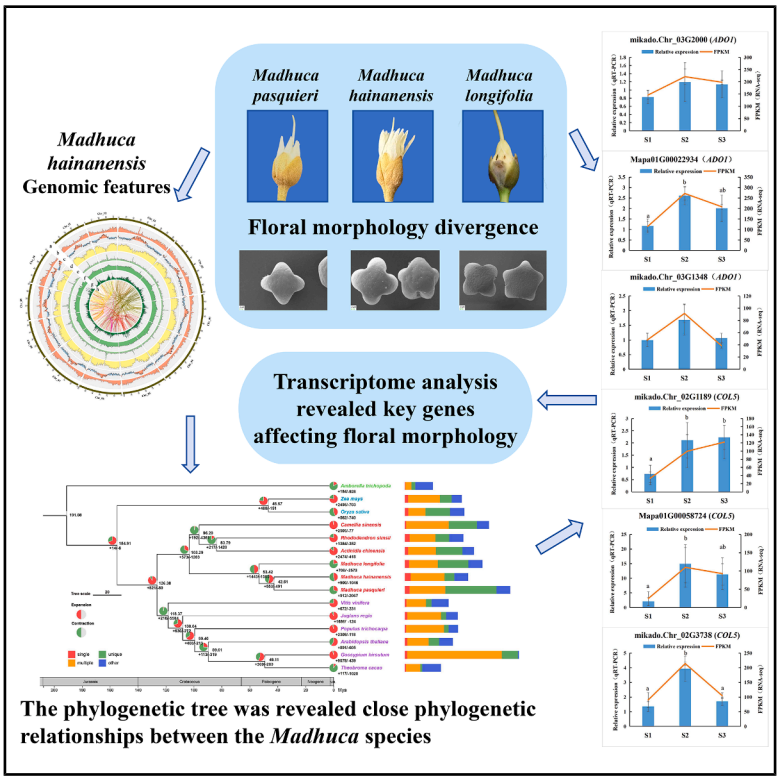
iScience

Article

Chromosome-level genome of *Madhuca hainanensis* reveals genomic evolution and floral divergence in *Madhuca*

Graphical abstract



Highlights

- Chromosome-level genome assembly of *Madhuca hainanensis* provides conservation data
- Reveals whole-genome triplication, shedding light on *Madhuca* genome evolution
- Identifies key genes (e.g., *ADO1*, *COL5*) linked to floral divergence in *Madhuca*
- Offers a valuable genomic resource for conservation of endangered *Madhuca* species

Authors

Yifei Ma, Ziyao Wang, Jingzhe Qiu, Shuiyun Ye, Guangzhou Chen, Jiayu Qin, Lu Zhang

Correspondence

zhanglu@scau.edu.cn

In brief

Genomics; Natural sciences; Plant biology; Plant development; Plant evolution; Plant genetics

Ma et al., 2026, iScience *29*, 115480
April 17, 2026
https://doi.org/10.1016/j.isci.2026.115480

CellPress

iScience

CellPress
OPEN ACCESS

Article

Chromosome-level genome of *Madhuca hainanensis* reveals genomic evolution and floral divergence in *Madhuca*

Yifei Ma,[1] Ziyao Wang,[1] Jingzhe Qiu,[1] Shuiyun Ye,[1] Guangzhou Chen,[2] Jiayu Qin,[1] and Lu Zhang[1,3,*]
[1]College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou 510642, China
[2]College of Fisheries and Life Science, Hainan Tropical Ocean University, Sanya 572000, China
[3]Lead contact
*Correspondence: zhanglu@scau.edu.cn
https://doi.org/10.1016/j.isci.2026.115480

SUMMARY

Madhuca hainanensis and *Madhuca pasquieri* are endangered species in China, classified as vulnerable on the IUCN red list. Compared to the fast-growing *Madhuca longifolia*, they exhibit lower fruit set and distinct floral morphology. Here, we present the chromosome-level genome assembly of *M. hainanensis* (760.86 Mb, 12 pseudochromosomes) using PacBio HiFi, Illumina, and Hi-C sequencing. Comparative genomics across floral developmental stages identified six key genes, including *ADO1* and *COL5*, associated with floral divergence. This study provides genomic insights critical for conservation efforts and elucidates molecular mechanisms contributing to the species' endangered status.

INTRODUCTION

The diversity in floral morphology is a result of evolutionary adaptation to different environments and is crucial for plant reproductive success.[1,2] Variations in floral morphology can affect pollinator attraction, pollen dispersal efficiency, and seed production, influencing plant survival and reproduction.[3] Pollen morphology studies are significant for plant species conservation.[4] While studies on model plants like *Arabidopsis thaliana* have revealed many key genes and regulatory networks controlling flower organ development,[5,6] the genetic control of floral morphology remains a complex and unresolved issue in plant biology. This is particularly true for non-model plants, where the genes and regulatory mechanisms affecting flower development are not well understood, limiting our understanding of floral morphology regulation. Recent genomic approaches have provided new pathways to decipher the molecular mechanisms underlying floral organ differentiation.[7] The genus *Madhuca* belongs to the family Sapotaceae and includes approximately 85 species worldwide. In China, the genus *Madhuca* comprises only two species, *Madhuca hainanensis* and *Madhuca pasquieri.* Both species are designated as national second-class protected wild plants and are classified as vulnerable on the IUCN red list. Effective conservation measures are required to prevent their exploitation and ensure their survival. *M. hainanensis* is a kind of precious wood with corrosion resistance in the southern and western parts of Hainan Island, China. It grows at elevations ranging from 70 to 1200 m in mountain evergreen forests and moist valleys. The wood is highly durable, resistant to decay, and shipbuilding, furniture, and more.[8] The seeds contain approximately 50% oil, which is edible and used in soap production. The tree also produces a viscous latex used for adhesive extraction, and its bark has medicinal properties.[9] *M. hainanensis* can flower and fruit normally at approximately 25 years old, demonstrating good resistance to drought, cold, and pests. However, detailed studies on floral morphology and the genome of *M. hainanensis* are lacking. *M. pasquieri* is valued for its wood and oil, with some medicinal uses.[10] It produces a latex that can be used to extract rubber. Our research group has previously analyzed the transcriptome and metabolome of *M. pasquieri* seedlings[11,12] and completed its whole-genome sequencing, providing high-quality data for future studies. *Madhuca longifolia*, native to India and Myanmar, is a fast-growing species with significant economic and medicinal value.[13] Its seeds are used for biodiesel production,[14,15] flowers are edible,[16] and leaves have pharmacological effects.[17,18] Our group has completed the genome sequencing of *M. longifolia*,[19] offering the reference genome for the genus *Madhuca* (GWHDTZT00000000). Our preliminary studies found that the two endangered species exhibited lower fruit set rates and significant morphological divergence in floral characteristics relative to the exotic, fast-growing *M. longifolia*, a species native to India and Myanmar. This study aims to perform whole-genome sequencing of *M. hainanensis* and conduct comparative genomic and transcriptome analyses to identify key genes influencing floral morphology divergence across three *Madhuca* species. This research fills the genomic data gap for *M. hainanensis* and provides insights into the molecular mechanisms of flower development and the conservation of endangered species.

Figure 1. Sampling stages of flora development in three *Madhuca* species, *M. hainanensis*, *M. pasquieri*, and *M. longifolia*
Bud exposure stage (S1), full bloom stage (S2), and late bloom stage (S3), Scale bars, 1 cm.

RESULTS

Comparative morphology of floral in three species of *Madhuca*

The floral morphology of the three *Madhuca* species at three developmental stages used for sampling is shown in Figure 1. The three species have bisexual flowers, generally arranged in axillary umbels and pendulous, with a few species having solitary axillary flowers. The observations (Figure 2A) indicate that the native endangered species *M. hainanensis* and *M. pasquieri* have similar floral morphologies, whereas the exotic fast-growing species *M. longifolia* exhibits distinct differences in floral morphology, particularly in petal arrangement and stamen configuration. *M. pasquieri* and *M. hainanensis* have petals that spread out during full bloom, with stamens arranged in a single whorl. In contrast, the petals of *M. longifolia* remain wrapped around the style during full bloom, with stamens arranged in two whorls. Additionally, the style of *M. longifolia* is significantly thicker than those of the other two species. In terms of flower color, *M. pasquieri* and *M. hainanensis* produce yellowish white or white flowers, while *M. longifolia* exhibits yellow-green flowers. *M. pasquieri* has the smallest flowers among the three species, with *M. hainanensis* and *M. longifolia* having similarly sized flowers. There are also minor differences in the pedicels and sepals among the three species. The flowering periods of the *Madhuca* species vary, with *M. pasquieri* and *M. hainanensis* typically flowering from July to September, whereas *M. longifolia* flowers earlier, from March to May, likely due to its tropical origin.

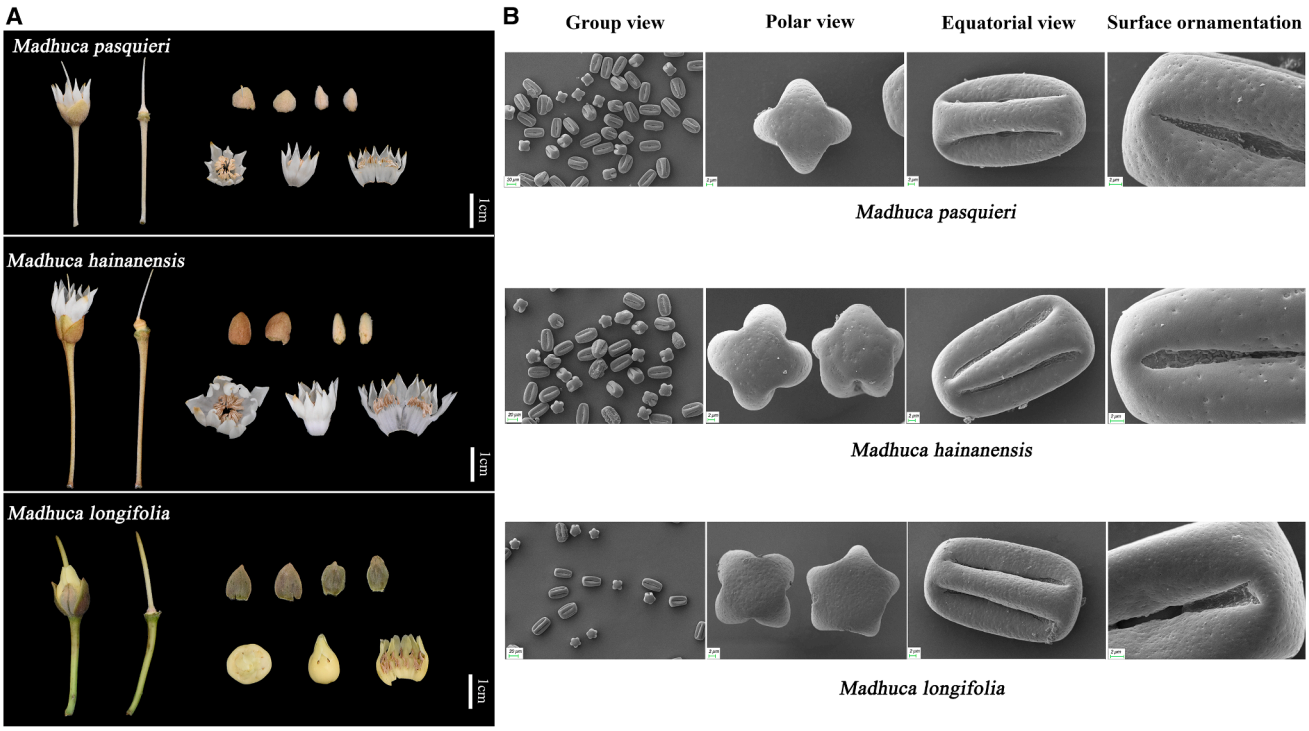


Figure 2. Floral morphology of three *Madhuca* species, *M. hainanensis*, *M. pasquieri*, and *M. longifolia*

(A) Floral morphology dissection of three *Madhuca* species.

