Journal of Experimental Botany, Vol. 74, No. 4 pp. 1275–1290, 2023
https://doi.org/10.1093/jxb/erac464 Advance Access Publication 26 November 2022

RESEARCH PAPER

A novel genome sequence of *Jasminum sambac* helps uncover the molecular mechanism underlying the accumulation of jasmonates

Min Xu[1,†], Qiang Gao[1,†], Mengwei Jiang[2], Wenling Wang[2], Juan Hu[2], Xiaojun Chang[2], Dinggao Liu[2], Yuwei Liang[1], Yifan Jiang[3], Fei Chen[4], Chuhao Li[5,6], Haoran Huang[5,6], Feng Chen[7], Fan Li[8], Robert N. Trigiano[9], Jihua Wang[8,*], Chen Jiao[1,*], Xiaofan Zhou[5,6,*] and Liangsheng Zhang[1,10,*]

[1] Genomics and Genetic Engineering Laboratory of Ornamental Plants, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China
[2] Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, College of Horticulture, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[3] College of Horticulture, Nanjing Agricultural University, Nanjing 210095, China
[4] College of Tropical Crops, Hainan University, Haikou 570228, China
[5] Guangdong Laboratory for Lingnan Modern Agriculture, Integrative Microbiology Research Center, South China Agricultural University, Guangzhou 510642, China
[6] Guangdong Province Key Laboratory of Microbial Signals and Disease Control, South China Agricultural University, Guangzhou 510642, China
[7] Department of Plant Sciences, University of Tennessee, Knoxville, TN 37996, USA
[8] Key Laboratory for Flower Breeding of Yunnan Province, National Engineering Research Center for Ornamental Horticulture, Floriculture Research Institute, Yunnan Academy of Agricultural Sciences, Kunming 650205, China
[9] Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, TN 37996, USA
[10] Hainan Institute of Zhejiang University, Sanya 572025, China.

[†] These authors contributed equally to this work.
* Correspondence: zls83@zju.edu.cn, xiaofan_zhou@scau.edu.cn, biochenjiao@zju.edu.cn, or wjh0505@gmail.com

Received 8 August 2022; Editorial decision 10 November 2022; Accepted 22 November 2022

Editor: Björn Usadel, Forschungszentrum Jülich, Germany

Abstract

***Jasminum sambac* is a well-known plant for its attractive and exceptional fragrance, the flowers of which are used to produce scented tea. Jasmonate (JA), an important plant hormone was first identified in *Jasminum* species. Jasmine plants contain abundant JA naturally, of which the molecular mechanisms of synthesis and accumulation are not clearly understood. Here, we report a telomere-to-telomere consensus assembly of a double-petal *J. sambac* genome along with two haplotype-resolved genomes. We found that gain-and-loss, positive selection, and allelic specific expression of aromatic volatile-related genes contributed to the stronger flower fragrance in double-petal *J. sambac* compared with single- and multi-petal jasmines. Through comprehensive comparative genomic, transcriptomic, and metabolomic analyses of double-petal *J. sambac*, we revealed the genetic basis of the production of aromatic volatiles and salicylic acid (SA), and the accumulation of JA under non-stress conditions. We identified several key genes associated with JA biosynthesis, and their non-stress related activities lead to extraordinarily high concentrations of**

© The Author(s) 2022. Published by Oxford University Press on behalf of the Society for Experimental Biology. All rights reserved. For permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

JA in tissues. High JA synthesis coupled with low degradation in *J. sambac* results in accumulation of high JA under typical environmental conditions, similar to the accumulation mechanism of SA. This study offers important insights into the biology of *J. sambac*, and provides valuable genomic resources for further utilization of natural products.

Keywords: Biosynthesis, consensus genome, consumption, fragrance, hormones, jasmonate.

Introduction

Arabian jasmine (*Jasminum sambac*, *Oleaceae*, $2n=2x=26$) is an evergreen shrub commonly cultivated for its beautiful white flowers and aroma. *J. sambac* is native to South and Southeast Asia and is widely cultivated for fine fragrances as well as jasmine scented tea (Zeng *et al.*, 2012). As one of the most popular beverages worldwide, *J. sambac* scented tea has a specific aroma derived from floral volatiles. The chemical composition of scent volatiles indicate that terpenoids, benzenoids/phenylpropanoids, and fatty acid derivatives contribute significantly to the floral scent of this species (Zeng *et al.*, 2012). Similarly, aromatic volatiles of *J. sambac* include three major terpenoids (linalool, α-farnesene, and (E)-β-ocimene) and three major benzenoid esters (benzyl acetate, methylbenzoate, and methylsalicylate) (Bera *et al.*, 2017). *J. sambac* presents single-petal (SP), double-petal (DP), and multi-petal (MP) flower phenotypes. Although SP, DP, and MP jasmines are diploid, there are great differences in their morphological characters (Supplementary Fig. S1). SP plants are relatively short (70–90 cm in height), the stems and branches are small, the buds and flowers are slight, small, and light, and the flower yield is lower than DP jasmine, but higher than MP jasmine. Additionally, SP plants are not resistant to cold or waterlogging, and have limited resistance to disease and insects. DP is the main variety of jasmine used to make scented tea. The plants are 1–1.5 m in height with many branches, thick and hard stems, and thick, dark green leaves. The buds and flowers of DP are larger with abundant floral compounds compared to SP jasmine. DP and MP jasmines have a stronger floral scent than SP varieties, and DP jasmines also have high flower yield and strong resistance to disease, cold, and insects (Wang *et al.*, 2022). The rich source of floral compounds, high yield, and strong disease, cold, and insect resistance of DP *J. sambac* makes it the leading commercial germplasm and a good material for the extraction of compounds for perfumes and the main aroma ingredient in jasmine tea. Although MP jasmine flowers are larger, their fragrance is weaker and yield lower compared to DP jasmine. Weak aroma, low yield, and long duration of blooms render MP plants not suitable for scented tea production.

J. sambac also contains jasmonate (JA), which was first discovered in *Jasmine* species, and its derivative, methyl jasmonate (MeJA), which was first purified from *Jasminum grandiflorum* in 1962 (Demole *et al.*, 1962). JA is the prototypical member of the oxylipin signalling molecules that have overlapping roles in regulating both stress responses and developmental processes in plants (Thines *et al.*, 2007; Kamali *et al.*, 2022, Yang *et al.*, 2022). JA biosynthesis begins with α-linolenic acid followed by serial, cascaded catalysed processes (Gfeller *et al.*, 2010). The core of JA signalling begins with jasmonoyl-isoleucine (JA-Ile), which is the bioactive form of JA and is controlled by a Skp1/Cullin/F-box (SCF^{COI1}) E3 ubiquitin ligase of which the F-box is encoded by *CORONTINE INSENSITIVE 1* (*COI1*). The jasmonate ZIM-domain (JAZ) repressor proteins interact with COI1 and are subsequently degraded by the 26S proteasome if the bioactive hormone JA-Ile is present. Once the JAZ proteins are eliminated, TOPLESS (TPL) and Novel Interactor of JAZ (NINJA) co-repressors release the transcription factor MYC2 (Pauwels *et al.*, 2010). MYC2 cooperates with a transcriptional mediator complex (MED25) to regulate the expression of early JA-responsive genes (Chini *et al.*, 2007; Pauwels *et al.*, 2010). JA accumulates in most plants in response to biotic or abiotic stresses, such as pathogen infection, wounding, insect herbivory, and ozone exposure (Mousavi *et al.*, 2013). However, *Jasminum* spp. can accrue abundant JA without stress (Zhu and Napier, 2017); however, the underlying molecular mechanism(s) have not been identified.

Salicylic acid (SA) is another crucial plant hormone that is important for resistance to biotrophic and semi-biotrophic pathogens and was first classified as a phytohormone in 1992 (Malamy *et al.*, 1990; Raskin, 1992; Fu and Dong, 2013). Previous studies suggested that SA is synthesized from phenylalanine to cinnamic acid via phenylalanine ammonia-lyase (PAL) (Chen *et al.*, 2022). An important milestone in SA biosynthesis was the discovery of PBS3 (avrPphB susceptible) that is involved in the isochorismate (IC) pathway, which is similar to SA biosynthesis in bacteria (Warren *et al.*, 1999). Since the discovery of the PAL and IC pathways, there are still some unanswered questions about how SA is synthesized in plants. For example, how do plants catalyse the biosynthesis of SA from benzoic acid via benzoic acid-2-hydroxylase (BA2H), enhanced Pseudomonas susceptibility 1 (EPS1), and avrPphB susceptible 3 (PBS3) in the IC pathway?

Here, we sequenced the DP *J. sambac* genome using Pacific Biosciences HiFi technology (PacBio, CA, USA), and obtained not only a telomere-to-telomere consensus genome assembly of *J. sambac*, but also two haplotype assemblies (HA and HB). Deciphering the genome of *J. sambac* is an important basis for further utilization of the floral fragrance, SA and JA. The genomic resources and discoveries we present here will be

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

valuable for future studies seeking to understand and utilize the biological characteristics of *J. sambac*.

Materials and methods

Plant material

Young leaves of DP *J. sambac* were used to isolate DNA for genome sequencing. Plants were cultivated under natural conditions in Fuzhou, Fujian Province of China. Leaves were collected, immediately frozen in liquid nitrogen, and stored at -80 °C for DNA extraction. Flower buds, as well as isolated pistils, petals, calyxes, and roots were used for RNA sequencing in this study.

