient of genetic relatedness

The coefficient of genetic relatedness was calculated based on 200-kb non-overlapping windows followed by the previous published method (Wu et al., 2018).

$$r = 0.5 * IBD1 + IBD2$$

where $IBD1$ and $IBD2$ indicate the genomic proportions sharing one or two haplotypes that are identical by descent.

$IBD0$ (sharing zero haplotype), $IBD1$ and $IBD2$ can be inferred by calculating the identical-by-state ratio (IBSR) and genomic distance D. The IBSR is calculated as the following formula:

$$IBSR = \frac{IBS2}{IBS2 + IBS0}$$

where $IBS0$ is the number of sites without allele sharing (i.e., number of joint-genotype AA|BB) within a non-overlapping window, while $IBS2$ is the number of sites sharing two different alleles (i.e., number of joint-genotype AB|AB).

The genomic distance D can be calculated using the following formula:

$$D = 1 - 0.25 * \frac{\pi_1 + \pi_2}{\pi_{12}}$$

where π_1 and π_2 are heterozygosity for each single individual and π_{12} is the pairwise sequence divergence between the two individuals.

If $IBSR < 0.95$, the genomic window is assigned $IBD0$. If $IBSR \geq 0.95$ and $D > 0.05$, the window is assigned $IBD1$. The last case ($IBSR \geq 0.95$ and $D < 0.05$) is inferred as $IBD2$.

ABBA-BABA analysis

To detect the introgression between section *Oreosycea* and other sections, we calculated the D statistic using the program doAbbababa2, implemented in ANGSD (Korneliussen et al., 2014). D-statistic is widely used to examine site patterns (also known as ABBA/ABAB patterns) in genome alignments for a specified four-taxon tree. Given four taxa with the relationship of “[P1,P2],P3,O,” a D-statistic significantly differed from zero indicate introgression between population P1 and P3 (negative D value) or between P2 and P3 (positive D value; Durand et al., 2011).

Dating the divergence time of the genus *Ficus*

The divergence time of the genus *Ficus* was estimated using MCMCTREE algorithm, implemented in PAML program (Yang, 2007). The MCMCTREE performed the Bayesian inference using SNPs located in single-copy genes identified from *F. microcarpa* reference genome with the following settings: 1) ‘GTR’ model; 2) the ‘rgene_gamma’ parameter; 3) branch lengths estimated by BASEML program; 4) Markov chain Monte Carlo (MCMC) analysis with 2,000,000 generations; 5) a burn-in of 1,000 iterations. The other parameters were left with default. Meanwhile, seven fossil evidences from previous studies (Craaud et al., 2012; Gardner et al., 2017)

were provided as soft calibration for the dating analysis, including 1) a maximum age of 100 Mya for the Crown of Moraceae; 2) 60-90 Mya for the Crown of the genus *Ficus*; 3) a minimum of 43 Mya for the stem of section *Urostigma*; 4) a minimum of 34 Mya for the stem of subgenus *Sycomorus*; 5) a minimum of 26 Mya for the stem of subgenus *Sycidium* and 6) subgenus *Synoecia*; 7) a minimum of 22 Mya for the stem of section *Americana*. Tracer v1.5 (Rambaut et al., 2018) was used to examine the sampling adequacy and convergence of the chains to a stationary distribution.

Cophylogenetic and codivergence analysis between figs and wasps

Cophylogenetic analysis between figs and wasps was performed using Jane 4 (Conow et al., 2010) in solve mode with 40 generations and a population size of 1,000. Statistical tests were performed by mapping 100 random tips with 40 generations and a population size of 1,000. Codivergence analysis between *Sycomorus* species and their pollinators relied on Tajima's D testing of selected genes that might be under pairwise selection in their obligate mutualism. The whole genomes of 14 *Sycomorus* figs and their pollinator wasps were sequenced on the Illumina X Ten platform at an average depth of 10 × for each accession. Variants were filtered and identified by mapping short reads to our newly sequenced *F. hispida* reference genome and the published *C. solmsi* wasp genome using the methods aforementioned. To identify selective sweeps, Tajima's *D* values were calculated using vcf.-kit for 10-kb sliding windows in 5-kp steps. We also performed SweeD analysis (Pavlidis et al., 2013), which utilized composite likelihood ratio (CLR) test to identify loci that display a strong skew in the Site Frequency Spectrum (SFS) toward rare variants in comparison with the genomic background. Both of the two analysis used the top 5th percentile as cutoff values to infer candidate selective sweeps and the overlapped results between the two approaches were considered as high-confident regions.

Identification of active volatiles compounds for pollinator attraction

To identified the main active volatiles compounds of subgenus *Sycomorus* figs responding to the specific pollinator attraction, three fig species and their obligate pollinator wasps, including *F. hispida* - *C. solmsi*, *F. auriculata* - *C. emarginatus* and *F. semicordata* - *C. gravelyi*, were chosen to represent three main sections (section *Sycocarpus*, *Sycomorus* and *Hemicardia*) of this subgenus. The fresh fig syconia arriving in receptive phase, when female flower mature for pollination and oviposition by pollinator, were detached from figs trees. And, within 30min, after washing the latex at the wound under running tap water to remove defensive volatiles, they were kept at ambient temperature in the glass beaker with deionized water covering the wound on carpodium to inhibit the production of defensive VOCs. The beaker with syconias were sealed using the odorless paraffin film. Odor from receptive-phase syconia were then collected using the solid phase micro-extraction (SPME) with a 65 μm PDMS/DVB fiber (Supelco, CA) over the headspace for 1 hour. The HS-SPME samples were used directly for compounds identification with the Gas Chromatography-Mass Spectrometry (GC-MS) and their electrophysiological stimulus to corresponding pollinator wasp with the gas chromatography-electroantennogram detection (GC-EAD). Pollinator wasps attracted to the receptive phase syconia of three host fig species in the field were collected, identified and used for GC-EAD analysis within 30 min. Previous behavior studies have demonstrated that pollinator wasps are only been attracted by fig syconia in receptive phase (Chen et al., 2009).