(B) Scanning electron microscopy of pollen in three *Madhuca* species. Scale bars: 20 μm (group view); 2 μm (polar view, equatorial view, surface ornamentation).

In addition to comparing the macro-level floral morphology, this study also analyzed the pollen of the three species using scanning electron microscopy (SEM). SEM observations revealed that the pollen of all three *Madhuca* species is monad (Figure 2B). Quantitative SEM measurements revealed clear interspecific differences in pollen size. The pollen grains of *M. hainanensis* were the largest, followed by *M. pasquieri*, whereas those of *M. longifolia* were the smallest. The average polar axis (P) lengths for *M. pasquieri*, *M. hainanensis*, and *M. longifolia* are 25.85 ± 0.67 μm, 28.35 ± 0.94 μm, and 22.60 ± 0.64 μm, respectively. The average equatorial axis (E) lengths are 42.84 ± 0.40 μm, 44.80 ± 1.01 μm, and 40.78 ± 1.03 μm, respectively. The pollen grains of the three species are similar in equatorial view, with rectangular outlines, but show significant differences in polar view. *M. pasquieri* has a star-shaped polar view with four grooves, *M. hainanensis* has a star-shaped polar view with four or five grooves, while *M. longifolia* has a star-shaped polar view with four or five grooves. The pollen surface ornamentation also varies, with *M. pasquieri* and *M. longifolia* having punctate surfaces with perforations, whereas *M. hainanensis* having vermiculate surfaces with perforations. Notably, *M. hainanensis* has more perforations than the other two species. Understanding the pollen morphological characteristics of *Madhuca* species can help reveal their potential pollination mechanisms, providing scientific basis for promoting their reproductive success and the conservation of endangered species.

Genome sequencing and assembly of *M. hainanensis*

Genome assemblies have been generated for all three *Madhuca* species in our research group; in this study, we present *M. hainanensis* as the primary chromosome-level reference, while the previously published *M. longifolia* genome and an in-house *M. pasquieri* assembly are used for comparative analyses.

The genome size of *M. hainanensis* was estimated to be 641.94 Mb based on k-mer analysis (k = 21), indicating a heterozygosity rate of 1.44% (Figure S1). From the Illumina sequencing platform, we obtained 193,224,666 raw reads totaling 28.98 Gb. After quality filtering, the effective data volume was reduced to 28.76 Gb clean reads, with a GC content of 37.10%. Utilizing PacBio HiFi sequencing, we obtained 25.25 Gb of high-quality clean reads, which played a crucial role in assembling the *M. hainanensis* genome into 725 contigs with a total length of 760.01 Mb. The scaffold N50 reached 48.10 Mb, and the largest contig measured 81.27 Mb, with an average GC content of 34.25% (Table S1). The alignment of Illumina clean reads to the preliminary assembled genome resulted in an alignment rate of 96.08% and a genome coverage of 99.58% (Table S2), indicating a high level of consistency between the reads and the assembled genome. Hi-C sequencing added an additional layer of structural insights, providing 67.44 Gb of high-quality filtered clean reads. This enabled the scaffolding of assembled contigs into 12 pseudochromosomes (Figure 3B), with sizes ranging from 45.43 Mb to 84.78 Mb (Table S3). The final assembly of the *M. hainanensis* genome was 760.86 Mb, with a contig N50 of 48.11 Mb and a

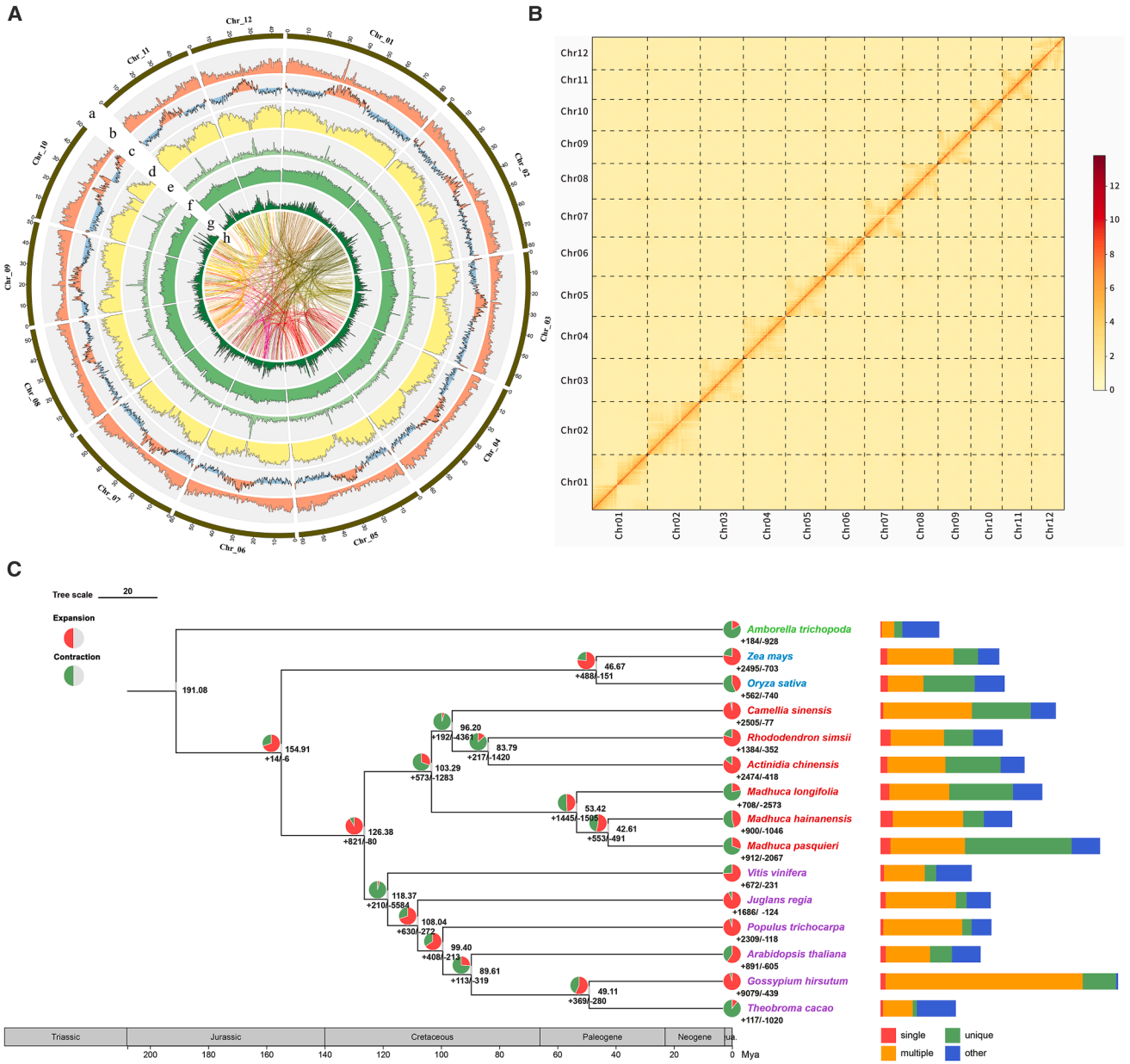


Figure 3. Integrated genomic and evolutionary analyses of three *Madhuca* species, *M. hainanensis*, *M. pasquieri* and *M. longifolia*

(A) The circos diagram of *M. hainanensis*. The circles from outer to inner separately represented: A-a. chromosome length; A-b. Gene density; A-c. GC content distribution; A-d. Transposon distribution density; A-e. LINE distribution density; A-f. distribution density of second-generation reads; A-g. LTR Copia distribution density; A-h. Colinearity of genome blocks.

(B) Genome-wide chromatin interactions of 12 pseudochromosomes.

(C) Phylogenetic tree of three species of *Madhuca* and 12 additional species. The phylogenetic relationships among 15 species are shown. The bar plot represents different types of gene families in each species. Species names are color-coded: green for basal plants, blue for monocots, red for superasterids, and purple for rosids. The numbers below the pie charts on the branches indicate the counts of gene families that underwent expansion/contraction at each niode.

scaffold N50 of 60.09 Mb (Table 1). The maximum contig and scaffold lengths attained were 81.24 Mb and 84.78 Mb, respectively, with an average GC content of 34.29%.(Figure 3A). The completeness of the final genome assembly was assessed using BUSCO (Table S4). The analysis revealed a high level of completeness, with 98.82% of the genes classified as complete. Among these complete genes, 94.55% were present as single copies and 4.28% were duplicated. Only 0.74% of the genes were fragmented, and 0.43% were missing. Genome quality was further evaluated using k-mer based analysis in Merqury and repeat continuity assessment using LTR_retriever. The assembly achieved a high Merqury QV value and high k-mer

Table 1. Genome assembly and annotation results of three species of *Madhuca*

	Madhuca species		
Genomic characteristics	*M. hainanensis*	*M. pasquieri*	*M. longifolia*
Genome Size (Mb)	760.86	736.46	737.92
GC content (%)	34.29	34.03	33.74
scaffold N50 (Mb)	60.09	57.83	60.05
contig N50 (Mb)	48.11	9.00	30.99
Maximum scaffold (Mb)	84.78	84.01	84.26
Chromosome karyotype	$2n = 24$	$2n = 24$	$2n = 24$
BUSCO Assessment (C) (%)	98.82	98.12	98.95
Repeat conten (%)	59.96	68.48	65.90
Number of annotated protein-coding genes	37962	63311	46610
Mean gene Length (bp)	6472.01	4621.82	5561.26
NR database (%)	92.89	93.77	74.93
SwissProt Database (%)	66.56	55.18	51.71
GO Database (%)	72.63	54.69	57.21
KEGG Database (%)	31.85	92.45	73.30
BUSCO Assessment (C) (%)	99.38	93.18	96.47

completeness, indicating high base-level accuracy (Tables S5 and S6). In addition, the whole-genome LAI score exceeded 20, suggesting reference-quality assembly of LTR retrotransposons (Table S7). These results further support the high fidelity and repeat completeness of the *M. hainanensis* genome assembly.

Genome annotation of *M. hainanensis*

Through detailed annotation of repetitive sequences in the genome of *M. hainanensis*, several important findings have been elucidated. Using TRF, 52,067,546 bp of tandem repeat sequences were annotated, constituting 6.84% of the genome. Additionally, RepeatMasker was employed to annotate scattered repeat sequences based on *de novo* annotation results, Repbase, and RepeatProteinMask using its inherent amino acid sequence library. Following integration and deduplication, a total of 456,195,252 bp of repetitive sequences were identified, accounting for 59.96% of the genome (Table S8). Further analysis of these repetitive sequences revealed that DNA transposons accounted for 19.21% of the genome, while LTR retrotransposons comprised 38.96%, making them the predominant transposon type in the genome of *M. hainanensis*, with LINE transposons accounting for 3.46% (Table S9). In terms of non-coding RNA annotation, 321 miRNAs, 4970 rRNAs, 432 snRNAs, and 7536 tRNAs were identified (Table S10). In the annotation of protein-coding gene structures, a total of 37962 protein-encoding genes were identified. Comparative analysis with closely related species revealed that *M. hainanensis* had fewer genes but longer gene lengths and more complex structures (Tables S11 and S12). Using the BUSCO assessment method, the completeness of the encoded genes in *M. hainanensis* reached 99.38% (Table S13). Finally, functional annotation was performed on the gene-coding sequences, with the highest number of annotated genes obtained from the NR database, totaling 35263, accounting for 92.89% of the total number of encoded genes. Additionally, functional annotations were acquired from SwissProt, GO, and KEGG databases (Table 1).