Illumina short read sequencing

Genomic sequencing was completed by Bio-Marker Technologies Company (Bejing, China). Genomic DNA was extracted from *J. sambac* petals using a Plant Genomic DNA Kit (TIANGEN, Beijing, China) following the manufacturer's protocol. The DNA extraction protocol is summarized: about 100 mg of fresh jasmine leaves were placed into a pre-cooled mortar and ground into a powder with liquid nitrogen. The powder was collected into a pre-cooled 2 ml microcentrifuge tube and 700 μl chloroform added and the contents vigorously mixed followed by centrifugation at 12 000 rpm for 5 min. The upper water phase was transferred into a clean 2 ml centrifuge tube with 700 μl GP2 buffer added and mixed thoroughly. The liquid was transferred to adsorption column, centrifuged for 30 sec at 12 000 rpm, and the waste liquid discarded. GD buffer (500 μl) was added to a CB3 adsorption column, centrifuged for 30 sec at 12 000 rpm, the waste liquid discarded, and the CB3 adsorption column placed into the collection tube. Following this, 600 μl of PW bleaching solution was added to the adsorption column and centrifuged at 12 000 rpm for 30 sec, the waste liquid discarded, and the CB3 adsorption column placed into the collection tube and repeated. The CB3 adsorption column was put back into the collection tube, centrifuged at 12 000 rpm for 2 min, and the waste liquid discarded. The CB3 adsorption column was placed at room temperature for several minutes to evaporate the residual bleaching solution from the adsorption material. Finally, DNA was dissolved in ddH_2O and stored at -80 °C until needed. The quality of the DNA was evaluated using 1% agarose gel electrophoresis and a Qubit 2.0 Fluorometer (Life Techologies, USA) for quality and concentration. Illumina data were obtained by using an Illumina HiSeq X Ten platform. Illumina data were used for genome survey and error correction for the PacBio data.

Genome size assessment

Illumina data and flow cytometry analyses were used for genome size assessment. Genome size and heterozygosity were evaluated by jellyfish software (version 2.1.4, Marçais and Kingsford, 2011) and GenomeScope (version 2.0, Vurture *et al.*, 2017). Genome size estimation was based on K-mer frequency analysis. The k-mer frequency distribution was calculated by jellyfish with the parameters 'jellyfish count -m 17 -o kmer17.jf -C -c 7 -s 1G -t 64 -F 2 < (zcat \WGS_1.fq.gz) < (zcat \WGS_2.fq.gz)'. In addition, findGSE, GCE, GenomeScope (version 1 and version 2), and the script estimate_genome_size.pl (https://github.com/josephryan/estimate_genome_size.pl) were also used for estimating genome size based on k-mer frequency (k=15, 16, 17, 18, 19). Genome sizes and ploidy levels were also analysed on a flow cytometer [BD FACSCalibur platform, 488 nm (BD Biosciences)]. Sample preparation was made according to procedures of the CyStain PI absolute reagent kit (Sysmex, Kobe, Japan). The *Solanum lycopersicum* (~900 Mb) and *Oryza sativa* (~460 Mb) genomes were used as internal standards.

PacBio and Hi-C library construction and sequencing

A DNA library was constructed following PacBio's standard protocol and sequenced by the Circular Consensus Sequencing (CCS) mode of the single-molecule real-time (SMRT) sequencing technology on the Sequel platform (PacBio, CA, USA). The process is summarized as follows: first, the genomic DNA (gDNA) was extracted from fresh jasmine leaves. The gDNA was broken by specialized g-TUBE holes and purified. Single hanging DNA was removed from the purified gDNA with the DNA prep enzyme (from the kit). Then, the DNA was repaired with the DNA Damage Repair Mix. Repaired DNA was ligated with adapters and treated with enzymes. Treated DNA was purified, size selected using a BluePippin DNA size selection system (Sage Science, Beverly, MA, USA) and purified again. The constructed library was bound with primers and polymerases using a PacBio Binding Kit. The final reaction products were purified by AMpure PB Beads (PacBio). The Sequencer Sequel II (PacBio) platform was used for SMRT sequencing. The long sequencing reads were initially corrected to obtain sub-reads. HiFi reads of *J. sambac* were generated by merged multiple sub-reads.

To anchor scaffolds onto chromosomes, Hi-C libraries were constructed with 150-bp paired-end reads generated by an Illumina HiSeq X Ten platform (Burton *et al.*, 2013). Petal samples were fixed with formaldehyde and lysed to extract DNA. Extracted cross-linked DNA was cut with restriction endonucleases overnight, resulting in pairs of distally located, but physically associated DNA molecules attached to one another. Fragments with sticky ends were biotinylated and ligated to form chimeric circles. Biotinylated circles were enriched and sheared again, and then processed to sequencing libraries in which individual templates were chimeras of the physically associated DNA molecules from the original cross-linking. Hi-C reads were aligned to *J. sambac* assembly genome.

Genome assembly and phasing using PacBio HiFi reads

A total of 22.04 Gb PacBio HiFi reads (CCS mode) were assembled using HiFi-asm (version 0.16.1, Cheng *et al.*, 2021) to obtain contigs. Hi-C sequencing was evaluated by HiC-Pro pipeline (Servant *et al.*, 2015). Hi-C data were mapped to the assembled contigs using the Burrows–Wheeler Alignment tool (BWA, version 0.7.17-r1188) with default parameters (Li and Durbin, 2010). Chromosome hierarchical clustering and sorting were performed using 3D-DNA (version 180922) (Dudchenko *et al.*, 2017). Lastly, scaffolds were manually assessed to obtain a high-quality consensus genome of *J. sambac* with gap-free sequences in nine chromosomes. The plants with telomere sequence structure characteristics were queried based on the telomere database (http://telomerase.asu.edu/sequences_telomere.html), and the telomeres were found through in-house python scripts in the head and tail of the chromosomes. The *J. sambac* genome sequence is available in the National Genomics Data Center (PRJCA006075). The Diploid Assembly (DipAsm) method (Garg *et al.*, 2020) was used for phasing genome assemblies into haplotigs using PacBio HiFi reads and Hi-C data. HiFi reads were mapped to the consensus assembly using minimap2; variants were called by Deep Variant (version 0.8.0); HiFi and Hi-C data were used to phase variants by using Whatshap and HapCUT2 (version 1.1); HiFi reads were phased by Whatshap; phased HiFi reads for HA and HB were assembled separately by HiFi-asm (version 0.16.1); the contig assemblies of HA and HB were scaffolded separately to generate the final chromosome-level haplotype-resolved assemblies.

Identification of repetitive elements and alleles between HA and HB

RepeatMasker (version 4.0.6) (Chen, 2004) was used to obtain long terminal repeat (LTR) repetitive elements. The LTR_retriever (version 2.9.0) package was used for LTR Assembly Index (LAI) (Ou *et al.*, 2018)

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

assessment of *J. sambac* haplotypes with default parameters. The identification of paired allele genes between two haplotypes was performed by MCScanX (Wang *et al.*, 2012).

Genome annotation and BUSCO assessment

Genome annotation was performed as follows. RNA-seq clean data were mapped to the assembled genome; the mapped bam files were used for AUGUSTUS (version 3.4.0, Stanke *et al.*, 2008) and GeneMark-ET (version 4.33, Lomsadze *et al.*, 2014) training and the RNA-seq assembled by Trinity (version 2.13.2) (https://anaconda.org/bioconda/trinity) and StringTie (version 1.2.3, Pertea *et al.*, 2015). The Program to Assemble Spliced Alignments (PASA, version 2.5.2) (Haas *et al.*, 2003) was used to combine the results of the transcriptome assembly. Lastly, the transcripts generated by PASA, the dicotyledon protein data of OrthoDB (version 10, https://www.orthodb.org), and trained results from AUGUSTUS (version 3.4.0) and GeneMark (version 4.33) were used for MAKER prediction (gmod.org/wiki/MAKER_Tutorial). The results of MAKER were corrected with PASA. Predicted proteins of *J. sambac* HA and HB haplotypes were functionally annotated with Non-Redundant Protein Sequence Database (Nr), *Arabidopsis thaliana*, and rice proteomes as well as Cluster of Orthologous Groups of proteins (COG), Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG). The Benchmarking Universal Single-Copy Ortholog (Simão *et al.*, 2015) (BUSCO, version 4.0.6, http://busco.ezlab.org) embryophyta_odb10 database (1614 conserved single-copy plant genes) with default parameters was used to assess the assembly and annotation quality of the genome and proteins.

Genomic synteny and Ka/Ks analysis

To identify syntenic gene pairs, all-against-all BLASTP alignments (version 2.10.1+) were performed with an e-value 1×10^{-5} using MCScanX (Wang *et al.*, 2012) with default parameters to calculate pairwise similarities located in collinear blocks. The *Ka*, *Ks*, and *Ka/Ks* of the syntenic gene pairs were analysed by the section 'Simple *Ka/Ks* Calculator' of TBtools (https://github.com/CJ-Chen/TBtools/releases) (Chen *et al.*, 2020).

Structural variation analysis

Structural Variation (SV) calling was carried out using the following parameters: NUCMER in MUMMER (Kurtz *et al.*, 2004) was used for alignment (nucmer --mum -c 100 HA.genome.fa HB.genome.fa --prefix HA.HB) of the subgenome sequences of HA and HB. The alignment result files were filtered using delta-filter (delta-filter -q -r -1 -l 100 HA.HB.delta > HA.HB.filter.delta), and the out file HA.HB.filter.delta was then used for downstream analysis. We further analysed the SVs using the online program ASSEMBLYTICS (http://assemblytics.com) with the parameters unique sequence length longer than 10 kbp and variant size range from 50 to 10 000 bp.