An HP7890-5975C (Agilent, US) GC-MS system was used with 1mL/min Helium as carrier gas. The splitless inlet with a small inner diameter liner for SPME samples (Supelco, CA) was heated to 250°C. A 30 m × 250 μm × 0.25 μm HP-5ms (Agilent, US) column was used. The oven ramp was set as 40°C for 1 min and then 6°C/min to 250°C for 5 min. The transfer line was heated to 150°C. In the quadrupole mass spectrometry, a 70 eV EI ion source was used and heated to 230°C. The mass range scanned was m/z 28.5–350 at a rate of 8 scan/s. The abundance threshold for detection was set to 10. Data were collected and analyzed using Chemstation software (Agilent Technologies, US). The GC-MS analysis was repeated nine times. We used a custom-made GC-EAD system, to test the antennal electrophysiological responses of wasps. Head of each wasp was cut down at cephalothorax using a scalpel. The tips of antennae were cut open using a pair of iris scissors under microscope. Sharpened glass electrodes filled with Ringer's solution was used to conduct the electro antennal signal. The opening at the cephalothorax were mounted to the reference electrode, the tip of each antennae was mounted to two probes respectively under universal stereo microscope with the help of 15 μm XYZR custom-made micromanipulator modified from optic micro-displacement platform (Shengning, CN). In the custom-made probe amplifier, the signal from antennae was amplified 21 times using LMP 7721 operational amplifier (TI, US) with a 0.01- 15Hz filter. Voltage signals were then followed and further filtered using an OPA 2228 operational amplifier (TI, US) at 0.01-10Hz. The probe was powered by battery. The outputted signal was then feed to a HP 34465A digital multimeter (Keysight, US) in DC measuring mode at 10 reads per second. Signal data were recorded using free BenchVue software (Keysight, US). A Trace GC (Thermo, US) was used to separate the SPME samples with 1ml/min Helium as carrier gas. The splitless inlet with a SPME inner was heated to 250°C. A 30 m × 320 μm × 0.32 μm HP-5(Agilent, US) column was used. The oven ramp was set as 50°C for 1 min and then 10°C/min to 280°C for 5 min. The EAD transfer line was heated to 250°C. The column effluents were cooled and diluted by a clean air flow in as odor pipette before it was delivered to the antennae preparation. The GC-EAD tests were repeated nine times.

Identification of Hormone levels in plants

Chemicals and reagents

HPLC-grade acetonitrile (ACN) and methanol (MeOH) were purchased from Merck (Darmstadt, Germany). MilliQ water (Millipore, Bradford, USA) was used in all experiments. All of the standards were purchased from Olchemim Ltd. (Olomouc, Czech Republic) and Sigma (St. Louis, MO, USA). Acetic acid was bought from Sinopharm ChemicalReagent (Shanghai, China). The stock solutions

of standards were prepared at the concentration of 10 mg/mL in ACN. All stock solutions were stored at –20°C. The stock solutions were diluted with ACN to working solutions before analysis.

Sample preparation and extraction

Fresh plant materials were harvested, weighted, immediately frozen in liquid nitrogen, and stored at –80°C until needed. Plant materials (50 mg fresh weight) were frozen in liquid nitrogen, ground into powder, and extracted with 0.5 mL methanol/water/formic acid (15:4:1,V/V/V) at 4°C . The extract was vortexed (10 min) and centrifuged at 14,000 rpm under 4°C for 5 min. The supernatants were collected, repeat the steps above, vortexed (5 min) and centrifuged(5 min). The combined extracts were evaporated to dryness under nitrogen gas stream, reconstituted in 80% methanol (V/V), ultraphoniced (1 min) and filtrated (PTFE, 0.22 µm; Anpel) before LC-MS/MS analysis.

HPLC Conditions

The sample extracts were analyzed using an LC-ESI-MS/MS system (HPLC, Shim-pack UFCL SHIMADZU CBM30A system; <https://www.shimadzu.com.cn/>; MS, Applied Biosystems 6500 Triple Quadrupole, <http://www.appliedbiosystems.com.cn/>). The analytical conditions were as follows, HPLC: column, Waters ACQUITY UPLC HSS T3 C18 (1.8 µm, 2.1 mm × 100 mm); solvent system, water (0.04% acetic acid): acetonitrile (0.04% acetic acid); gradient program, 90:10V/V at 0min, 40:60 V/V at 5.0min, 40:60 V/V at 7.0min, 90:10 V/V at 7min, 90:10 V/V at 10 min; flow rate, 0.35 mL/min; temperature, 40°C; injection volume: 2 µL. The effluent was alternatively connected to an ESI-triple quadrupole-linear ion trap (Q TRAP)-MS.