Genome assembly and annotation statistics for the three *Madhuca* species are summarized in Table 1. Comparative analysis revealed that the three species exhibited close similarity in genome size. *M. hainanensis* possessed a slightly larger genome (760.86 Mb), whereas *M. pasquieri* and *M. longifolia* showed relatively smaller genomes (736.46 Mb and 737.92 Mb, respectively). The GC content among the three species was relatively consistent. *M. hainanensis* displayed the highest scaffold N50 and contig N50 values, indicating superior continuity in genome assembly. The longest scaffolds in all three species exceeded 84 Mb, suggesting the acquisition of relatively long contiguous sequences during genome assembly. In terms of annotation results, *M. pasquieri* exhibited the highest proportion of repetitive sequences in its genome, while *M. hainanensis* had the lowest. Discrepancies were observed in the annotation of protein-coding genes, with *M. pasquieri* possessing the highest number and *M. hainanensis* the lowest. Additionally, differences were noted in the annotation proportions across databases (NR, SwissProt, GO, and KEGG), reflecting potential species-specific variations in gene function and metabolic pathways. In summary, while the three species exhibited similarities in genome size and chromosome number, they displayed significant divergence in genome assembly quality, gene annotation, and genome completeness. These distinctions establish the basis for enhanced comprehension of genetic diversity and adaptive traits in these species.

Demographic history of *M. hainanensis*

Based on the demographic history analysis for the species *M. hainanensis* (Figure S2), it was observed that the effective population size (Ne) of *M. hainanensis* maintained a relatively high level until approximately 22 million years ago. Subsequently, there was a significant decline in the effective population size from about 20 million years to 440,000 years ago, eventually nearing extinction. Several notable fluctuations were observed within this decline, indicating possible responses to past environmental changes.

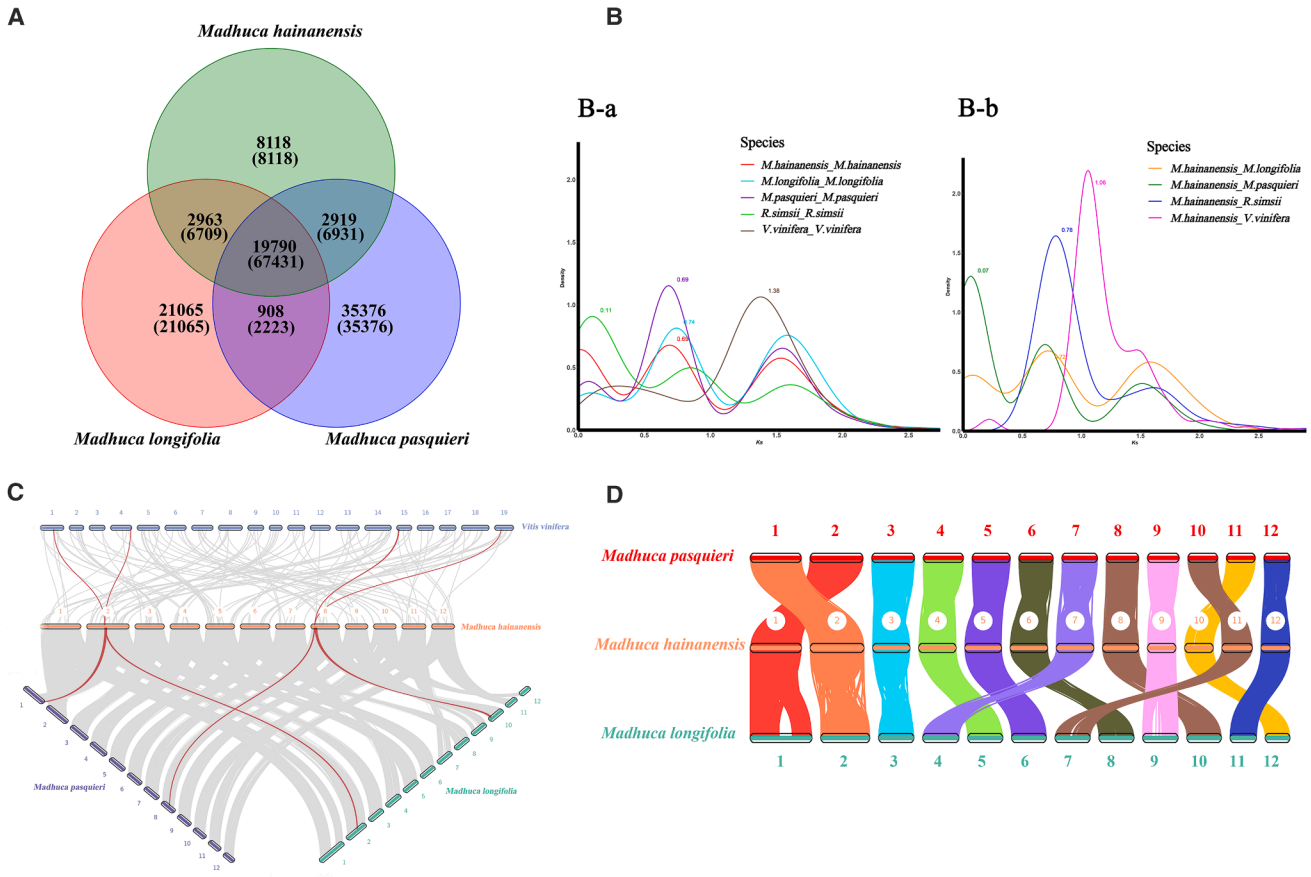


Figure 4. Comparative genomic analysis of *Madhuca hainanensis* with other plants

(A) Venn diagram of unique and common gene families of the three species of *Madhuca*, *M. hainanensis*, *M. pasquieri*, and *M. longifolia*.

(B) Distribution of synonymous substitution rates (Ks) for homologous genomes, shown separately for intraspecific (Figures 4B-a) and interspecific (Figures 4B-b) comparisons.

(C) Synteny analysis between *Vitis vinifera* and three species of *Madhuca*, *M. hainanensis*, *M. pasquieri*, and *M. longifolia*.

(D) Synteny analysis among three species of *Madhuca*, *M. hainanensis*, *M. pasquieri*, and *M. longifolia*.

Genome evolution and comparative genomics

Gene family clustering

To elucidate the evolutionary relationships among the three *Madhuca* species and contrast these phylogenies with 12 representative species, we performed a comparative genomics analysis. Clustering of gene families showed that *M. pasquieri* possesses the largest number of gene families among the 15 species studied (Table S14). Clustering analysis further identified a total of 19,790 shared gene families among the three *Madhuca* species, with species-specific gene families comprising 8,118 for *M. hainanensis*, 35,376 for *M. pasquieri*, and 21,065 for *M. longifolia*, respectively (Figure 4A).

Phylogenetic analysis and species divergence

Based on 42 single-copy gene families identified across the 15 species, a phylogenetic tree was constructed (Figure 3C), revealing close phylogenetic relationships between the *Madhuca* species and other wood plants like *Camellia sinensis*, *Rhododendron simsii*, and *Actinidia chinensis*. The divergence time estimates suggest that *M. longifolia* diverged from *M. pasquieri* and *M. hainanensis* around 53.42 million years ago, while *M. pasquieri* and *M. hainanensis* diverged approximately 42.61 million years ago.

Gene family expansion and contraction

The evolutionary process has driven dynamic the expansion and contraction of gene families in response to environmental pressures, which is closely associated with adaptive evolution among related species. Using CAFÉ, we identified 900, 912, and 708 gene families, respectively, while 1,046, 2,067, and 2,573 in *M. hainanensis*, *M. pasquieri*, and *M. longifolia*, respectively (Figure 3C). GO and KEGG enrichment analyses of the contracted gene families in *M. hainanensis* showed significant enrichment in functions related to external membrane (GO:0019898), transposase activity (GO:0004803), and DNA-mediated transposition (GO:0006313), as well as KEGG pathways such as plant-pathogen interaction (ko04626) and phosphatidylinositol signaling system (ko04070) (Figure S3). In contrast, the expanded gene families in *M. hainanensis* were significantly enriched in photosynthesis- and energy-related functions, including photosynthetic

membrane (GO:0034357), electron transport chain (GO:0022900), and photosynthesis (GO:0015979), while KEGG analysis indicated high enrichment in metabolic pathways including photosynthesis (ko00195) and oxidative phosphorylation (ko00190) (Figure S4). These results indicate that gene family expansion in *M. hainanensis* is closely associated with energy metabolism and photosynthetic processes.

To examine whether these expanded energy-related genes also show expression divergence, we compared their expression levels with their one-to-one orthologs in the closely related species *M. pasquieri* across the three floral developmental stages (S1–S3). At the gene-set level, multiple enriched pathways, including photosynthesis, photosynthetic membrane, electron transport chain, and oxidative phosphorylation, showed stage-dependent expression divergence between the two species (Table S15).

These results suggest that the photosynthesis- and energy metabolism-related gene families expanded in *M. hainanensis* have also undergone divergence in their expression patterns during floral development, relative to their orthologs in *M. pasquieri*. All GO and KEGG terms mentioned previously are significant at FDR <0.05 (Table S16).

Whole-genome duplication and collinearity analysis

Ks distributions indicated shared ancient polyploidy signals among *Madhuca* and related species (Figure 4B). A shared Ks peak at approximately 1.5 suggests a core eudicot shared whole-genome triplication event. Additionally, *M. hainanensis* showed a distinct peak at Ks ≈ 0.69, suggesting a subsequent WGD event following the earlier triplication. The Ks values reflect divergence points relative to other species, consistent with phylogenetic evidence and supporting the hypothesis that the *M. hainanensis* lineage underwent an ancestral whole-genome triplication (WGT) followed by a more recent WGD. The proportion of genomic covariate blocks can be inferred as the number of WGD. Since the genome of *Vitis vinifera* is relatively ancient and has undergone only the WGT shared by core-dicotyledonous plants, we selected *Vitis vinifera* as the reference species and three plants of the genus *Madhuca* for the covariate analysis in this study.

The results showed that there was an obvious 2:1 linear relationship between *Vitis vinifera* and all plants of the genus *Madhuca* (Figure 4C); in addition, the results of the previous study of our group showed that both *Vitis vinifera* and *M. pasquieri* experienced a WGT and a WGD, which is consistent with the results of the Ks analysis of *M. pasquieri*, and further confirming the conclusion that *M. hainanensis* experienced a WGT event and a WGD event. Next, we performed covariance analysis on the genomes of three *Madhuca* species. The results demonstrated that the chromosomes of *M. hainanensis*, *M. pasquieri*, and *M. longifolia* exhibited a high degree of covariance, forming robust 1:1 chromosomal correspondences (Figure 4D), which reflects that the genome structures of the three *Madhuca* species are conserved to a certain degree. Furthermore, the high collinearity confirms the superior assembly quality of the *M. hainanensis* genome. To investigate whether post-WGD retained genes are associated with floral development, we extracted paralogous gene pairs corresponding to the Ks ≈ 0.69 WGD peak and performed GO/KEGG enrichment analyses. Several significantly enriched terms were directly related to floral and reproductive development (e.g., flower development and reproductive shoot system development) and regulatory pathways such as plant hormone signal transduction and MAPK signaling (Table S17).