RNA-seq and analysis

Total RNA was extracted from flower buds, fully-opened flowers, pistils, petals, calyxes, and roots using a Plant Total RNA Isolation Kit (Biomarker, Beijing, China) and DNA contamination was eliminated using a RNase-Free DNase I Kit (Takara, Dalian, China). RNA was evaluated using the same method as with DNA. RNA was used for cDNA library construction after testing with a NanoDrop spectrophotometer (ThermoFisher, Waltham, USA) and a Qubit 2.0 fluorometer (ThermoFisher). After the PCR products were purified, the library was assessed on an Agilent Bioanalyzer 2100. The assessed libraries were sequenced on an Illumina HiSeq 2500 platform (San Diego, CA, USA) according to the manufacturer's instructions. The raw RNA-Seq data were filtered to obtain clean data by removing low-quality bases, adaptors, duplications, and potential contaminants using Trimommatic (Bolger *et al.*, 2014). Clean data reads were mapped into HA and HB using TopHat2 (version 2.0.14, Kim *et al.*, 2013) using the default parameters. The expression levels fragments per kilobase million (FPKM) and transcripts per million (TPM) were obtained using StringTie (version 1.2.3). All the sequencing data have been deposited in the National Genomics Data Center (https://ngdc.cncb.ac.cn) with Genome Sequence Archive (GSA) accession number CRA007626 (BioProject accession number: PRJCA010780). In addition, RNA-seq data of SP and DP jasmines with project number PRJCA006739 (https://bigd.big.ac.cn/) (Wang *et al.*, 2022), and RNA-seq data of MP jasmines (Xu *et al.*, 2022) under accessions SRR14317414–SRR14317417 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA723725) were also used for analysis (Supplementary Table S1).

Targeted metabolomics analysis of phytohormones in J. sambac

Three duplicates of nine samples of *J. sambac* young leaves, buds, and fully-opened flowers were collected for targeted metabolomic analysis using ultra-high-performance liquid chromatography-tandem mass spectrometry (UHPLC-MS/MS). First, the samples were ground into a fine powder in liquid nitrogen. An aliquot of each individual sample was accurately weighed and transferred to a sterile Eppendorf tube. After the addition of 1000 µl of extract solution (50% acetonitrile in water, precooled at -40 °C, containing an isotopically-labelled internal standard mixture), the samples were vortexed for 30 sec and sonicated for 5 min in an ice-water bath, and then homogenized at 40 Hz for 4 min. Homogenization and sonication were repeated three times. The samples were centrifuged at 13 000 rpm for 15 min at 4 °C, and an 800 µl aliquot of the supernatant was evaporated to dryness under a gentle stream of nitrogen, and finally reconstituted in 80 µl of 10% acetonitrile/H_2O (v/v). All the samples were vortexed for 30 s and sonicated for 5 min in an ice-water bath. Reconstituted samples were transferred to an Eppendorf tube with a filter membrane for filtration. After the samples were centrifuged at 12 000 rpm for 15 min at 4 °C, the clear supernatant was subjected to UHPLC-MS/MS analysis. Next, stock solutions were individually prepared by dissolving or diluting each standard substance to yield a final concentration of 10 mmol l^{-1}. An aliquot of each of the stock solutions was transferred to a 10 ml flask to form a mixed working standard solution. A series of calibration standard solutions were then prepared by stepwise dilution of the mixed standard solution (containing an isotopically-labelled internal standard mixture in identical concentrations to the samples). The UHPLC separation was carried out using an EXIONLC System (Sciex), equipped with a Waters ACQUITY UPLC CSH C18 column (150 × 2.1 mm, 1.7 µm).

Comparative genomics and gene family analysis

Rooted species trees of *J. sambac*, *Fraxinus excelsior*, *Olea europaea*, *Osmanthus fragrans*, *Mimulus guttatus*, *Sesamum indicum*, and *Coffea canephora* in the *Lamiales* were constructed by a single copy homologous gene with OrthoFinder (version 2.5.1, Emms and Kelly, 2015, 2019). The orthogroups (OGs) of the different species of jasmine genomes (SP, DP, MP, HA, and HB) were also obtained by OrthoFinder. The comparison of the jasmine genomes was mainly based on the results of OrthoFinder. Aromatic volatile-related candidate genes, such as *HMGR*, *AACT*, and *BEAT* gene families, were identified by running BLASTP against both haplotype-specific genomes and the monoploid genome with a filtered parameter (e-value $<10^{-5}$, identity ≥40%, and coverage ≥30%) (Altschul *et al.*, 1990). To identify the terpene synthase (*TPS*) genes, the Hidden Markov Model (HMM) profiles of TPS domains (PF01397.22 and

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

PF03936.17) were queried against the haplotypes and genome protein sequences using the hmmsearch tool HMMER (version 3.2.1) with an e-value $<10^{-5}$ (Wheeler and Eddy, 2013). Multiple sequence alignment was carried out using MAFFT (version 7.471, Katoh and Standley, 2013). Similar strategies were used to identify genes in other families.

Results and discussion

The nearly complete genome sequence of J. sambac

A total of 73.8 Gb (~148×) of Illumina data, 22.04 Gb (~41×) of PacBio HiFi data, and 96.13 Gb (~179×) of Hi-C data were generated. We *de novo* assembled the HiFi reads using HiFi-asm (Cheng *et al.*, 2021) (Supplementary Fig. S2), which produced a consensus genome assembly sequence of 537.99 Mb. The assembly size was similar to the estimated genome size of *J. sambac* based on flow cytometry analysis (499.48 Mb) (Supplementary Fig. S2; Supplementary Table S2) and k-mer analysis of Illumina short reads (429 Mb to 735 Mb) (Supplementary Table S3; Supplementary Figs S3, S4). The predicted genome size ranged from 429 Mb to 735 Mb with different tools. The GCE results showed that there is little difference (514.7~525 Mb) in genome size estimation.

The assembly of 866 contigs with an N50 length of 35.96 Mb was anchored into pseudo-chromosomes with the Hi-C data. The final assembly consisted of 397 scaffolds with 91.95% (494.66 of 537.99 Mb) of the assembled sequences placed on 13 chromosomes, among which nine were gap-free (Fig. 1A) and harboured telomeric repeats at both ends (Fig. 1B). The assessment of the assembly sequence using BUSCO (Simão *et al.*, 2015) (completeness value: 97.6, Supplementary Fig. S5; Supplementary Table S4) and the LTR Assembly Index (Ou *et al.*, 2018) (LAI score: 16.56) supported the high completeness of the *J. sambac* genome assembly sequence.

Haplotype-resolved genome assembly and structural variations analysis of J. sambac

In addition to the consensus assembly sequence, we generated phased, chromosome-scale genome assemblies for both haplotypes A (HA) and B (HB) of *J. sambac* (Supplementary Fig. S6) using the DipAsm method (Garg *et al.*, 2020) (Supplementary Fig. S1). The sizes of both phased assembly sequences (HA: 537.30 Mb; HB: 537.60 Mb) were very close to the consensus assembly sequence. The N50 values of the contigs/scaffolds were 0.49/40.49 Mb in HA and 0.53/41.55 Mb in HB (Supplementary Table S5). We identified 39 381 protein-coding genes in HA, 39 201 in HB, and 39 843 in the consensus genome sequence. The distribution of gene density and syntenic blocks are shown in circos (Fig. 1C). A total of 31 864 pairs of allelic genes from 247 syntenic blocks (Fig. 1B) were identified in HA and HB. Moreover, the estimated SNP switch error rate was low (7.65%) between HA and HB, indicating a high accuracy of phasing (Fig. 1D). Both HA and HB assemblies had BUSCO completeness values greater than 97% (Supplementary Fig. S5) and LAI scores greater than 15 (HA: 15.02; HB: 15.25), affirming their high quality. Furthermore, 11.11% Ty1-*copia*, 14.24% Ty3-*gypsy*, and 14.18% unknown LTR superfamilies were detected in HA, whereas 9.09% Ty1-*copia*, 14.55% Ty3-*gypsy*, and 13.81% unknown LTR superfamilies were identified in HB (Fig. 2A). The CRM and Galadriel clades of the *Gypsy* superfamily, as well as the Bianca, Ikeros, SIRE, and TAR clades of the *Copia* superfamily, displayed a differential insertion time (Fig. 2B).

The phased assemblies enabled genome-wide allele-specific expression (ASE) analyses. We conducted ASE analysis (Supplementary Fig. S7) based on RNA-seq on vegetative buds, flowers, pistils, calyxes, petals, and roots of *J. sambac*, and found 1701 HA-preferentially expressed genes and 1767 HB-preferentially expressed genes in *J. sambac* buds (adjusted $P \leq 0.05$, fold change ≥2) (Fig. 2C). Moreover, a total of 1758 HA-preferentially expressed genes and 1862 HB-preferentially expressed genes were identified in *J. sambac* flowers (adjusted $P \leq 0.05$, fold change ≥2) (Fig. 2D). The expression patterns of homologous genes in flowers were divided into different patterns consistent with the ASE results (Fig. 2E). Structure variation (SV) is an important source of genetic variation causing ASE. We found 16 752 SVs, ranging from 50 bp to 10 kb, between HA and HB, among which ~15.8% were located in the coding regions of annotated genes (Fig. 1E). The statistics show that 9.32% (3672/39 381) SV-associated genes in HA and 9.0% (3532/39 201) SV-associated genes in HB were identified. GO and KEGG pathway enrichment analyses showed that SV-affected genes are mainly involved in the adaptability and resistance of jasmines, such as ageing, response to abiotic stimulus GO terms, and amino acid metabolism pathways (Supplementary Fig. S8). SVs have an impact on plant agronomic traits, for example, a 5576-bp deletion results a flower developmental gene *APETALA2* (*AP2*) JaA8G35779 being expressed in HA but not in HB (Fig. 1F).