ESI-Q TRAP-MS/MS

API 6500 Q TRAP LC/MS/MS System, equipped with an ESI Turbo Ion-Spray interface, operating in both positive and negative ion modes and controlled by Analyst 1.6 software (AB Sciex). The ESI source operation parameters were as follows: ion source, turbo spray; source temperature 500°C; ion spray voltage (IS) 4500 V; curtain gas (CUR) were set at 35.0 psi; the collision gas (CAD) was medium. DP and CE for individual MRM transitions was done with further DP and CE optimization. A specific set of MRM transitions were monitored for each period according to the plant hormones eluted within this period.

QUANTIFICATION AND STATISTICAL ANALYSES

All details of the statistics applied in this study are provided alongside the respective analysis in the [Method Details](#) section. The t test and Fisher exact test were performed using R 3.26 package.

Supplemental Figures

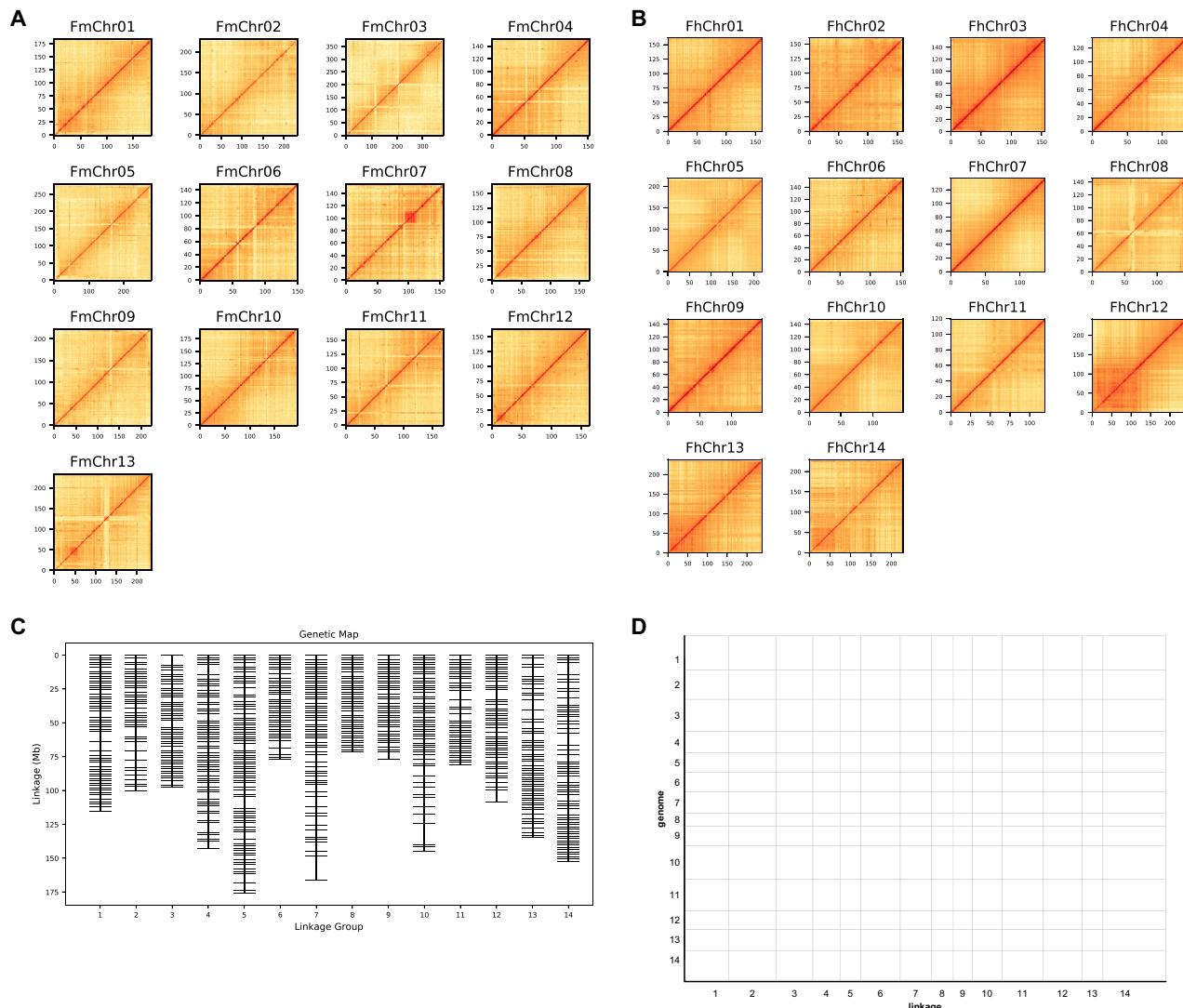


Figure S1. Assessment of Chromosome-Level Genome Assemblies of the Two *Ficus* Species, Related to Figure 1

Genome-wide analysis of chromatin interactions at 150-kb resolution in *F. microcarpa* (A) and *F. hispida* (B) genomes. (C) A high-density genetic map for *F. hispida* F1 population. (D) Comparison of F1 genetic map with Hi-C assembly in *F. hispida*.