Importantly, a substantial proportion of these WGD-retained genes showed systematic expression changes across the three floral developmental stages (S1–S3) in *M. hainanensis*, with many genes displaying coordinated increases or stage-specific expression patterns (Table S17), indicating that post-WGD gene retention is accompanied by developmentally regulated transcriptional dynamics during floral development.

Transcriptome analysis revealed key genes affecting floral morphology

The gene expression patterns in *M. pasquieri* and *M. hainanensis* during the Bud exposure stage (S1) showed significant differences compared to those observed during the full bloom stage (S2) and late bloom stage (S3), with the latter two stages exhibiting greater similarity to each other. This correlation between gene expression patterns and floral morphology was also evident. However, the exotic species *M. longifolia* exhibited distinct gene expression patterns during the full bloom stage (S2) and late bloom stage (S3), which might be related to its unique floral morphology as reflected by the markedly different numbers of differentially expressed genes between these stages (Figure 5).

GO enrichment analysis indicated that DEGs in the floral development of all three species were primarily enriched in the biological process and molecular function categories. The DEGs were mainly associated with cell processes (GO: 0009987) and metabolic processes (GO: 0008152) in the biological process category, suggesting that cellular functions and metabolic activities play crucial roles in floral morphogenesis and functionality. In the molecular function category, the DEGs were enriched in catalytic activity (GO: 0003824) and binding (GO: 0005488), indicating the involvement of these genes in various biochemical reactions and molecular interactions essential for floral physiology and development. Additionally, in the cellular component category, DEGs were enriched in anatomical structures (GO: 0110165), reflecting the adjustment and optimization of cellular structures and tissues during floral development. KEGG enrichment results showed that DEGs in the three species during the three developmental stages were significantly enriched in key pathways, including metabolic pathways, biosynthesis of secondary metabolites, plant hormone signal transduction, and starch and sucrose metabolism. These findings suggest that these conserved pathways may play critical regulatory roles in the fundamental biological processes governing floral development. Through trend analysis, comprehensive investigation of differential gene expression patterns across floral developmental stages was conducted in the three species (Figure 6). The results showed specific expression patterns of differentially expressed genes across the species. By integrating phenotypic observations, key expression profiles exhibiting highly correlated with floral developmental stages were selected for GO enrichment analysis. This analysis focused on pathways within the biological process category that are closely implicated in floral

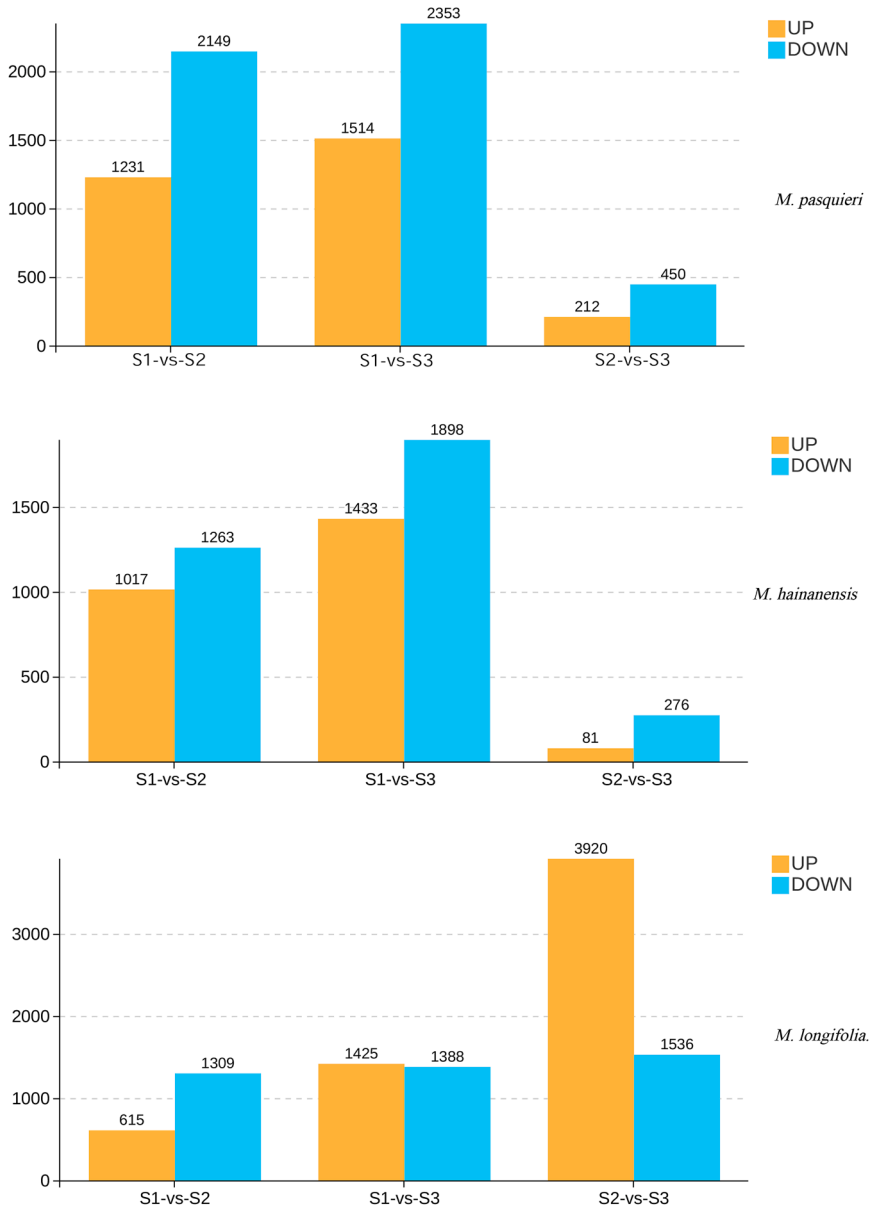


Figure 5. Numbers of differentially expressed genes (DEGs) (up- and downregulated) across three floral developmental stages (S1 vs. S2, S1 vs. S3, and S2 vs. S3) in the three species of *Madhuca*, *M. hainanensis*, *M. pasquieri*, and *M. longifolia*

morphogenesis, identifying key genes associated with floral morphological differences in the three species. Consequently, there are five, five, and nine key genes affecting floral morphology were identified in *M. hainanensis*, *M. pasquieri*, and *M. longifolia*, respectively (Tables S18–S20).

It is noteworthy that the key genes in *M. hainanensis* and *M. pasquieri* were classified into trend analysis profile six (initial upregulation followed by stabilization), whereas the key genes in *M. longifolia* were classified into profile two (initial downregulation followed by upregulation), showing different expression trends from the former two species. After identifying the key genes affecting floral morphogenesis in the three *Madhuca* species, the homologous gene expression patterns were compared across the species to discover the genes influencing floral morphological differences. When homologous genes showed inconsistent expression trends between species, it indicated their potential role in determining floral morphological differences. Based on the heatmap of key gene expression and their homologs in the three species (Figure 7A), genes such as Mha_mikado.Chr_03G2000 (*ADO1*) and Mpa_Mapa01G00022934 (*ADO1*) showed similar trends (initial upregulation followed by stabilization), whereas Mlo_mikado.Chr_03G1348 (*ADO1*) displayed a different trend (initial upregulation followed by downregulation). Similarly, homologous genes like Mha_mikado.Chr_02G1189 (*COL5*), Mpa_Mapa01G00058724 (*COL5*), and Mlo_mikado.Chr_02G3738 (*COL5*) showed differing expression patterns across the species. Overall, these genes exhibited similar expression patterns in *M. hainanensis* and *M. pasquier*, but displayed distinct patterns in *M. longifolia*, which correlated with the observed phenotypic differences in floral characteristics. This finding indicates that these genes constitute key regulators underlying the divergent floral morphologies between *M. longifolia* and the other two *Madhuca* species. To validate these findings, RT-qPCR was performed on two groups of six genes exhibiting similar expression patterns in *M. pasquieri* and *M. hainanensis*, but divergent expression patterns in *M. longifolia*. The RT-qPCR results (Figure 7B) aligned with the RNA-seq data, thereby corroborating the reliability of the transcriptome sequencing analysis.

DISCUSSION

Floral morphology variations and reproductive implications in *Madhuca* species

The variations in floral morphology among three species of the genus *Madhuca* may influence their reproductive strategies. In *M. longifolia*, the flowers do not fully open during peak blooming periods, and the stamens arranged in two whorls. This floral configuration serves to shield pollen and nectar within the corolla from adverse meteorological conditions, thereby attracting more pollinators and enhancing pollination

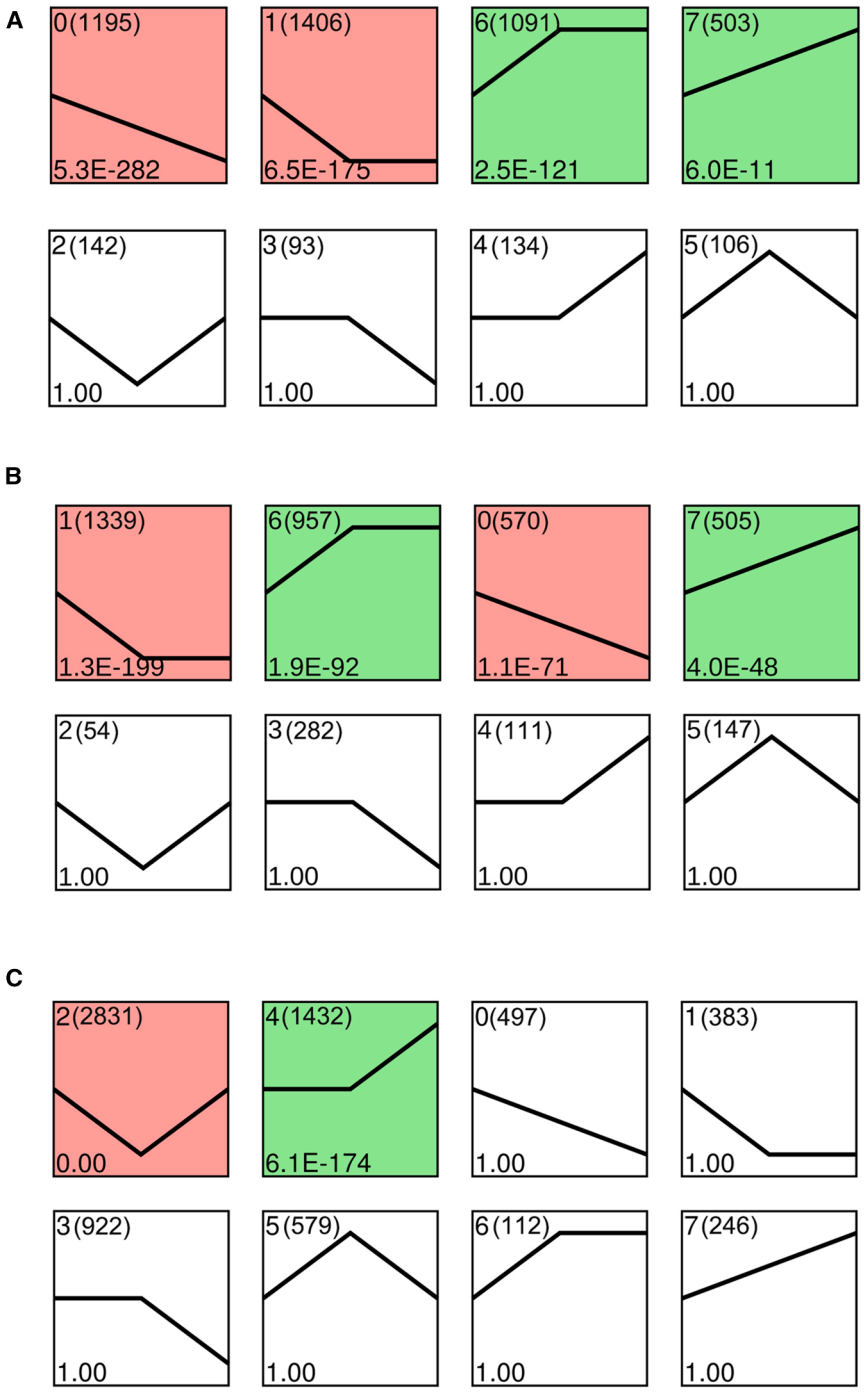


Figure 6. Trend analysis of DEGs across three floral developmental stages in the three species of *Madhuca*, *M. hainanensis*, *M. pasquieri*, and *M. longifolia*

(A) Trend profiles for *M. pasquieri*.