J. sambac *has not experienced recent polyploidization since the* Oleaceae*-wide WGT event*

Previous genomic comparisons of *F. excelsior* (Sollars *et al.*, 2017), *Olea europaea* (Unver *et al.*, 2017), *Osmanthus fragrans* (Yang *et al.*, 2018), and *J. sambac* cv. *trifoliatum* (Xu *et al.*, 2022) revealed that these species have experienced two polyploidization events, including an ancestral WGT event common to plants in the *Oleaceae* and a more recent WGD event not shared by *J. sambac*. We identified 13 991 common OGs among the *Oleaceae* species (Supplementary Fig. S9), and constructed a phylogenetic tree of the seven *Lamiales* species along with *C. canephora* as an outgroup using 1327 single-copy genes (Supplementary Fig. S10). Our *Ks* analysis of the collinear gene pairs of *J. sambac* and pairwise synteny visualization of *J. sambac* versus *O. europaea* also confirmed that *J. sambac* did not experience any recent

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

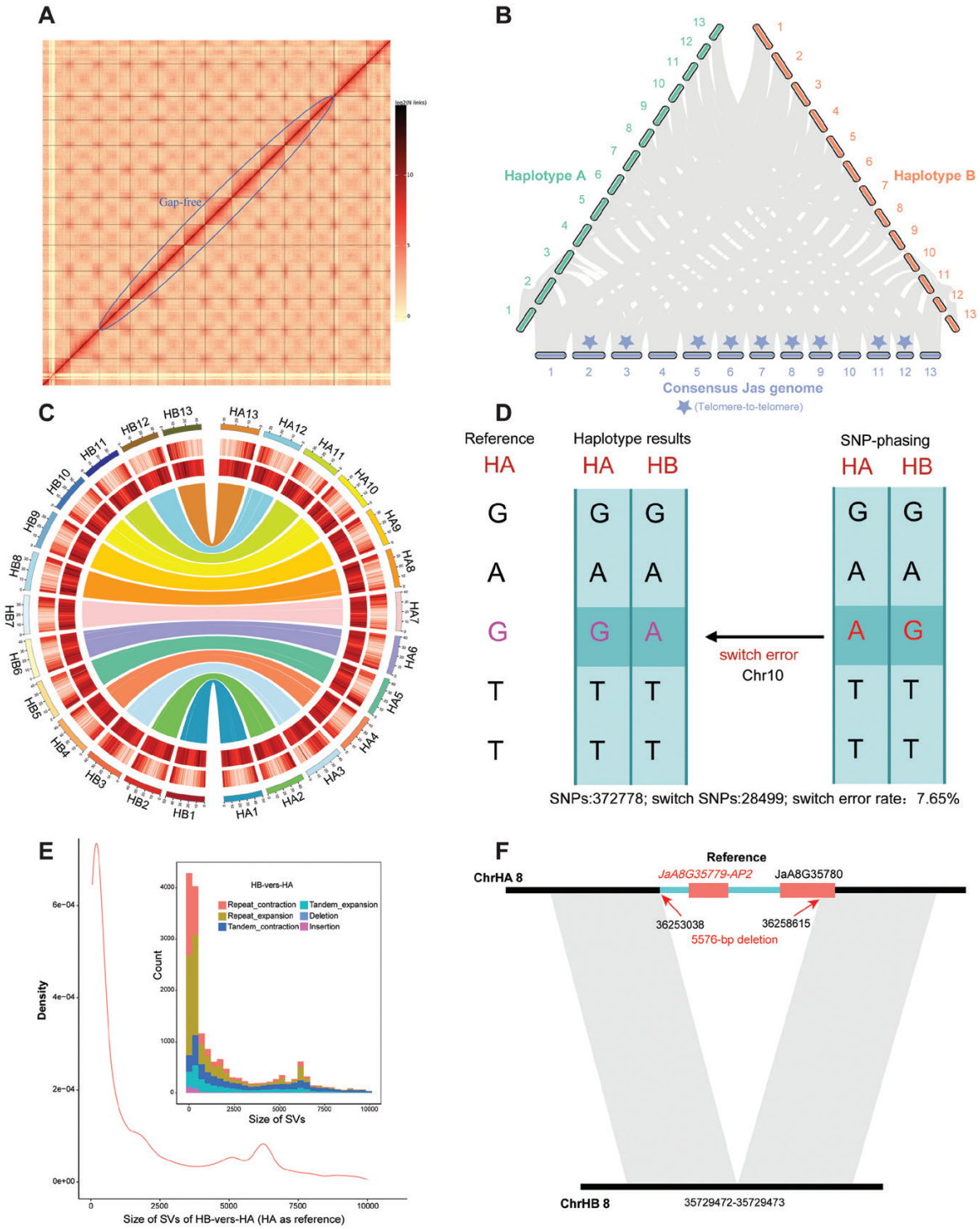

Fig. 1. Analysis of the *J. sambac* genome. (A) A Hi-C contact matrix visualization for individual chromosomes of the *J. sambac* diploid genome assembly. The pixel intensity represents the count of Hi-C links at 150-Kb-sized windows on the chromosomes on a logarithmic scale. A darker red colour indicates higher contact probability, whereas yellow space represents fewer or no contacts. Nine chromosomes with no gaps are shown with a blue ellipse. (B) Collinearity of the diploid (HA and HB) and consensus genomes (Jas) of *J. sambac*. The 13 chromosomes belonging to HA, HB, and Jas are shown in green, blue, and orange, respectively. Syntenic blocks between homoeologous regions are shown in grey. Blue pentacles represent telomere-to-telomere assembled chromosomes. (C) A circos diagram of HA and HB haplotype pseudochromosomes of *J. sambac*. From the outer to the inner, circles represent pseudochromosomes, gene densities, transposons, and the collinearity of the HA and HB haplotypes. (D) SNP analysis to assess haplotype assembly quality. (E) Statistics of SVs in the haplotypes (HB versus HA). (F) A 5576-bp deletion gives rise to *AP2* gene *JaA8G35779* expression in HA, shown in red font.

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

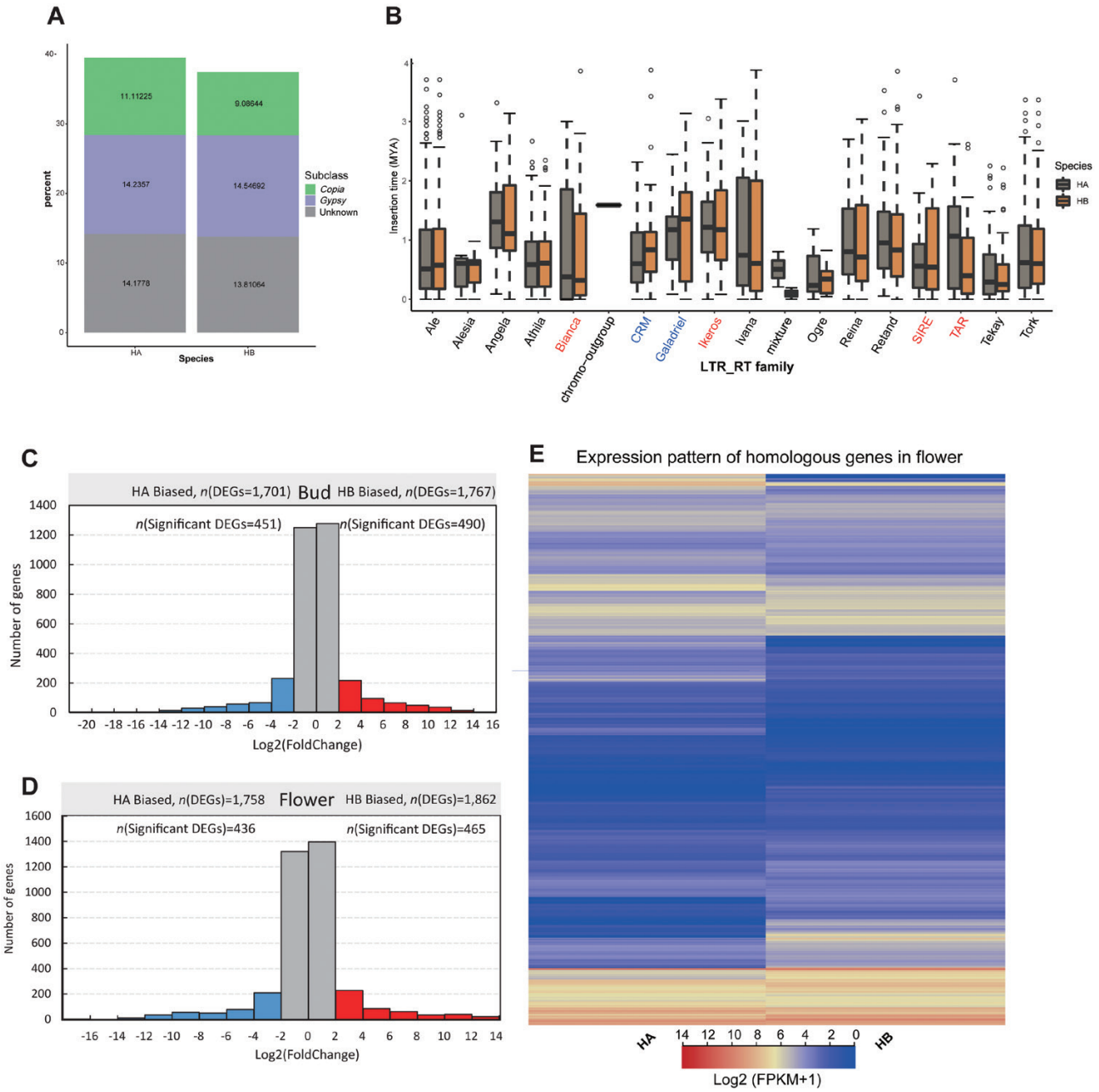

Fig. 2. The HA and HB haplotype of *J. sambac*. (A) Distribution of the *Copia* and *Gypsy* superfamily of LTR retrotransposons. (B) The insertion time (MYA) of the top 16 clades of LTR. The CRM and Galadriel clades of *Gypsy* (blue font); the Bianca, Ikeros, SIRE, and TAR clades of *Copia* (red font) displayed a higher differential insert time. (C) and (D) Histograms of biallelic expression of divergent alleles in buds and flowers of *J. sambac*. (E) Expression pattern of homologous genes in flowers (a total of 3868 pairs).

polyploidization after the *Oleaceae*-wide WGT event (Supplementary Fig. S10).