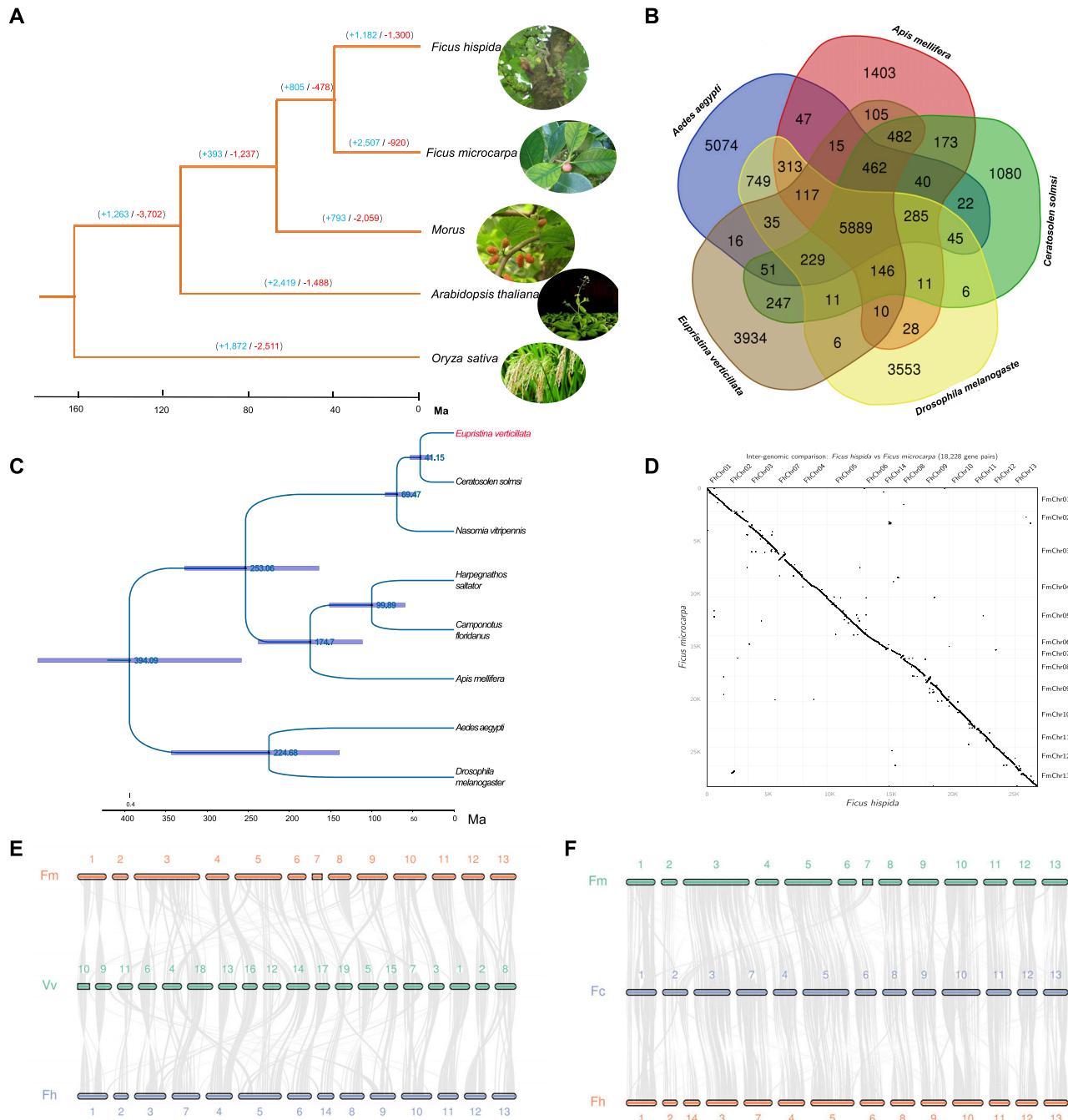


Figure S2. Phylogenetic and Comparative Analysis of Our Two De-Novo-Assembled *Ficus* Species and One Wasp Pollinator, Related to Figure 1

(A) Phylogenetic relationship of *F. microcarpa*, *F. hispida*, *M. notabilis*, *A. thaliana* and *O. sativa*. The divergence times among different plant species are labeled in the bottom. The numbers on each branch represent expansion (blue) and contraction (red) of gene families. (B) Venn diagram of orthologous and species-specific genes in different insect genomes. (C) Phylogenetic analysis and estimation of divergence time among multiple insect genomes. Divergence time were labeled on each branch. (D) Dot plot analysis between *F. microcarpa* and *F. hispida* genomes. (E-F) Syntenic analysis of *F. microcarpa* (Fm) and *F. hispida* (Fh) with *Vitis vinifera* (Vv) and *F. carica* (Fc).