(B) Trend profiles for *M. hainanensis*.

(C) Trend profiles for *M. longifolia.*

Note: For each profile, the bottom-left number indicates the *p*-value, the top-left number is the profile ID, the number in parentheses is the gene count assigned to that profile, and colored profiles denote significance ($p < 0.05$).

efficiency. In contrast, the flowers of *M. hainanensis* and *M. pasquieri* exhibit laterally expanded petals and pedicels that bend downwards during full anthesis. This morphology renders pollen more vulnerable to meteorological conditions such as wind and precipitation, which can reduce pollination efficiency. Additionally, their flowering phenology frequently coincides with the rainy season, which may negatively affect pollen viability and hinder successful sexual reproduction, leading to reduced seed set and constrain population recruitment.

Pollen morphology variations not only reflect genetic diversity among species but also correlate with their reproductive strategies.[20] The number and shape of germination pores can affect pollen germination efficiency and pollen tube growth, thereby influencing reproductive success rates.[21] *M. hainanensis* and *M. pasquieri* typically have fewer germination pores than *M. longifolia*, potentially lowering their reproductive success rates. Phylogenetic analysis indicates that *M. longifolia* diverged first among the three species, followed by *M. hainanensis* and *M. pasquieri*. The pentagonal shape of the pollen in *M. longifolia* is likely a primitive trait, while the tetragonal pollen shape observed in *M.hainanensis* and *M. pasquieri. M. hainanensis* may be an adaptation to specific environmental conditions during their evolutionary process. The SEM comparison of pollen among these species highlights differences that not only reflect genetic diversity and ecological adaptability,[22] but also provide important scientific bases for the conservation of endangered species.

Genomic insights from *M. hainanensis*

This study generated a high-quality genome assembly for *M. hainanensis* utilizing Illumina, PacBio HiFi, and Hi-C technologies, establishing a new reference genome for the genus *Madhuca*. Comparative genomic analysis revealed that the genome size of *M. hainanensis* (760.86 Mb) is slightly larger than that of *M. pasquieri* (736.46 Mb) and *M. longifolia* (737.92 Mb). The contig N50 of *M. hainanensis* (48.11 Mb) is higher than that of *M. pasquieri* (9.00 Mb) and *M. longifolia* (30.99 Mb), indicating higher assembly continuity. Repetitive sequences, particularly transposable elements, significantly influence genome size and

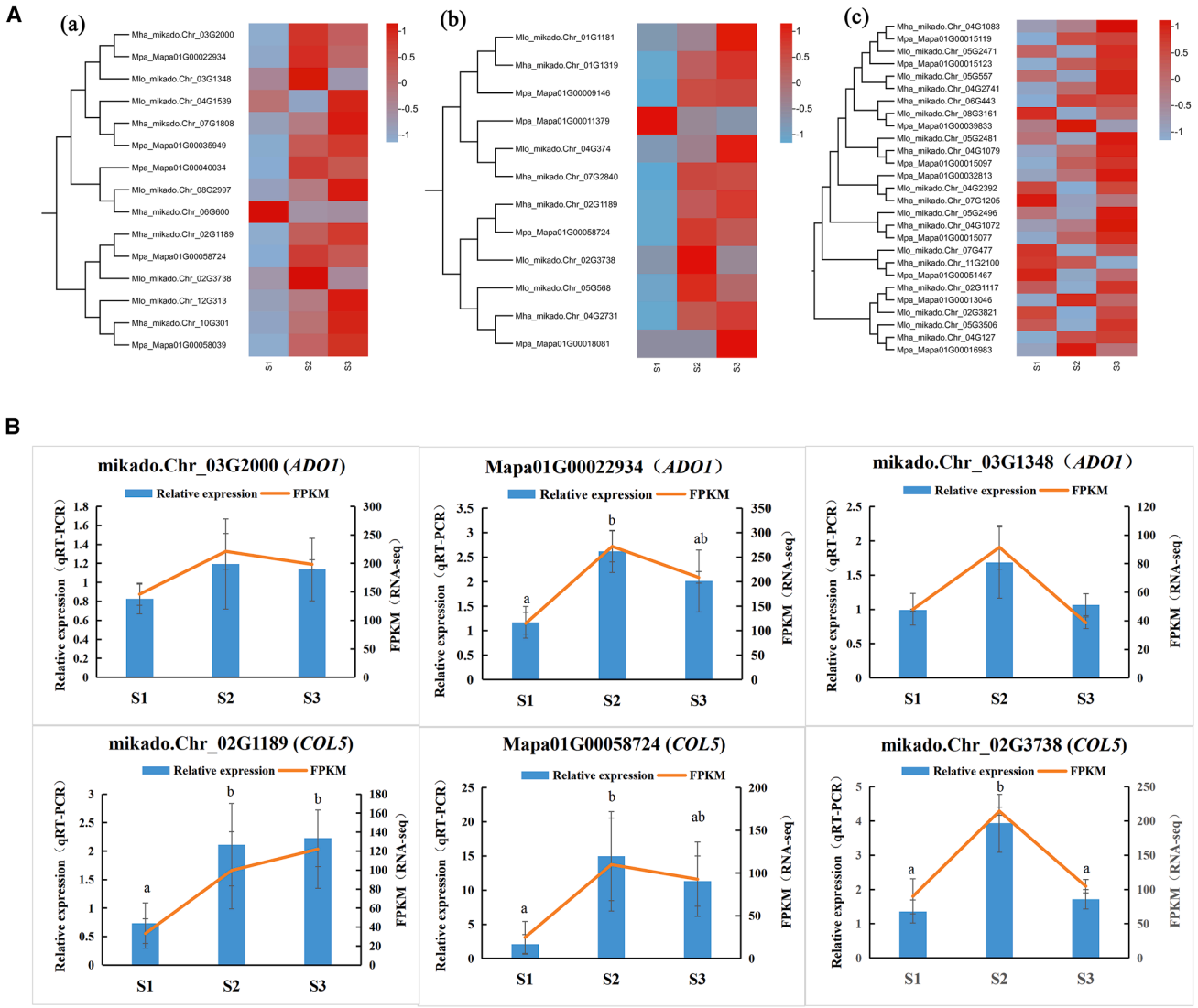


Figure 7. Expression validation of floral morphology-associated genes in *Madhuca* species

(A) Heatmap of key genes influencing floral morphology and their homologous genes in three species of *Madhuca*, *M. hainanensis*, *M. pasquieri*, and *M. longifolia*.
(B) RT-qPCR validation of six key genes influencing floral morphology in three species of *Madhuca*.
Data are represented as mean ± SEM. Different letters indicate statistically significant differences, ($p < 0.05$, one-way ANOVA followed by Tukey's post-hoc test).

complexity. The proportions of repetitive sequences in *M. pasquieri*, *M. hainanensis*, and *M. longifolia* are 68.48%, 59.96%, and 65.90%, respectively. Notably, *M. hainanensis* and *M. pasquieri* have a higher proportion of LTR retrotransposons, while *M. longifolia* is dominated by DNA transposons, especially Helitron types, which account for 62.31% of its genome. These differences may be related to the species-specific evolutionary history and ecological adaptability. During evolution, species may have experienced different transposon expansion events, which are influenced by species-specific ecological pressures and natural selection.[23] Gene family and evolutionary analyses reveal that *M. pasquieri* has the greatest counts of unique paralogous genes (30,744) among the 15 compared species, and the highest number of protein-coding genes (63,311) and unique gene families (35,376). Furthermore, this species exhibits the most significantly expanded (912) and contracted (2,067) gene families among the three analyzed species.

The GO enrichment analysis of shared and unique gene families obtained from clustering analysis of the three species revealed that the unique gene families in *M. hainanensis* and *M. pasquieri* were enriched in flower-related GO terms, such as flower development (GO:0009908), regulation of floral development (GO:0009909), floral whorl development (GO:0048438), floral organ development (GO:0048437), and floral organ morphogenesis (GO:0048444). However, GO enrichment analysis of the unique gene families in *M. longifolia* did not demonstrate significant enrichment in these floral-related pathways. This finding suggests that

during the evolutionary trajectory of *M. longifolia*, functional divergence within these gene families may have occurred, potentially attributable to differential ecological adaptations and reproductive strategies associated with its rapid growth as an exotic species. PSMC analysis indicates a significant decline in the effective population size of *M. hainanensis* between approximately 2.0×10^{-7} to 4.4×10^{-5} years ago, corresponding to major geological and climatic events. The uplift of the Qinghai-Tibet Plateau around 8 million years ago, which initiated the South Asian monsoon and contributed to inland aridification,[24,25] likely played a pivotal role. Additional contributing factors include the increase in loess mass accumulation rate in China around 3 million years ago,[26] and the onset of the Xixiabangma Glaciation from 0.8 to 1.2 million years ago,[27] both of which may have exacerbated the long-term population decline and shaped the current endangered status of the species.

Key genes influencing floral morphology

During the identification of key genes influencing floral morphology, the homologous gene Mlo_mikado.Chr_04G1539 (*MYB62*) was found to be highly expressed in *M. longifolia*, potentially contributing to its unique floral development. *MYB62* is known to function in responses to nutrient deficiency, gibberellin metabolism,[28] flowering time regulation,[29] and anthocyan in biosynthesis,[30,31] highlighting its multifaceted significance in plant growth and adaptation. Similarly, the gene Mha_mikado.Chr_06G600 (*AGL12*) shows high expression in *M. hainanensis*, correlating with its specific floral morphology. *AGL12*, a member of the MIKC-type MADS-box subfamily, plays a pivotal role in root meristem cell proliferation and flowering transition.[32] Heterologous expression of the *AGL12* gene in *Vitis vinifera*, accelerates flowering initiation, enhances vegetative growth, and increased yield.[33]

Comparative transcriptome analyses across *Hypericum* species in the Changbai Mountains found that delayed flowering phenology correlates with higher expression levels of *SVP* and *AGL12*, and lower expression levels of *FUL*.[34] In another group of homologous genes, Mha_mikado.Chr_02G1189 (*COL5*) and Mpa_Mapa01G00058724 (*COL5*) exhibit congruent expression patterns, whereas Mlo_mikado.Chr_02G3738 (*COL5*) displays divergent expression trend, which may be related to the unique floral development pattern of *M. longifolia*. Empirical evidence indicates that *COL5* overexpression induces precocious flowering in *Arabidopsis thaliana* under short-day photoperiods, thus demonstrating its function as a key regulator of flowering time.[35] After identifying the key genes *ADO1* and *COL5* that influencing floral morphology and validating their expression levels via RT-qPCR, further molecular biology experiments are needed to elucidate the specific roles in floral morphology development. In conclusion, these genes play diverse roles in regulating floral morphology, responding to environmental signals, and adaptive evolution, thereby providing novel insights into plant adaptive mechanisms and endangered species conservation. Future population-level resequencing across multiple individuals will further enable investigation of genetic diversity, population structure, and adaptive evolution in *Madhuca* species.