Comparison of different jasmine genomes

Two haplotype (HA and HB) genomes and a nearly complete consensus genome were obtained in this study and compared to the previous published DP jasmine genome (Wang *et al.*, 2022). The faint perfume of jasmine forms during the course of flowering. We compared the assemblies of haplotypes A and B reported in this study, and the published genomes for SP, DP, and MP jasmines (Wang *et al*, 2022; Xu *et al.*, 2022) (Supplementary Table S6). As a result, a total of 29 797 OGs were identified (Fig. 3A), among which 654 OGs were present in all but SP jasmines. GO enrichment analysis showed that these OGs were mainly associated with aromatic volatile processes

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

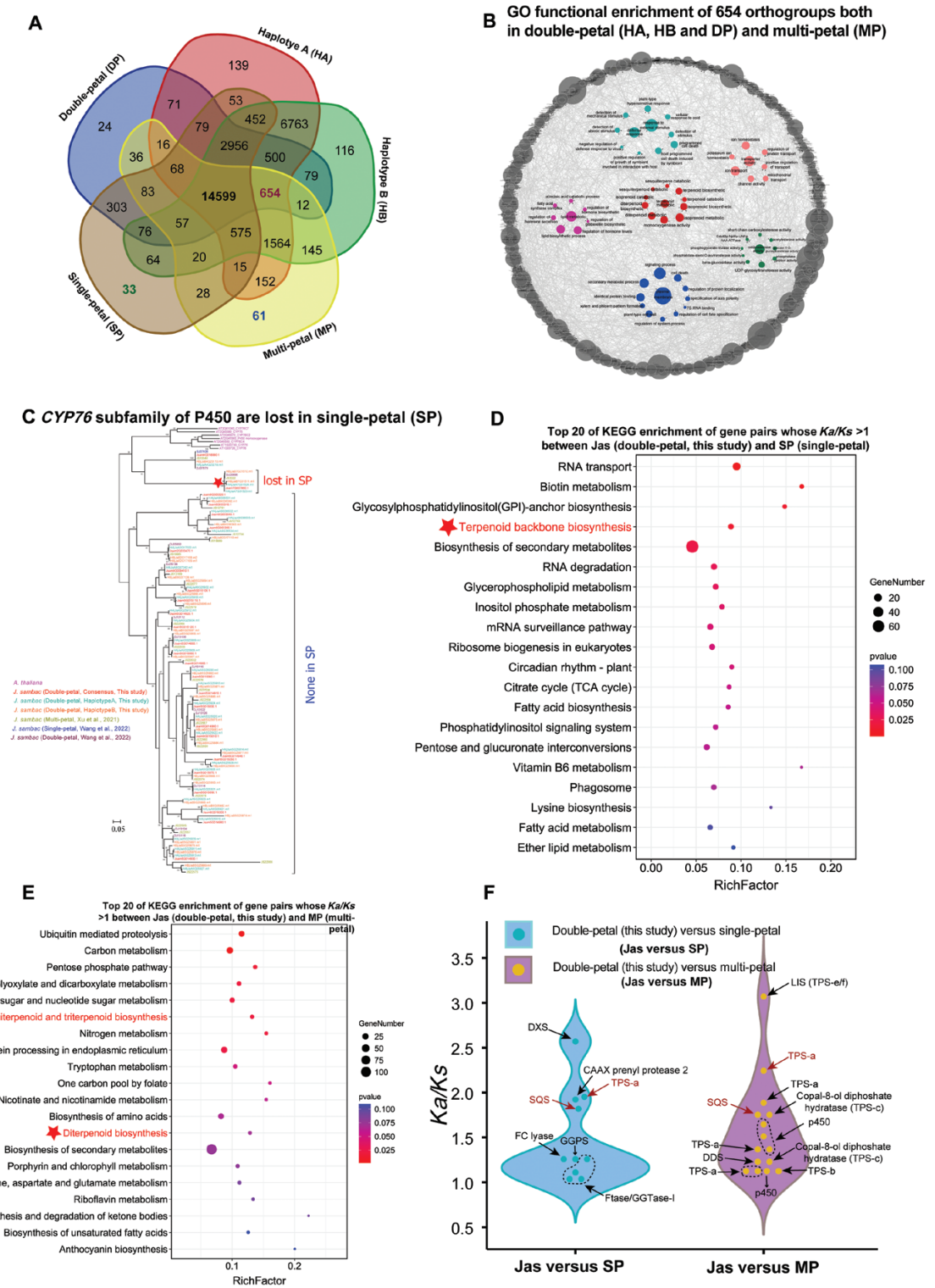

Fig. 3. Comparative genomics reveals the mechanism of the aromatic volatile accumulation in DP and MP *J. sambac*. (A) Venn diagram of OGs among haplotype A (HA) and haplotype B (HB) of DP (this study, Jas), and SP, DP (from Wang *et al.* 2022), and MP (from Xu *et al.* 2022) of *J. sambac*. (B) Go functional enrichment of 654 common OGs which all occur in double-petal (HA, HB, and DP) and multi-petal (MP) of *J. sambac*. (C) *CYP76* subfamily genes which associated with linalool metabolism of P450 are lost in SP *J. sambac*. (D) and (E), Top 20 of KEGG enrichment of gene pairs whose *Ka/Ks* >1 in DP (this study, Jas) versus SP, and DP (this study, Jas) versus MP, respectively. (F) *Ka/Ks* shows genes associated with aromatic volatile biosynthesis in DP *J. sambac* experienced positive selection (*Ka/Ks* >1). Abbreviations: DXS: 1-deoxy-D-xylulose-5-phosphate synthase; LIS: S-linalool synthase; TPS: terpene synthase; SQS: squalene synthase; GGPS: geranylgeranyl pyrophosphate synthase; DDS: dammarenediol II synthase-like; FC lyase: farnesylcysteine lyase; Ftase/GGTase-I: farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha.

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

(terpenoid biosynthetic and terpenoid catabolic processes), stress response (plant-hypersensitive response, response to cold, virus, and stimulus), ion homeostasis, abscisic acid catabolic processes, and lipid and fatty acid metabolic processes (Fig. 3B). For instance, the CYP76 subfamily of P450, which is involved in linalool biosynthesis, has only one member in SP jasmines, but more than 20 copies in DP and MP jasmines (Fig. 3C).

The *Ka/Ks* of homologous gene pairs was calculated in our consensus genome versus SP [DP (Jas)-vs-SP] and MP [DP (Jas)-vs-MP], respectively. In the Jas-vs-SP comparison, genes with signatures of positive selection (*Ka/Ks*>1) were mainly enriched in KEGG pathways related to the biosynthesis of secondary metabolites, terpenoid backbone biosynthesis, and fatty acid biosynthesis and metabolism (Fig. 3D). Comparatively, positively selected genes in the DP (Jas)-vs-MP comparison were mainly enriched in biosynthesis of secondary metabolites, diterpenoid biosynthesis, sesquiterpenoid and triterpenoid biosynthesis, and carbon metabolism pathways (Fig. 3E). Some exemplar genes under positive selection include: (1) in DP (Jas)-vs-SP: 1-deoxy-D-xylulose-5-phosphate synthase (*DXS*), geranylgeranyl pyrophosphate synthase (*GGPS*), farnesylcysteine lyase (*FC lyase*), farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha (*Ftase/GGTase-I*), and CAAX prenyl protease 2; (2) in Jas-vs-MP: S-linalool synthase (*LIS*), terpene synthases (*TPS- b,c,e/f*) and the *p450* genes; and (3) in both comparisons: a squalene synthase (*SQS*, *Jsam5G004870.1*) and a terpene synthase (*TPS-a*, *Jsam9G014600.1*) (Fig. 3F) (Supplementary Table S7). *Ka/Ks* analysis indicated that some genes (e.g. *TPS*, *p450*, *SQS*, *LIS*) associated with aromatic volatile biosynthesis were positively selected in DP jasmines. This also could explain why DP jasmines are the most fragrant, whereas the MP jasmines are intermediate in fragrance, and SP jasmines are the least fragrant.