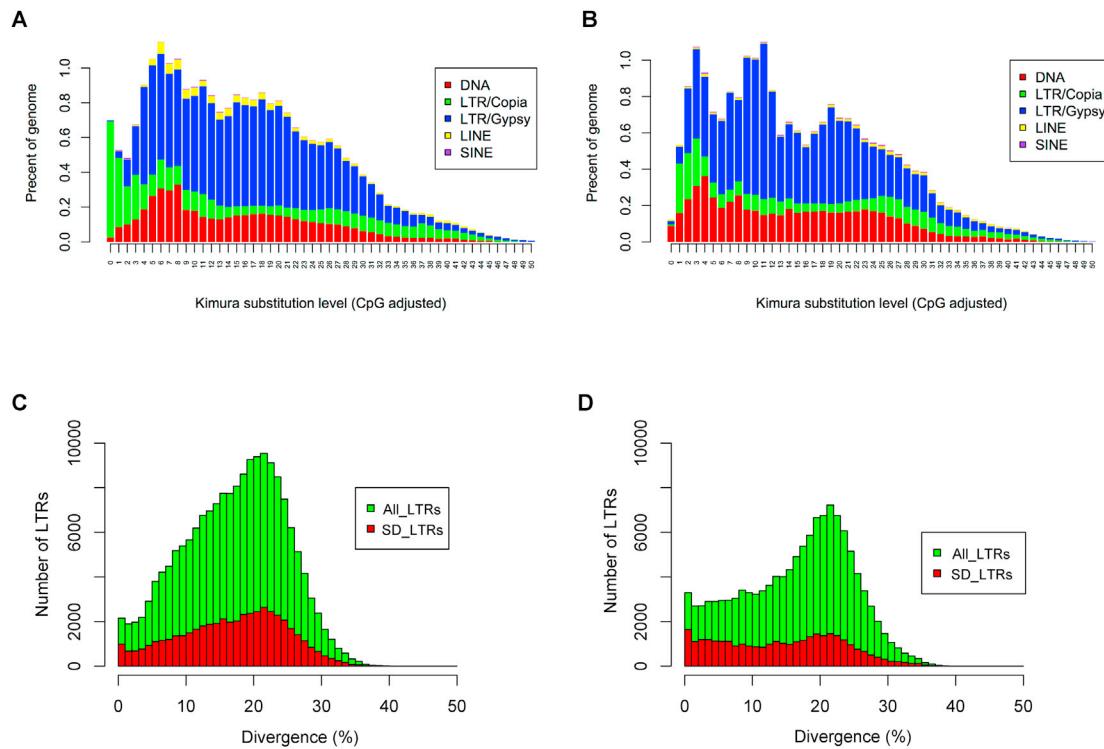


Figure S3. TE Distribution of Fm and Fh, Related to Figure 1

Kimura analysis of different type of TEs in *F. microcarpa* (A) and *F. hispida* (B). Number of LTRs and segmental duplication (SD) overlapped LTRs in *F. microcarpa* (C) and *F. hispida* (D). Green histograms indicate all of identified LTRs and red histograms represent SD-overlapped LTRs.

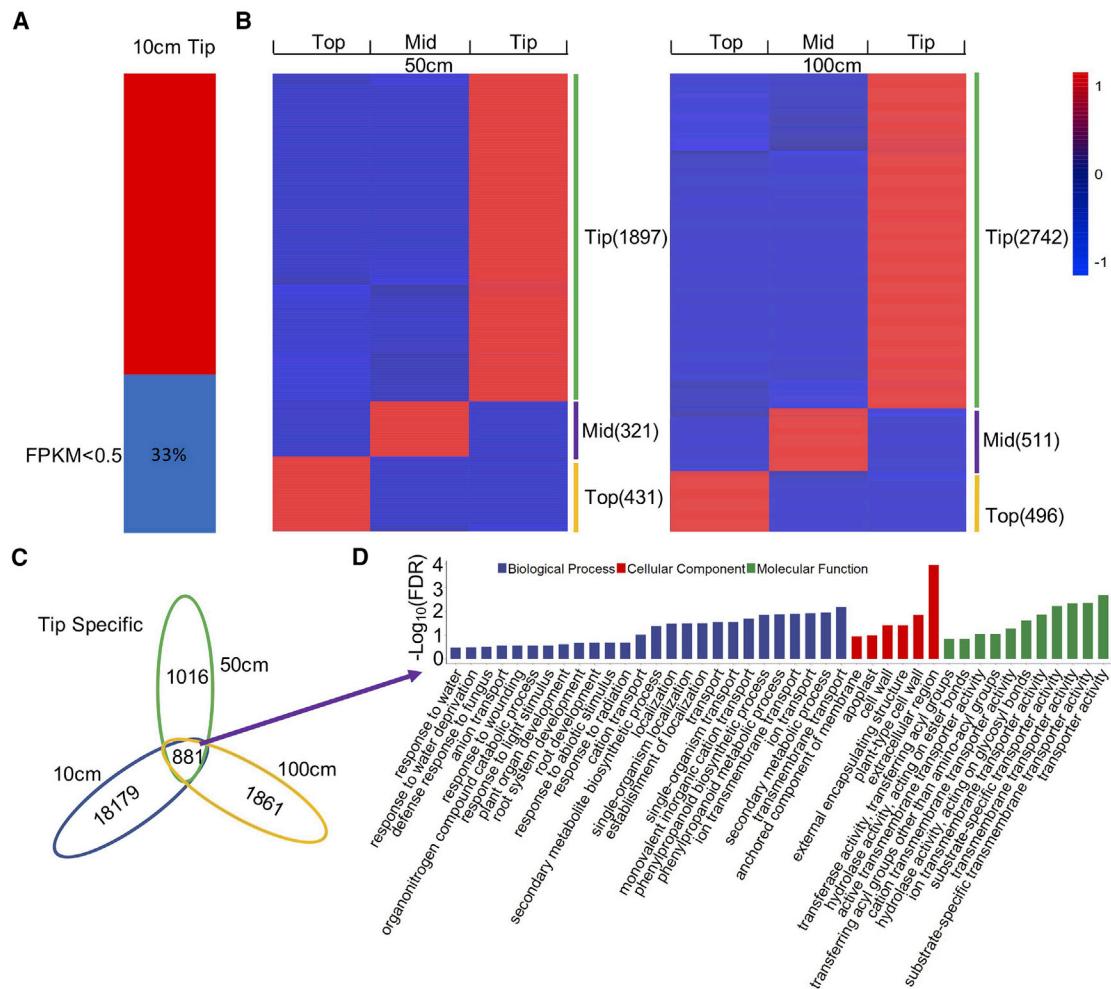


Figure S4. Identification of Highly Expressed Genes in Fm AR Tips, Related to Figure 2