Limitations of the study

This study provides a chromosome-level reference genome for *M. hainanensis* and comparative insights into floral divergence, but several limitations should be noted. First, population resequencing data were not included due to the protected status of *M. hainanensis*, limiting analyses of genetic diversity and selection signatures. Second, only floral developmental stages were analyzed for transcriptomics, and other tissues (e.g., leaves, seeds) could provide additional functional context. Third, the functional validation of key floral genes (e.g., *ADO1*, *COL5*) was not performed, which would strengthen their causal link to morphological traits. Future work addressing these aspects will further advance our understanding of *Madhuca* evolution and conservation.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Lu Zhang, (zhanglu@scau.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All genome assembly data generated in this study have been deposited in the Genome Warehouse (GWH) of the National Genomics Data Center (NGDC,https://ngdc.cncb.ac.cn), China National Center for Bioinformation (CNCB), under the following accession numbers: GWHHKZZ00000000.1 (*M. hainanensis*), GWHDTYC00000000 (*M. pasquieri*), and GWHDTZT00000000 (*M. longifolia*). These datasets are publicly accessible without embargo.
- Transcriptome sequencing (RNA-seq) datasets of *M. hainanensis*, *M. pasquieri*, and *M. longifolia* have been deposited in the Genome Sequence Archive (GSA, https://ngdc.cncb.ac.cn/gsa/) of the NGDC (CNCB) under the BIG Sub-GSA accession number CRA038128 (BioProject ID: PRJCA057057). All transcriptome data are publicly available with no access restrictions.
- Other data: All other data supporting the findings of this study are available within the article and its supplemental information.
- Code: This study did not generate new custom code.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (32371742) and the Wildlife Conservation and Management Projects of the Guangdong Forestry Bureau in 2024 and 2025. We are grateful to thank Guangzhou Genedenovo Biotechnology Co., Ltd for assisting in sequencing analysis.

AUTHOR CONTRIBUTIONS

Writing – original draft, visualization, methodology, investigation, formal analysis, data curation, Y.M.; writing – review and editing, software, formal analysis, data curation, Z.W.; **:** writing – review and editing, validation, data curation, J.Q.; **:** writing – original draft, visualization, methodology, S.Y.; writing – review and editing, investigation, G.C.; writing – review and editing, investigation, J.Q.; conceptualization, writing – review and editing, resources, supervision, project administration, funding acquisition, L.Z.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

STAR★METHODS

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

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Plant materials
- METHOD DETAILS
  - Floral morphology and pollen scanning electron microscope
  - Genomic sequencing, RNA-Seq and Hi-C
  - Genome assembly and annotation
  - Genome annotation
  - Demographic history of M. hainanensis
  - Comparative phylogenomics and whole genome duplication analysis
  - RNA sequencing and gene expression analysis
  - qRT-PCR analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Differential expression analysis
  - qRT-PCR data analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2026.115480.

Received: September 25, 2025
Revised: November 27, 2025
Accepted: March 23, 2026
Published: March 27, 2026

REFERENCES

1. Brophy, J.A.N., Larue, T., and Dinneny, J.R. (2018). Understanding and engineering plant form. Semin. Cell Dev. Biol. *79*, 68–77. https://doi.org/10.1016/j.semcdb.2017.08.051.
2. Gehring, M. (2019). Epigenetic dynamics during flowering plant reproduction: evidence for reprogramming? New Phytol. *224*, 91–96. https://doi.org/10.1111/nph.15856.
3. Mander, L., Julier, A.C.M., and Jaramillo, C. (2025). The architecture of floral diversity: a bipartite network of flower morphology in the neotropical rainforest of barro colorado island, panama. Bot. J. Linn. Soc. *210*, 1–11. https://doi.org/10.1093/botlinnean/boaf032.
4. Leal, L.C., and Koski, M.H. (2024). Linking pollen limitation and seed dispersal effectiveness. Ecol. Lett. *27*, e14347. https://doi.org/10.1111/ele.14347.
5. Quiroz, S., Yustis, J.C., Chávez-Hernández, E.C., Martínez, T., Sanchez, M.D.L.P., Garay-Arroyo, A., álvarez-Buylla, E.R., and García-Ponce, B. (2021). Beyond the genetic pathways, flowering regulation complexity in *Arabidopsis thaliana*. Int. J. Mol. Sci. *22*, 5716. https://doi.org/10.3390/ijms22115716.
6. Hawar, A., Chen, W., Zhu, T., Wang, X., Liu, J., Xiong, S., Ito, T., Chen, D., and Sun, B. (2025). The histone acetyltransferase GCN5 regulates floral meristem activity and flower development in Arabidopsis. Plant Cell *37*, koaf135. https://doi.org/10.1093/plcell/koaf135.
7. Yang, J., Chen, J., He, X., Wang, G., Barrett, S.C.H., and Li, Z. (2025). The *Monochoria* genome provides insights into the molecular mechanisms underlying floral heteranthery. J. Genet. Genomics *52*, 826–838. https://doi.org/10.1016/j.jgg.2025.02.008.
8. Dai, Z.C., Si, C.C., Zhai, D.L., Huang, P., Qi, S.S., Zhong, Q.X., Hu, X., Li, H.M., and Du, D.L. (2013). Human impacts on genetic diversity and differentiation in six natural populations of *Madhuca hainanensis*, an endemic and endangered timber species in China. Biochem. Systemat. Ecol. *50*, 212–219. https://doi.org/10.1016/j.bse.2013.01.008.
9. Wang, Y., Wang, H.T., Chen, Y.K., Yu, J., and Yang, Y. (2021). The complete chloroplast genome of *Madhuca hainanensis* (Sapotaceae), an endemic and endangered timber species in Hainan Island, China. Mitochondrial DNA. B Resour. *6*, 755–757. https://doi.org/10.1080/23802359.2021.1878954.
10. Hoang, L.S., Tran, M.H., Lee, J.S., To, D.C., Nguyen, V.T., Kim, J.A., Lee, J.H., Woo, M.H., and Min, B.S. (2015). Anti-inflammatory activity of pyrrolizidine alkaloids from the leaves of *Madhuca pasquieri* (Dubard). Chem. Pharm. Bull. (Tokyo) *63*, 481–484. https://doi.org/10.1248/cpb.c14-00855.
11. Kan, L., Liao, Q., Chen, Z., Wang, S., Ma, Y., Su, Z., and Zhang, L. (2021). Dynamic transcriptomic and metabolomic analyses of *Madhuca pasquieri* (Dubard) H. J. Lam during the post-germination stages. Front. Plant Sci. *12*, 731203. https://doi.org/10.3389/fpls.2021.731203.
12. Kan, L., Liao, Q., Su, Z., Tan, Y., Wang, S., and Zhang, L. (2020). Single-molecule real-time sequencing of the *Madhuca pasquieri* (Dubard) Lam. Transcriptome reveals the diversity of full-length transcripts. Forests *11*, 866. https://doi.org/10.3390/f11080866.
13. Awasthi, Y.C., Bhatnagar, S.C., and Mitra, C.R. (1975). Chemurgy of sapotaceous plants: *Madhuca* species of India. Econ. Bot. *29*, 380–389. https://doi.org/10.1007/BF02862185.
14. Mani, Y., Devaraj, T., Devaraj, K., Abdurrawoof, S.A., and Subramanian, S. (2020). Experimental investigation of biodiesel production from *Madhuca longifolia* seed through in situ transesterification and its kinetics and thermodynamic studies. Environ. Sci. Pollut. Res. Int. *27*, 36450–36462. https://doi.org/10.1007/s11356-020-09626-y.
15. Khan, I.U. (2024). Biodiesel production and selected fuel qualities from prospective non-edible oils: *Hevea brasiliensis*, *Madhuca longifolia*, *Azadirachta indica*, and *Gossypium hirsutum*. Int. J. Green Energy *21*, 3054–3071. https://doi.org/10.1080/15435075.2024.2349194.
16. Pinakin, D.J., Kumar, V., Suri, S., Sharma, R., and Kaushal, M. (2020). Nutraceutical potential of tree flowers: a comprehensive review on biochemical profile, health benefits, and utilization. Food Res. Int. *127*, 108724. https://doi.org/10.1016/j.foodres.2019.108724.
17. Simon, J.P., and Evan Prince, S. (2021). Ameliorative activity of aqueous leaf extract from *Madhuca longifolia* against diclofenac-administered toxicity on rat stomach and intestine. J. Histotechnol. *44*, 114–126. https://doi.org/10.1080/01478885.2020.1861917.
18. Jha, D., Prajapati, S.K., Deb, P.K., Jaiswal, M., and Mazumder, P.M. (2024). *Madhuca longifolia*-hydro-ethanolic-fraction reverses mitochondrial dysfunction and modulates selective glut expression in diabetic mice fed with high fat diet. Mol. Biol. Rep. *51*, 209. https://doi.org/10.1007/s11033-023-08962-9.
19. Wang, S., Lin, H., Ye, S., Jiao, Z., Chen, Z., Ma, Y., and Zhang, L. (2024). High-quality chromosome-level genomic insights into molecular adaptation to low-temperature stress in *Madhuca longifolia* in southern subtropical China. BMC Genom. *25*, 877. https://doi.org/10.1186/s12864-024-10769-2.
20. Naghiloo, S., and Nikzat Siahkolaee, S. (2019). Does breeding system affect pollen morphology? A case study in Zygophylloideae (Zygophyllaceae). Plant Reprod. *32*, 381–390. https://doi.org/10.1007/s00497-019-00379-4.
21. de Oliveira Souza, S., Souza de Oliveira, R., Aona, L.Y.S., Souza, F.V.D., Soares, T.L., Rossi, M.L., and de Souza, E.H. (2021). Pollen morphology and viability of *Tillandsia* (Bromeliaceae) species by light microscopy and scanning electron microscopy. Microsc. Res. Tech. *84*, 441–459. https://doi.org/10.1002/jemt.23601.
22. Kadluczka, D., Sliwinska, E., and Grzebelus, E. (2022). Combining genome size and pollen morphology data to study species relationships in the genus *Daucus* (Apiaceae). BMC Plant Biol. *22*, 382. https://doi.org/10.1186/s12870-022-03743-1.
23. Bozan, I., Achakkagari, S.R., Anglin, N.L., Ellis, D., Tai, H.H., and Strömvik, M.V. (2023). Pangenome analyses reveal impact of transposable elements and ploidy on the evolution of potato species. Proc. Natl. Acad. Sci. USA *120*, e2211117120. https://doi.org/10.1073/pnas.2211117120.

24. Li, J., Fang, X., Song, C., Pan, B., Ma, Y., and Yan, M. (2014). Late Miocene–Quaternary rapid stepwise uplift of the NE Tibetan Plateau and its effects on climatic and environmental changes. Quat. Res. *81*, 400–423. https://doi.org/10.1016/j.yqres.2014.01.002.

25. Zheng, B., Xu, Q., and Shen, Y. (2002). The relationship between climate change and Quaternary glacial cycles on the Qinghai-Tibetan Plateau: review and speculation. Quat. Int. *97–98*, 93–101. https://doi.org/10.1016/S1040-6182(02)00054-X.

26. Sun, Y., and An, Z. (2005). Late pliocene-pleistocene changes in mass accumulation rates of eolian deposits on the central Chinese loess plateau. J. Geophys. Res. *110*. https://doi.org/10.1029/2005JD006064.

27. Zheng, B. (1988). Quaternary glaciation of mt. Qomolangma-xixabangma region. Geojournal *17*, 525–543. https://doi.org/10.1007/BF00209441.