Floral C-class MADS-box genes *AGAMOUS (AG)* or *AGAMOUS-like (AGL)* specify carpel identity, whereas class A and B (*APETALA3*, *AP3*; *PISTILLATA*, *PI*) genes specify petal identity (Zhang *et al.*, 2020). The *AG* and *AGL1/5* genes were highly expressed during the flower development stages S1–S4 of SP (e.g. AG FPKM =~200 at S2 in SP), but weakly expressed in DP and almost non-expressed in MP (e.g. AG FPKM =~23 at S2 in DP and FPKM =~14 in MP) (Supplementary Fig. S11). However, the *PI* and *AP3* genes were both highly expressed and have similar expression trends in SP and DP jasmines (e.g. PI FPKM =~3436 at S5 in SP, ~3100 at S5 in DP) (Supplementary Fig. S11). The expression of these MADS-box genes in jasmine showed that the low or non-expression of C-class *AG* might be involved in MP formation.

Accumulation of aromatic volatiles in J. sambac

GC-MS analyses of jasmine buds and flowers showed that (E)-3-hexen-1-ol acetate, benzyl acetate, linalool, and α-farnesene were the primary metabolites that contributed to the biosynthesis of aromatic volatiles (Fig. 4A). Genes associated with aromatic volatiles, such as those participating in terpenoid backbone biosynthesis, which mainly included 2-C-methyl-D-erythritol 4-phosphate (MEP) and mevalonate (MVA) pathways, generally showed high expression in buds, flowers, and petals (Fig. 4B). Genes associated with another important floral scent volatile, 3-hexen-1-ol-acetate biosynthesis (*HPL1*, *ADH1*, and *CHAT*), also showed high expression (Fig. 4C) in the same samples. Also noteworthy, the terpene synthase (*TPS*) genes were involved in the last catalytic reaction in the MEP and MVA pathways to generate terpenoids (e.g. linalool and α-farnesene) and separated into *TPS-a, b, g, c, e/f* (Fig. 4D). The *TPS* genes showed extensive tandem-duplications across the genome, especially in chromosomes 12 and 13 (Supplementary Figs S12–S13). Most of the *TPS* genes were more highly expressed in buds, flowers, and petals compared with other tissues (Fig. 4E). Notably, the HA haplotype had more TPS genes than the HB haplotype (87 versus 79, respectively) due to a SV imbalance, but these TPS genes showed a HB-biased expression pattern (Fig. 4E). Benzyl alcohol acetyltransferase (*BEAT*) is the key enzyme that catalyses the substrates of benzyl alcohol and acetyl-CoA to generate benzyl acetate and CoA (Dudareva *et al.*, 1998; Bera *et al.*, 2017). A pair of *BEAT* allelic genes (*JaA11G08221* and *JaB11G08390*) were also highly expressed (Supplementary Fig. S14). Overall, our results showed that the critical genes involved in aromatic volatile biosynthesis in *J. sambac* mainly included *HMGS*, *HMGR*, and MVD in the MVA pathway; *DXS*, *DXR*, *HDS*, *TPS*, and *FPS* in the MEP pathway; and *HPL*, *ADH*, and *CHAT* in the 3-hexen-1-ol-acetate biosynthesis pathway (Fig. 4C).

JA biosynthesis and consumption of J. sambac

In this study, we examined JA accumulation and consumption, and expression of relevant genes in *J. sambac*. Twelve phytohormones, including JA, SA, and abscisic acid (ABA), were identified by UHPLC-MS/MS from young leaves, buds, and fully-opened flowers of jasmine (Supplementary Fig. S15). The concentration of JA was up to 4186.85 ng g^{-1}, and that of the JA consumption formation moiety (JA-Ile) was only about 37.64 ng g^{-1} in fully-opened flowers (Fig. 5B), indicating that JA is accumulated in fully-opened flowers. Genes encoding key JA biosynthetic enzymes were mainly highly expressed in buds, flowers, and petals, especially *AOC* (FPKM=~1400 in buds, ~275 in flowers), *OPR*, and *OPCL1*, and those associated with β-oxidation [e.g. *ACX* (FPKM=~462 in buds, ~480 in petals), *MFP* and *KAT* (FPKM=~1525 in buds, ~657 in petals)]. In *J. sambac*, *JMT* (FPKM=~5 in buds, ~38 in flowers, ~44 in petals) had lower expression under natural conditions (Fig. 5C). However, methyl jasmonate esterase (*MJE*) which has a role in converting MeJA to JA had higher expression (FPKM=~4112 in flowers, ~536 in petals) than other genes in the JA biosynthesis pathway of *J. sambac*. *MJE* is a subfamily of the methyl esterase (*MES*) family (Supplementary Fig. S16). This suggests that most of the *MJE* is used for JA and MeJA conversion, and only part of it is involved in MeSA biosynthesis.