Samples are collected from aerial root tissues at different stages, including ~10 cm (A), ~50 cm and ~200 cm (B) aerial root. For 50 cm and 200 cm long aerial roots, three parts are sampled, including top, middle (Mid) and tip. Tissue specifically expressed genes are identified using RNA-seq data. (C) Venn diagram shows the numbers of tissue specifically expressed genes and co-expressed genes in the tip tissue of all periods. (D) GO enrichment of the 881 highly expressed genes in aerial root tips ($FDR < 0.05$).

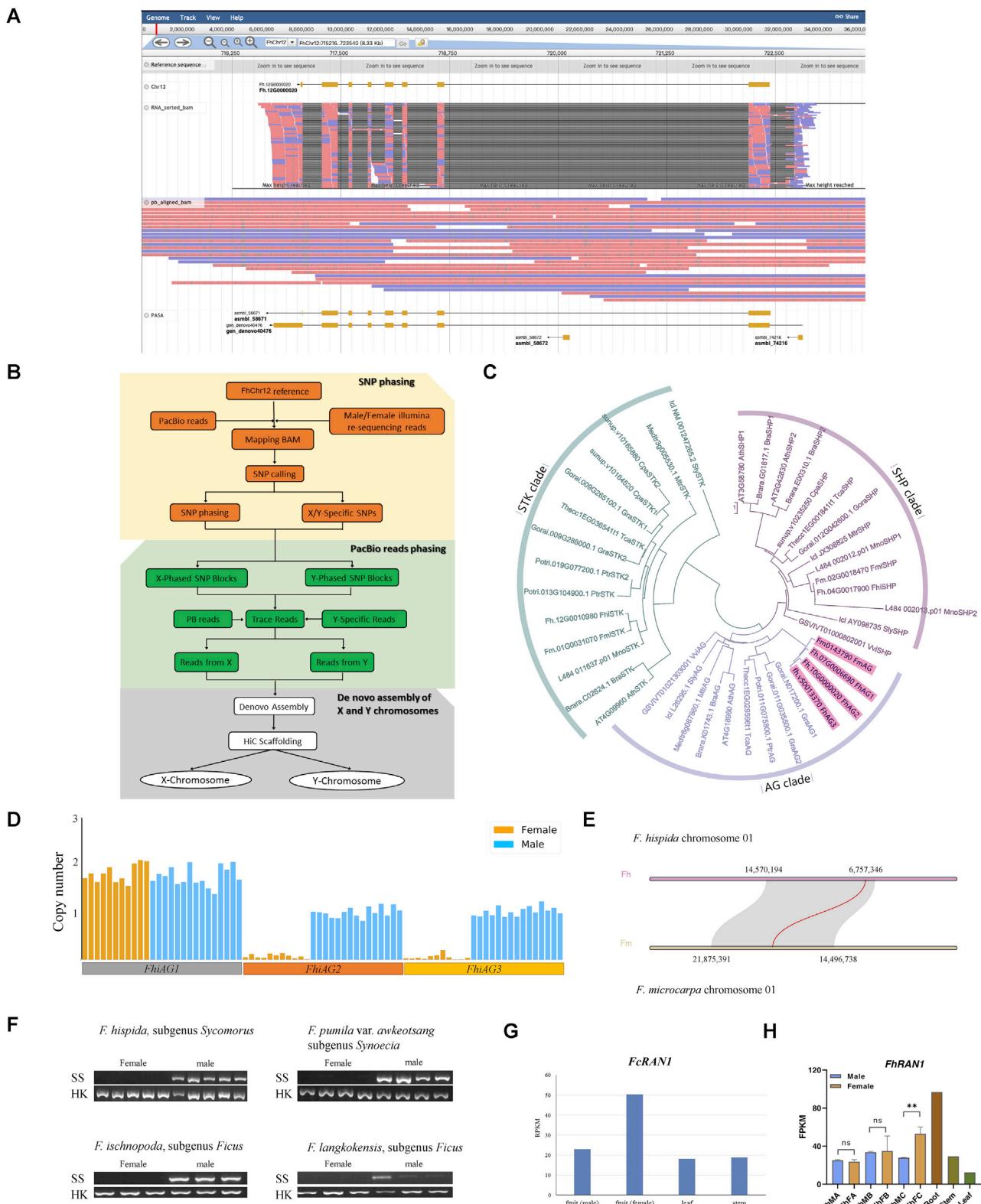


Figure S5. Sex Determination and Evolutionary History of AG Genes in Fh, Related to Figure 3