28. Devaiah, B.N., Madhuvanthi, R., Karthikeyan, A.S., and Raghothama, K.G. (2009). Phosphate starvation responses and gibberellic acid biosynthesis are regulated by the *MYB62* transcription factor in *Arabidopsis*. Mol. Plant *2*, 43–58. https://doi.org/10.1093/mp/ssn081.

29. Qi, X., Tang, W., Li, W., He, Z., Xu, W., Fan, Z., Zhou, Y., Wang, C., Xu, Z., Chen, J., et al. (2021). *Arabidopsis* G-Protein β subunit AGB1 negatively regulates DNA binding of MYB62, a suppressor in the Gibberellin pathway. Int. J. Mol. Sci. *22*, 8270. https://doi.org/10.3390/ijms22158270.

30. Meng, J., Wang, H., Chi, R., Qiao, Y., Wei, J., Zhang, Y., Han, M., Wang, Y., and Li, H. (2023). The eTM–miR858–*MYB62*-like module regulates anthocyanin biosynthesis under low-nitrogen conditions in *Malus spectabilis*. New Phytol. *238*, 2524–2544. https://doi.org/10.1111/nph.18894.

31. Su, L., Zhang, M., Zhang, Y., Chen, Y., Yang, L., Wang, Y., Song, Y., and Gong, L. (2023). Transcriptome analysis reveals the crucial function of hyperoside in inhibiting anthocyanin accumulation in grape (*Vitis vinifera* L.) fruits by inducing *VvMYB62*. Front. Plant Sci. *14*, 1119749. https://doi.org/10.3389/fpls.2023.1119749.

32. Tapia-López, R., García-Ponce, B., Dubrovsky, J.G., Garay-Arroyo, A., Pérez-Ruíz, R.V., Kim, S.H., Acevedo, F., Pelaz, S., and Alvarez-Buylla, E.R. (2008). An *AGAMOUS*-Related MADS-box gene, *XAL1* (*AGL12*), regulates root meristem cell proliferation and flowering transition in *Arabidopsis*. Plant Physiol. *146*, 1182–1192. https://doi.org/10.1104/pp.107.108647.

33. Mao, T., Wang, X., Gao, H., Gong, Z., Liu, R., Jiang, N., Zhang, Y., Zhang, H., Guo, X., and Yu, C. (2023). Ectopic expression of MADS-box transcription factor *VvAGL12* from grape promotes early flowering, plant growth, and production by regulating cell-wall architecture in *Arabidopsis*. Genes *14*, 2078. https://doi.org/10.3390/genes14112078.

34. Yunrui, X., Rui, S., Xing, Y., Zhe, Z., Keqin, Z., and Nanyi, Z. (2023). Comparative transcriptomic analysis reveals differences in *MADS-box* genes of different hypericum in Changbai Mountains. Ecol. Evol. *13*, e10196. https://doi.org/10.1002/ece3.10196.

35. Hassidim, M., Harir, Y., Yakir, E., Kron, I., and Green, R.M. (2009). Over-expression of *CONSTANS-LIKE 5* can induce flowering in short-day grown *Arabidopsis*. Planta *230*, 481–491. https://doi.org/10.1007/s00425-009-0958-7.

36. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, 884–890. https://doi.org/10.1093/bioinformatics/bty560.

37. Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764–770. https://doi.org/10.1093/bioinformatics/btr011.

38. Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). Genomescope: fast reference-free genome profiling from short reads. Bioinformatics *33*, 2202–2204. https://doi.org/10.1093/bioinformatics/btx153.

39. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods *18*, 170–175. https://doi.org/10.1038/s41592-020-01056-5.

40. Utturkar, S.M., Klingeman, D.M., Jr., Brown, S.D., and Hurt, R.A., Jr. (2017). A case study into microbial genome assembly gap sequences and finishing strategies. Front. Microbiol. *8*, 1272. https://doi.org/10.3389/fmicb.2017.01272.

41. Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., and Shendure, J. (2013). Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. Nat. Biotechnol. *31*, 1119–1125. https://doi.org/10.1038/nbt.2727.

42. 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. https://doi.org/10.1038/s41477-019-0487-8.

43. 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. https://doi.org/10.1126/science.aal3327.

44. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

45. Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics *23*, 1061–1067. https://doi.org/10.1093/bioinformatics/btm071.

46. 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. https://doi.org/10.1093/bioinformatics/btv351.

47. Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res *35*, 3100–3108. https://doi.org/10.1093/nar/gkm160.

48. Chan, P.P., and Lowe, T.M. (2019). tRNAscan-SE: Searching for tRNA genes in genomic sequences. Methods Mol. Biol. *1962*, 1–14. https://doi.org/10.1007/978-1-4939-9173-0_1.

49. Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., and Petrov, A.I. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. *46*, D335–D342. https://doi.org/10.1093/nar/gkx1038.

50. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580.

51. Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. *35*, W265–W268. https://doi.org/10.1093/nar/gkm286.

52. Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. BMC Bioinf. *9*, 18. https://doi.org/10.1186/1471-2105-9-18.

53. Yan, H., Bombarely, A., Li, S., and Valencia, A. (2020). DeepTE: a computational method for *de novo* classification of transposons with convolutional neural network. Bioinformatics *36*, 4269–4275. https://doi.org/10.1093/bioinformatics/btaa519.

54. Xiong, W., He, L., Lai, J., Dooner, H.K., and Du, C. (2014). HelitronScanner uncovers a large overlooked cache of *Helitron* transposons in many plant genomes. Proc. Natl. Acad. Sci. USA *111*, 10263–10268. https://doi.org/10.1073/pnas.1410068111.

55. Tarailo-Graovac, M., and Chen, N. (2009). Using repeatmasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics *4*, 4–10. https://doi.org/10.1002/0471250953.bi0410s25.

56. Tempel, S. (2012). Using and understanding repeatmasker (Humana Press), pp. 29–51.

57. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360. https://doi.org/10.1038/nmeth.3317.

58. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/nbt.3122.

59. Li, H. (2023). Protein-to-genome alignment with miniprot. Bioinformatics *39*, btad014–btad016. https://doi.org/10.1093/bioinformatics/btad014.

60. Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: *ab* initio prediction of alternative transcripts. Nucleic Acids Res. *34*, W435–W439. https://doi.org/10.1093/nar/gkl200.

61. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., Wortman, J.R., and Lawrence Livermore National Laboratory Llnl, L.C.U.S. (2008). Automated eukaryotic gene structure annotation using EVidencemodeler and the program to assemble spliced alignments. Genome Biol. *9*, R7. https://doi.org/10.1186/gb-2008-9-1-r7.

62. Venturini, L., Caim, S., Kaithakottil, G.G., Mapleson, D.L., and Swarbreck, D. (2018). Leveraging multiple transcriptome assembly methods for improved gene structure annotation. GigaScience *7*, 1–15. https://doi.org/10.1093/gigascience/giy093.

63. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60. https://doi.org/10.1038/nmeth.3176.

64. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645. https://doi.org/10.1101/gr.092759.109.

65. 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. https://doi.org/10.1093/molbev/msu300.

66. Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Bioinformatics *13*, 555–556. https://doi.org/10.1093/bioinformatics/13.5.555.

67. Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178–2189. https://doi.org/10.1101/gr.1224503.

68. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340.

69. De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics *22*, 1269–1271. https://doi.org/10.1093/bioinformatics/btl097.

70. Zhang, Z., Li, J., Zhao, X.Q., Wang, J., Wong, G.K.S., and Yu, J. (2006). KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genom. Proteom. Bioinform. *4*, 259–263. https://doi.org/10.1016/S1672-0229(07)60007-2.

71. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. BMC Bioinf. *12*, 323. https://doi.org/10.1186/1471-2105-12-323.

72. Ernst, J., and Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. BMC Bioinf. *7*, 191. https://doi.org/10.1186/1471-2105-7-191.

73. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. *35*, 1547–1549. https://doi.org/10.1093/molbev/msy096.

74. 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. https://doi.org/10.1186/s13059-014-0550-8.

75. Qiu, Z., Zhao, Y., Qian, Y., Tan, C., Tang, C., and Chen, X. (2023). Seed and pollen morphology in nine native Chinese species of *Sorbus* (Rosaceae). Int. J. Fruit Sci. *23*, 87–101. https://doi.org/10.1080/15538362.2023.2212057.

76. Talip, N., Ayob, S.N.A., Kader Maideen, H.M., Bunawan, H., and Abdul Rahman, M.R. (2021). Kesignifikan sistematik ciri anatomi pina dan stip angiopteris angustifolia c. Presl dan A. evecta (G. Forst.) Hoffm. [Marattiaceae] di semenanjung Malaysia. Sains Malays. *50*, 1243–1254. https://doi.org/10.17576/jsm-2021-5005-05.

77. Feng, C., Dai, M., Liu, Y., and Chen, M. (2021). Sequence repetitiveness quantification and *de novo* repeat detection by weighted k-mer coverage. Briefings Bioinf. *22*, bbaa086. https://doi.org/10.1092/bib/bbaa086.

78. Korf, I. (2004). Gene finding in novel genomes. BMC Bioinf. *5*, 59. https://doi.org/10.1186/1471-2105-5-59.

79. Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol. Biol. Evol. *38*, 5825–5829. https://doi.org/10.1093/molbev/msab293.

80. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. *47*, D309–D314. https://doi.org/10.1093/nar/gky1085.

81. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496. https://doi.org/10.1038/nature10231.

82. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B, Methodological *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Genome assembly data of *Madhuca hainanensis*	This paper	NGDC: GWHHKZZ00000000.1
Genome assembly data of *Madhuca pasquieri*	This paper	NGDC: GWHDTYC00000000
Genome assembly data of *Madhuca longifolia*	Wang et al.[19]	NGDC: GWHDTZT00000000
RNA-seq data of *Madhuca hainanensis, Madhuca pasquieri and Madhuca longifolia*	This paper	NGDC: PRJCA057057
Software and algorithms		
fastp (v0.18.0)	Chen et al.[36]	https://github.com/OpenGene/fastp
jellyfish (v2.2.6)	Marcais and Kingsford[37]	https://github.com/gmarcais/Jellyfish
GenomeScope (v1.0.0)	Vurture et al.[38]	https://github.com/schatzlab/genomescope
Hifiasm (v0.15.1-r334)	Cheng et al.[39]	https://github.com/chhylp123/hifiasm
Pilon (v2.10.4)	Utturkar et al.[40]	https://github.com/broadinstitute/pilon
Lachesis (v2.1)	Burton et al.[41]	https://github.com/shendurelab/LACHESIS
Allhic (v0.9.8)	Zhang et al.[42]	https://github.com/tangerzhang/ALLHiC
3D-DNA (v180114)	Dudchenko et al.[43]	https://github.com/theaidenlab/3d-dna
BWA (v0.7.15)	Li and Durbin[44]	https://github.com/lh3/bwa
CEGMA (v2.5)	Parra et al.[45]	https://github.com/KorfLab/CEGMA_v2
BUSCO (v4)	Simao et al.[46]	https://busco.ezlab.org/busco_v4.html
RNAmmer (v1.2)	Lagesen et al.[47]	https://services.healthtech.dtu.dk/services/RNAmmer-1.2/
tRNAscan-SE (v2.0)	Chan and Lowe[48]	https://trna.ucsc.edu/software/
Rfam (v13)	Kalvari et al.[49]	http://rfam.xfam.org/
Tandem Repeats Finder (TRF) (v4.09)	Benson[50]	https://tandem.bu.edu/trf/home
LTR_FINDER (v1.07)	Xu and Wang[51]	https://github.com/oushujun/LTR_FINDER
LTR_harvest (v1.6.2)	Ellinghaus et al.[52]	https://genometools.org/
DeepTE (GitHub implementation)	Yan et al.[53]	https://github.com/LiLabAtVT/DeepTE
HelitronScanner (v1.0)	Xiong et al.[54]	https://github.com/Kapeel/Helitron
RepeatMasker (v4.0.6)	Tarailo-Graovac and Chen[55]	http://www.repeatmasker.org/
RepeatProteinMask (bundled with RepeatMasker , v4.0.6)	Tempel[56]	https://www.repeatmasker.org/RepeatProteinMask.html
HISAT2 (v2.2.2)	Kim et al.[57]	https://github.com/DaehwanKimLab/hisat2
StringTie (v2.1.5)	Pertea et al.[58]	https://github.com/gpertea/stringtie
miniprot (v2023)	Li[59]	https://github.com/lh3/miniprot
Augustus (v3.3.3)	Stanke et al.[60]	http://bioinf.uni-greifswald.de/augustus/
EvidenceModeler (v1.1.1)	Haas et al.[61]	https://github.com/EVidenceModeler/EVidenceModeler/wiki
Mikado (v2.0.0)	Venturini et al.[62]	https://www.earlham.ac.uk/mikado
DIAMOND (v2.1.18)	Buchfink et al.[63]	https://github.com/bbuchfink/diamond
Circos (v0.69-3)	Krzywinski et al.[64]	http://circos.ca/
IQ-TREE (v1.6.3).	Nguyen et al.[65]	http://www.iqtree.org/
PAML (v4.10.8)	Yang[66]	https://github.com/abacus-gene/paml/releases/tag/v4.10.8
OrthoMCL (v2.0.9)	Li et al.[67]	https://www.orthomcl.org/
Muscle (v3.8.31)	Edgar[68]	http://www.drive5.com/muscle/
CAFÉ (v5.0)	De Bie et al.[69]	https://www.iresearchgroup.com/cafe/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
KaKs Calculator (v2.0)	Zhang et al.[70]	https://ngdc.cncb.ac.cn/biocode/tools/BT000001
RSEM (v1.3.3)	Li and Dewey[71]	https://github.com/deweylab/RSEM
STEM (v1.3.13)	Ernst and Bar-Joseph[72]	http://www.cs.cmu.edu/~jernst/stem/
MEGA X (v2018)	Kumar et al.[73]	https://www.megasoftware.net/
DESeq2 (v1.46.0)	Love et al.[74]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Plant materials

Fresh leaves of *M. hainanensis*, flowers of *M. hainanensis*, *M. pasquieri* , and *M. longifolia* were collected in Guangzhou (N 23°9′28″, E 113°21′12″), Guangdong Province, China. Flowers of three *Madhuca* species were sampled at three developmental stages for transcriptome sequencing, namely: Bud exposure stage (S1), Full bloom stage (S2), and Late bloom stage (S3)(Figure 1). Each stage included three replicates of ten flowers. Samples were wrapped in aluminum foil, flash-frozen in liquid nitrogen, and stored at -80°C.

METHOD DETAILS

Floral morphology and pollen scanning electron microscope

Flowers from the three *Madhuca* species were dissected and photographed to observe morphological divergence. Fresh pollen from healthy flower buds was collected, air-dried, and sputter-coated with gold using an EM ACE600 vacuum coater (Leica, Hessian, GER). Pollen samples were then examined using an EVO MA 15 scanning electron microscope (ZEISS, Oberkochen, and GER, respectively) to capture representative images of pollen grains, including polar view, equatorial view, and exine ornamentation.[75] Pollen grain measurements (polar axis length P and equatorial axis length E) were taken using SmartSEM software,[76] with ten pollen grains measured per species and reported as mean ± standard error.

Genomic sequencing, RNA-Seq and Hi-C

Genomic DNA was extracted from 5 g of fresh young leaves of *M. hainanensis* using the CTAB method. The DNA quality was assessed using a Qubit Fluorometer (Thermo Fisher Scientific, Waltham, MA), a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA), and 1% agarose gel electrophoresis. Only high-quality DNA was used for subsequent library construction and sequencing. PacBio HiFi Sequencing: High-quality DNA was sheared into 8-10 kb fragments, end-repaired, and ligated to SMRTbell adapters. Sequencing was performed on the PacBio platform. Illumina Sequencing: DNA was sonicated, end-repaired, adapter-ligated, PCR amplified, and purified. Library quality was checked using an Agilent 2100 (Agilent, Santa Clara, CA) bioanalyzer and real-time PCR. Sequencing was performed on the Novaseq 6000 system with a PE150 strategy, and data were processed with fastp (v0.18.0)[36] to obtain clean reads. High-throughput chromosome conformation capture (Hi-C) Sequencing: Cells were cross-linked, lysed, digested with restriction enzymes, and the DNA was ligated and purified. Sequencing was performed on the Illumina HiSeq X Ten platform with a PE150 strategy. Genome size was estimated using K-mer analysis with jellyfish (v2.2.6)[37] and GenomeScope (v1.0.0) based on PacBio sequencing data.[38]

Genome assembly and annotation

Genome Assembly: HiFi reads were assembled using Hifiasm (v0.15.1-r334),[39] and the assembly was polished using Illumina data with Pilon.[40] Hi-C Assisted Assembly: Hi-C interaction data were used to anchor scaffolds to chromosomes using Lachesis (v2.1),[41] Allhic (v0.9.8),[42] and 3D-DNA (v180114).[43] Quality Assessment: The assembly quality was evaluated using BWA (v0.7.15)[44] for sequence consistency, and CEGMA (v2.5)[45] and BUSCO (v4)[46] for sequence completeness.

Note: the following genome assembly and polishing pipeline refers to *M. hainanensis.* Genomic resources for *M. pasquieri* and *M. longifolia* were obtained as described above and were used only for downstream comparative analyses.

For comparative analyses, assembly and annotation statistics of *M. pasquieri* and *M. longifolia* were obtained from their published genome resources and summarized alongside *M. hainanensis* (Table 1).

Clarification: the chromosome-level assembly workflow described below was performed for *M. hainanensis*. The *M. longifolia* genome assembly was obtained from the published resource (cited in the manuscript), and the *M. pasquieri* assembly was generated by our research group but has not yet been presented in a standalone publication; both were included here solely for comparative analyses and are provided with appropriate data attribution in Data and Code Availability.

Genome annotation

Non-coding RNA Annotation: Non-coding RNAs were predicted using RNAmmer (v1.2)[47] for rRNA, tRNAscan-SE (v2.0)[48] for tRNA, and Rfam (v13)[49] for sRNA and miRNA. Repeat Sequence Annotation: Tandem repeats were identified with TRF (v4.09),[50] and dispersed repeats were detected using LTR_FINDER(v1.07),[51] LTR_harvest (v1.6.2),[52] DeepTE,[53] HelitronScanner (v1.0).[54] And RepLoc.[77] RepeatMasker (v4.0.6)[55] and RepeatProteinMask[56] were used for further identification.

Protein-Coding Gene Annotation: Protein-coding genes were predicted using RNA-seq data with HISAT2[57] and Stringtie[58] for transcript assembly, and homology-based predictions with miniprot.[59] *De novo* prediction was performed using Augustus[60] and SNAP,[78] and results were integrated with EVM.[61] Final gene models were refined using Mikado.[62]

Functional Annotation: Predicted genes were functionally annotated using Diamond[63] against NCBI non-redundant protein database (NR), clusters of orthologous groups for eukaryotic complete genomes (KOG), SwissProt, and Kyoto encyclopedia of genes and genomes (KEGG) databases, with Gene Ontology (GO) annotation done using eggNOG-mapper[79] and eggNOG.[80] Visualization of genome features for *M. hainanensis* was performed using Circos software (v0.69-3).[64]

Demographic history of M. hainanensis

Demographic history was inferred using Pairwise Sequentially Markovian Coalescent (PSMC) analysis with Illumina sequencing data, assuming a generation time of approximately 25 years and a mutation rate of 3.39×10^{-9}.[81]

Comparative phylogenomics and whole genome duplication analysis

Phylogenetic analysis and divergence time estimation

Single-copy orthologous genes were identified across the selected plant species using DIAMOND (v2.1.18)[63] with the sensitive mode. Multiple sequence alignments were performed for each orthologous gene set, and concatenated alignments were used to construct a maximum-likelihood phylogenetic tree using IQ-TREE (v1.6.3). The best-fitting substitution model was determined as GTR+F+R4 according to the Bayesian Information Criterion.

Divergence times were estimated using the MCMCTree program implemented in the PAML (v4.10.8) package.[66] Multiple fossil- and TimeTree-based calibration points were applied, with detailed calibration nodes, mean ages, and 95% confidence intervals provided in Table S21.

Gene Family Analysis: Gene families of *M. hainanensis*, *M. pasquieri*, *M. longifolia*, and 12 other angiosperms were identified using DIAMOND[63] and OrthoMCL.[67] Single-copy orthologous genes from the 15 species were aligned using Muscle (v3.8.31).[68] The phylogenetic tree was constructed using IQ-TREE[65] with the GTR+F+R4 model. Divergence times were estimated using the MCMCTree tool in the PAML package,[66] with calibration points obtained from the TimeTree database. The expansion and contraction of gene families relative to the common ancestor were analyzed using the Computational Analysis of gene Family Evolution (CAFÉ) software (v5.0).[69] Whole-Genome Duplication (WGD) events were inferred based on Ks distributions, and Synteny was analyzed using BLASTP, ParaAT, and KaKs Calculator.[70]

RNA sequencing and gene expression analysis

Total RNA was extracted from flowers at different developmental stages using the Trizol kit (Invitrogen, Carlsbad, CA, USA), and RNA-seq libraries were constructed using the NEBNext Ultra RNA Library Prep Kit (NEB #7530, New England Biolabs, Ipswich, MA, USA). Sequencing was performed on an Illumina platform, and clean data were obtained using fastp (v0.18.0).[36] Differential gene expression analysis was performed using HISAT2[57] for alignment, Stringtie[58] for transcript reconstruction, RSEM[71] for expression quantification, and DESeq2[74] for normalization and statistical testing. For trend analysis and data visualization of differentially expressed genes (DEGs), the Short Time-series Expression Miner (STEM) software was used.[72] The expression data files for all DEGs were input into STEM, and trend analysis was conducted. GO and KEGG enrichment analyses were performed on genes within significant trends using hypothesis testing to calculate *P*-values, followed by FDR correction.[82] A Q-value $\leq$0.05 was used as the threshold for defining significant enrichment in GO and KEGG categories.

qRT-PCR analysis

Homologous genes were identified using Protein-Protein BLAST (v2.6.0), multiple sequence alignments were performed with MEGA X,[73] and phylogenetic trees were constructed. Gene expression was validated by qRT-PCR using HiScript® Q RT SuperMix (R123-01) and QuantStudio™ 5 Real-Time PCR System (Applied Biosystems). Reference genes for *M. pasquieri*, *M. hainanensis*, and *M. longifolia* were Mapa01G00002908, mikado.Chr_08G1687, and mikado.Chr_01G1533, respectively. qRT-PCR data were analyzed using the $2^{-\Delta\Delta Ct}$ method to calculate the relative expression levels of genes at different developmental stages.

QUANTIFICATION AND STATISTICAL ANALYSIS

Differential expression analysis

Differential expression analysis was performed using DESeq2 with multiple-testing correction; genes and functional terms were considered significant at FDR < 0.05 unless otherwise specified.

qRT-PCR data analysis

Statistical significance for qRT-PCR data was determined by one-way ANOVA followed by Tukey's post-hoc test, with $P < 0.05$ considered significant (different letters indicate significant differences).