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

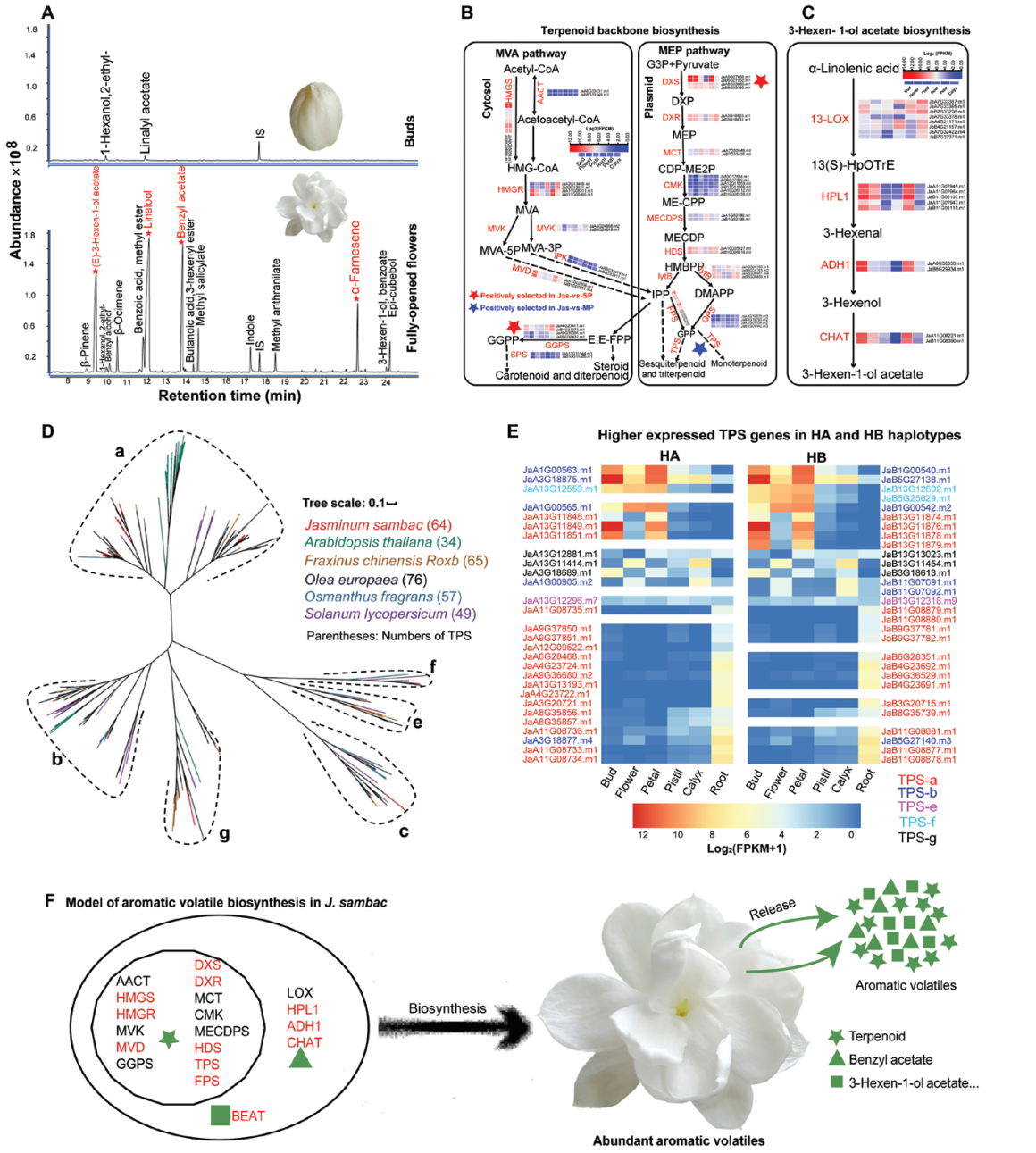

Fig. 4. Aromatic volatile biosynthesis in *J. sambac*. (A) GS-MS analysis of buds and fully-opened flowers of *J. sambac*. (B) Terpenoid backbone biosynthesis involved in aromatic volatiles. (C) 3-hexen-1-ol acetate biosynthesis of aromatic volatiles in *J. sambac*. DXS and GGPS genes were positively selected in Jas-vs-SP (red pentacle), TPS genes were positively selected in Jas-vs-MP (blue pentacle). (D) Phylogenetic analysis and subfamily classification of TPS genes (87 in HA and 79 in HB), the letters a-c, e-g: the subfamily of TPS genes. (E) Homologue expression bias (HEB) analysis of TPS genes in HA and HB, TPS genes are biased HB. (F) Model of aromatic volatile biosynthesis. Red font represents highly expressed genes associated with aromatic volatiles. Abbreviations: AACT: acetyl-CoA acetyltransferase; HMGS: 3-hydroxyl-3-methylglutaryl-CoA synthase/ HMG-CoA synthase; HMGR: 3-hydroxyl-3-methylglutaryl-CoA reductase; MVK: phosphokinase: mevalonate phosphokinase (MVA) kinase; MVD: diphosphomevalonate decarboxylase; IPK: isopentenyl phosphate kinase; GGPS: geranylgeranyl diphosphate synthase/ geranylgeranyl pyrophosphate synthase; SPS: solanesyl diphosphate synthase; DXS: 1-deoxy-D-xylulose-5-phosphate synthase/DXP-synthase; DXR: 1-deoxy-D-xylulose-5-phosphate reductoisomerase; MCT: 2*C*-methyl-D-erythritol-4-phosphate (MEP) cytidylyltransferase; CMK: 4-(cytidine-5'-diphospho)-2*C*-methyl-D-erythritol kinase; MECDPS: 2*C*-phospho-4-(cytidine-5'-diphospho)-2*C*-methyl-D-erythritol synthase/2*C*-methyl-D-erythritol 2, 4-cyclodiphosphate synthase; HDS: (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase; lytB: (E)-4-hydroxy-3-methylbut-2-en-1-yl-diphosphate reductase; FPS: farnesyl diphosphate synthase; GPS: geranyl diphosphate synthase; TPS: terpene synthase; HMG-CoA: 3-hydroxyl-3-methylglutaryl-CoA; G3P: glyceraldehyde-3-phosphate; DXP: 1-deoxy-D-xylulose-5-phosphate; MEP: 2*C*-methyl-D-erythritol-4-phosphate; CDP-ME2P: 4-(cytidine-5'-diphospho)-2*C*-methyl-D-erythritol; ME-CPP: 2-C-phospho-4-(cytidine-5'-diphospho)-2*C*-methyl-D-erythritol; HMBPP: 1-hydroxyl-2-methyl-2-butenyl-4-diphosphate; DMAPP: dimethylallyl-PP; IPP: isopentenyl-diphosphate; GPP: geranyl diphosphate; FPP: farnesyl diphosphate; GGPP: geranylgeranyl diphosphate/geranylgeranyl pyrophosphate; MVA: mevalonate; MVA-3P: mevalonate-3P.

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

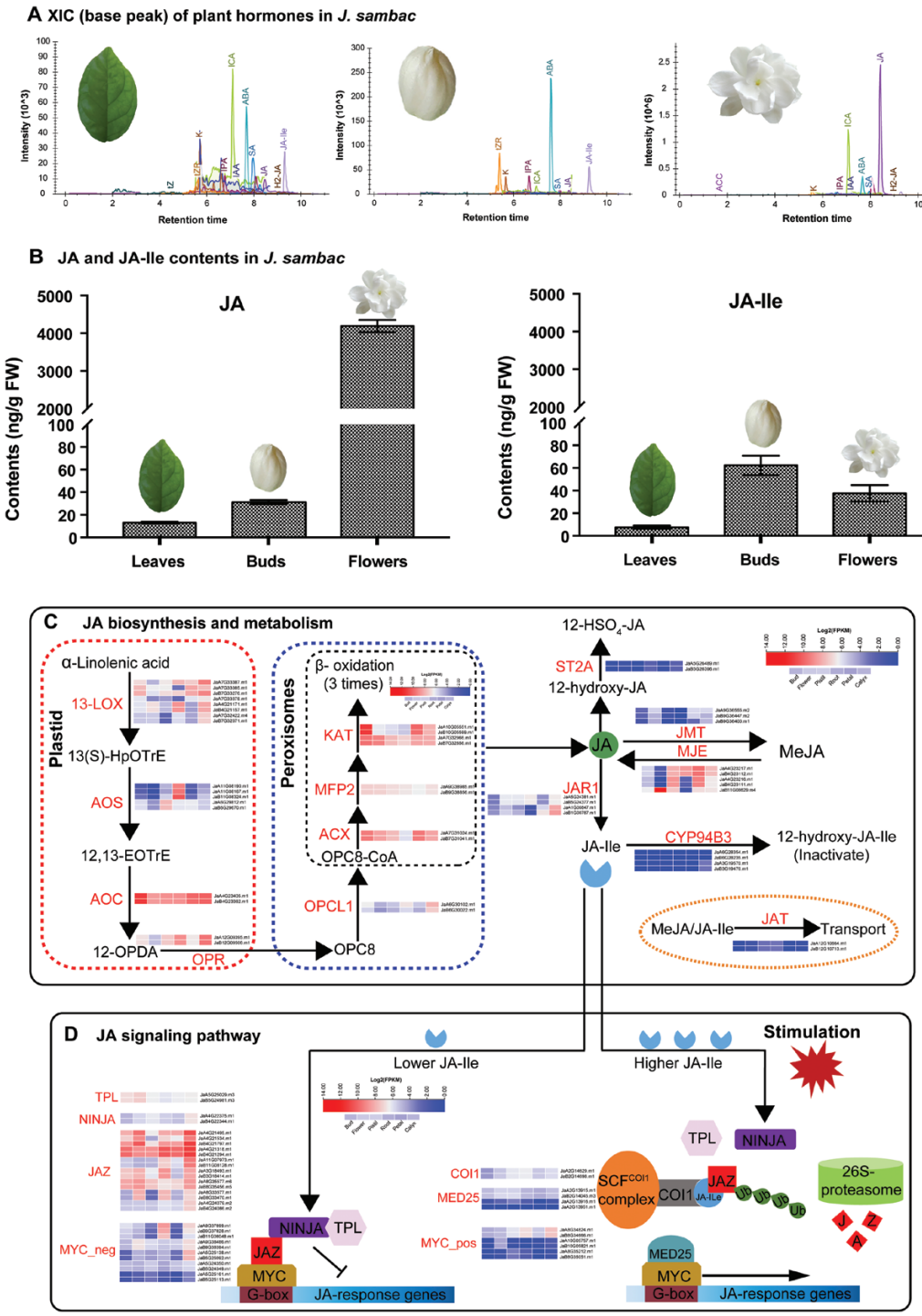

Fig. 5. JA biosynthesis and consumption of *J. sambac*. (A) XIC (base peak of extracted ion chromatogram) of plant hormones in young leaves, buds, and fully-opened flowers of *J. sambac* by LC-MS. (B) JA and JA-Ile content (ng/g) in *J. sambac*. (C) Genes and their expression associated with JA biosynthesis and metabolism. (D) Genes and their expression associated with JA signalling. Abbreviations: LOX: lipoxidase; AOS: allene oxide synthase; AOC: allene oxide cyclase; OPR: 12-oxophytodienoic acid reductase; OPCL1: OPC-8:0 CoA ligase 1; ACX: acyl-CoA oxidase; MFP: multifunctional proteins/3-hydroxyacyl-CoA dehydrogenase; KAT: 3-ketoacyl-CoA thiolase; JAR: JA-amino acid synthetase; JA-Ile: jasmonoyl-L-isoleucine; ST2A: sulfotransferase; JMT: jasmonic acid carboxyl methyltransferase; MJE: methyl jasmonate esterase; 13(S)-HpOTrE/13-HPOT: 13s-hydroperoxyotadecatrienoic acid; 12-OPDA: 12-oxophytodienoic acid; OPC8: OPC-8:0 CoA; JAT: jasmonate transporter; TPL: TOPLESS; NINJA: Novel Interactor of JAZ; COI1: CORONATINE INSENSITIVE1; JAZ: Jasmonate ZIM domain proteins, mainly consist of ZIM (zinc-finger-inflorescence meristem, domain of JAZ for binding to NINJA and for homo- and heteromerization) and Jas (domain of JAZ for binding to COI1, MYC2 and other TFs) domains; SCF: SKP/CULLIN/F-BOX; MED25: Mediator complex 25.

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

In the initial steps of the JA signalling pathway, jasmonoyl-L-isoleucine synthetase 1 (*JAR1*), which contributes to the formation of the JA-Ile conjugate (bioactive JA) (Hui *et al.*, 2019), had lower expression (*JAR1* FPKM=~10 in buds, ~56 in flowers, ~78 in petals) compared to genes associated with JA biosynthesis (Fig. 5C). Genes encoding the SCF^{COI1} complex, which perceives JA-Ile conjugate ligands via *COI1*, also showed lower expression compared to genes associated with JA biosynthesis (Fig. 5C). The TIFY motif of JAZ proteins, which is necessary for binding with the NINJA-TPL co-repressor complex to repress JA signalling via repressing the transcription factor *MYC2* to bind with a G-box promoter (Pauwels *et al.*, 2010), showed higher expression compared with genes associated with JA biosynthesis (Fig. 5C). *JAZ* is a repressor of JA signalling and the synthetic construct based on Jas9-VENUS is used as a biosensor in rice to detect the content and distribution of JA (Li *et al.*, 2021). High expression of *JAZ* with the TIFY10 domain in roots and calyxes of *J. sambac* may play important roles in maintaining the balance of JA and storage of resources used in response to various stimuli. The *MYC* family of genes not only have positive roles in activating JA-response genes, but also in repressing the JA-response genes targeted by JAZ proteins (Fig. 5D). We extracted the upstream 2000 bp sequence of the promoters of the *MYC* family genes and predicted the *cis*-acting elements of the promoters. Results showed that most MYC promoters contained a G-box element and a CAAT-box/CAT-box (Supplementary Fig. S17). These *cis*-acting elements are mainly involved in abscisic acid responsiveness, MeJA responsiveness, wound responsiveness, salicylic acid responsiveness, etc. (Supplementary Fig. S17). Notably, genes encoding *NINJA*, *TPL*, and *MED25* co-repressors also showed low expression. All of these results indicated that the JA-mediated signalling pathway was usually maintained in the resting stage (Fig. 5C). The jasmonate transporter gene (*JAT*), which mediates MeJA/JA-Ile nuclear entry and maintains critical nuclear JA-Ile concentration for activating JA signalling (Li *et al.*, 2017), showed almost no expression in *J. sambac* under typical growing conditions. High JA synthesis coupled with low degradation resulted in accumulation of abundant JA under typical environmental conditions in *J. sambac* (Fig. 5D).