(A) Validation of *Fh.12G0000020* (*FhAG2*) gene in JBrowse. The top panel (Chr12) presents structure of this gene. The panel “RNA_sorted.bam” indicates alignment of RNA-seq reads, supporting alternative splicing sites for the gene structure. The panel “pb_aligned_bam” is alignment of PacBio corrected reads,

(legend continued on next page)

supporting accurate assembly of the genomic sequences. The final panel "PASA" includes RNA-seq assembled transcripts, also supporting the existence of this gene. (B) Assembly and phasing of X and Y chromosomes. (C) Phylogenetic tree of AG, STK and SHP genes in multiple species, including *F. microcarpa*, *F. hispida*, *A. thaliana*, *P. trichocarpa*, *T. cacao*, *B. rapa*, *M. truncatula*, *S. lycopersicum*, *C. papaya*, *M. notabilis* and *V. vinifera*. AG, STK and SHP clades are indicated with different colors. (D) Copy number variations of three AG genes in 11 female and 15 male individuals. (E) Syntenic analysis of *FhAG1* and *FmAG* flanking regions. AG gene pairs were marked by red line. (F) PCR analysis of sex specific primers in four *Ficus* species, *F. hispida*, *F. pumila* var. *awkeotsang*, *F. ischnopoda* and *F. langkokensis*. SS is sex-specific primer and HK indicates primer of house keeping gene alpha tubulin. (G) Expression of *FcRAN1* in *F. carica*. This figure is originally from *F. carica* genome publication (Mori et al., 2017) in Figure S5, showing differential expression of *FcRAN1* between male and female fruits. (H) Expression of *FhRAN1* gene in different tissues of *F. hispida*, including root, stem, leaf male and female inflorescences at variable developmental stages. FhMA, FhMB and FhMC indicate male inflorescences at stage A, B and C, respectively. FhFA, FhFB and FhFC indicate female inflorescences at stage A, B and C, respectively. Significance was tested by Student's t test; ** p < 0.01.

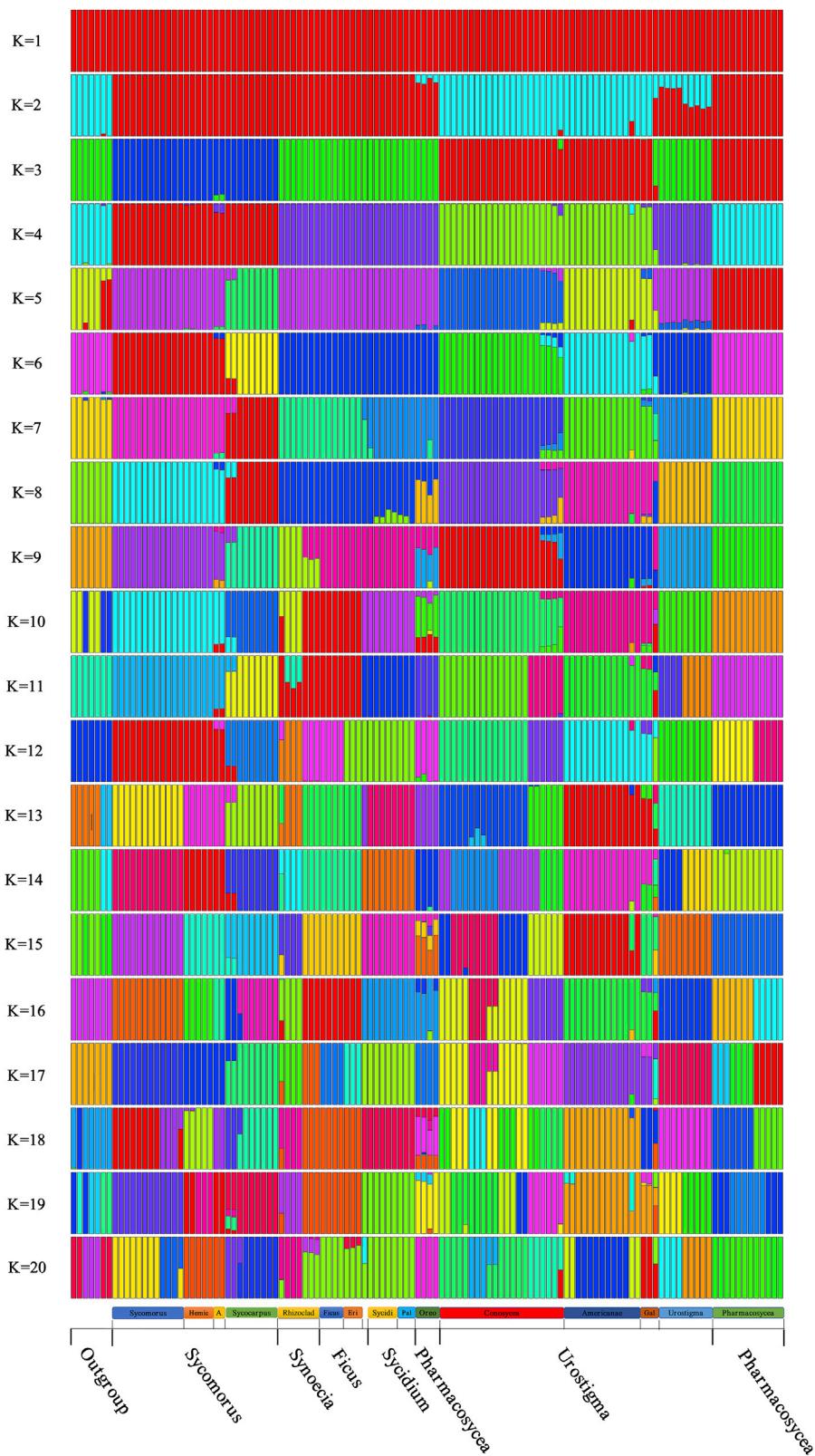


Figure S6. Admixture Analysis from K = 1 to K = 20, Related to Figure 4

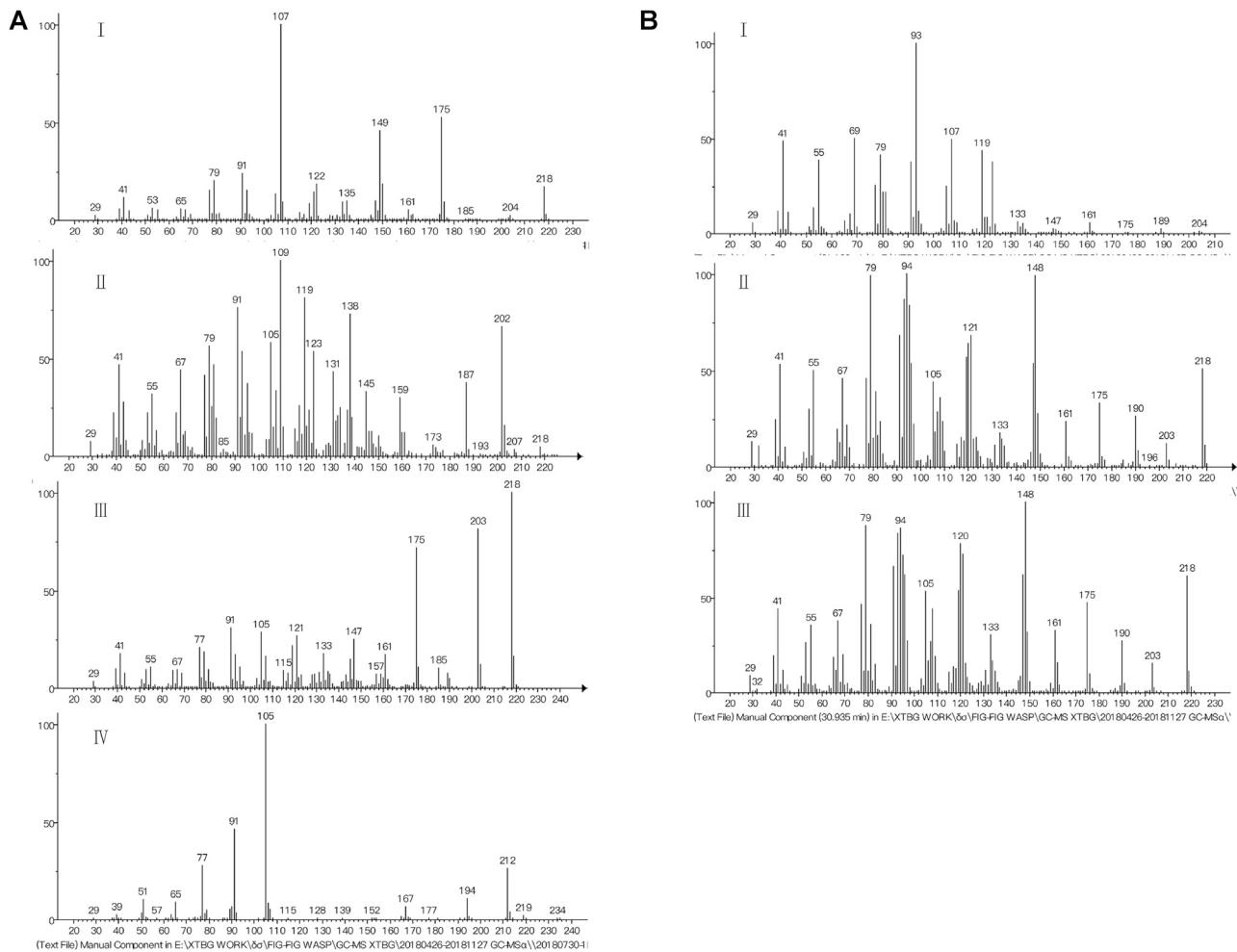


Figure S7. Mass Spectrum Analysis of Two *Ficus* Species, Related to Figure 5

(A) Mass spectra of the 4 strongest EAD active compounds in *F. hispida*. (I-IV) EAD active 1-4. In *F. hispida*, the mass spectra of the EAD major active components 1, 2 and 3 shows a molecular ion m/z 218 for the formula as $C_{15}H_{22}O$, indicating structures as sesquiterpenones, which is also supported by the similar fragmentation pattern of analogs in NIST MS DATA base. EAD active 4 (IV) has the same (> 90%) mass spectrum as benzyl benzoate. (B) Mass spectra of the 3 strongest EAD active compounds in *F. auriculata*. (I) EAD active 1 has the similar (< 70%) mass spectrum as α -Farnesene according data base searching; (II-III) the mass spectra of the EAD major active component 2, 3 shows a molecular ion m/z 218 for the formula as $C_{15}H_{22}O$, indicating a structure as sesquiterpenone, which is also supported by the similar fragmentation pattern of analogs in NIST MS DATA base.