Furthermore, the expression profiles of genes related to JA biosynthesis were consistent with that of aromatic genes in the flower organs (buds, flowers, pistils, and petals), suggesting that JAs might also participate in flower development and floral scent production. Abundant aromatic volatiles and JA in *J. sambac* may be associated with anatomical development, robust resistance to abiotic and biotic stresses, and environmental adaptation. *Ka/Ks* analysis showed that the genes associated with JA were non-positively selected, and there were no significant differences in gene expression in SP, DP, and MP jasmines (Supplementary Fig. S18). The expression results in *J. sambac* indicated that the jasmine plants had abundant JA, regardless of SP, DP, or MP flowers. A cross-species comparison of the key genes involved in JA biosynthesis and consumption showed that there are similar expression patterns in *J. sambac* and *O. fragrans* (Fig. 6; Supplementary Fig. S19).

Evolution of JA metabolic genes in J. sambac

Ash (*F. excelsior*; *Oleaceae*) also accumulated abundant JAs (Chen *et al.*, 2016), whereas sesame (*S. indicum*; *Lamiales*) did not (Mehmood *et al.*, 2021) (Supplementary Table S8). Phylogenetic analysis revealed expansion of key JA biosynthesis genes such as *LOX* and *KAT*, possibly due to the *Oleaceae*-wide WGT event (Supplementary Fig. S20A, No. 1). By contrast, the key JA consumption gene *JAR1* showed significant contraction in species of the *Oleaceae* compared with sesame in the *Lamiales* (Supplementary Fig. S20C, No. 3). Moreover, the expanded and contracted genes showed higher and lower expression in *J. sambac*, respectively, compared with genes without significant copy-number variations (Fig. 5C). These results suggested that both gene gain/loss events (e.g. the WGT) and regulatory changes contributed to JA accumulation in *J. sambac*.

The PAL pathway contributes to SA accumulation in J. sambac

Salicylic acid (SA) is also an essential signalling molecule in seed germination, flower development, and responses to biotic and abiotic stresses (Zheng *et al.*, 2009). Young leaves and fully-opened flowers of *J. sambac* had high SA levels under typical growing conditions (Supplementary Figs S15, S21A). SA is synthesized either from PAL or from chorismate via isochorismate in some bacteria under the catalysis of isochorismate synthase (ICS) and isochorismate pyruvate lyase (IPL) (Shah, 2003). The recent discovery of PBS3 and EPS1 completed the missing steps of salicylic acid biosynthesis from isochorismate (IC) in Arabidopsis (Torrens-Spence *et al.*, 2019). We identified the genes associated with SA biosynthesis in *J. sambac* (Supplementary Fig. S21B) in both the PAL and IC pathways. Our results revealed that PAL, abnormal inflorescence meristem 1 (AIM1, Supplementary Fig. S21C) in the PAL pathway, and the *MES* (Supplementary Fig. S16) and SA methyltransferase (*BSMT*) genes, which catalyse the interconversion of SA and MeSA, were highly expressed. However, the *ICS*, *PBS3*, and *EPS1* genes in the alternative IC pathway of SA biosynthesis showed lower expression levels. MEME analysis identified a motif (ASINQYKTPRCVK-FAPIVELLBSRVVSTHFSPKCPKWTPER) which may be critical for *PBS3* function (Supplementary Fig. S22), a GH3 acyl adenylase-family enzyme for SA accumulation (Nobuta *et al.*, 2007). Biosynthetic SA can be converted to 2, 3-dihydroxy-benzaic acid (2, 3-DHBA) and 2, 5-dihydroxy-benzaic acid (2, 5-DHBA) by the SA 3-hydroxylase (S3H) and SA 5-hydroxylase (S5H) enzymes, respectively (Supplementary Fig. S23). Moreover, SA also can be converted to SA-glucose ester (SGE) or SA-glucoside/SA O-β-glucoside (SAG) by SA glucosyltransferase (SAGT) enzymes. In this study, the genes

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

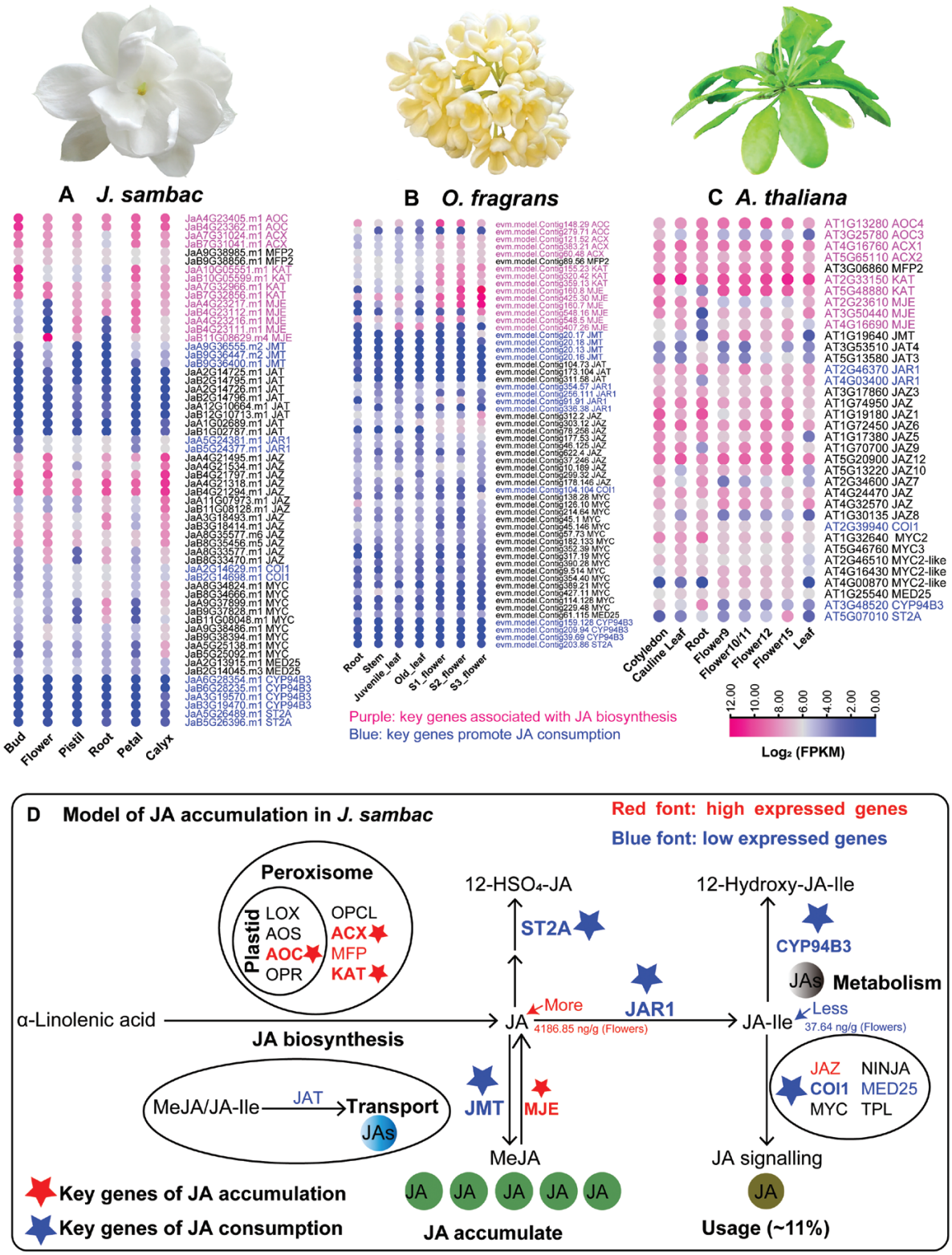

Fig. 6. Key genes involved in JA biosynthesis and consumption in *J. sambac* (A), *O. fragrans* (B), and *A. thaliana* (C). (D) Model of JA accumulation in *J. sambac*.

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

encoding *S3H*, *S5H*, and *SAGT* (*UGT74F1/UGT74F2*) had low expression in *J. sambac* (Supplementary Fig. S21B). In addition, the SA receptors of the non-expresser of pathogenesis related family (NPR) also displayed a low expression level (Supplementary Figs S21B, S24). These results demonstrated that the PAL pathway probably contributes to SA accumulation in *J. sambac.* Similar to the molecular mechanism of elevated levels of JA in *J. sambac*, the accumulation of SA is also a net result of more production and less consumption.

In most plants, JA or its derivatives will typically increase when the plant is faced with various stresses increasing JA biosynthesis related genes, whereas in *J. sambac* JA can accumulate under normal growing conditions. Our results showed that JA accumulation in *J. sambac* was possibly caused by increased biosynthesis accompanied with decreased consumption degradation in *J. sambac*. High JA synthesis coupled with low degradation resulted in accumulation of abundant JA in *J. sambac* under typical environmental conditions. The accumulation of SA also utilized the mechanism of more accumulation and less consumption, which may be the apparatus for more hormone accumulation under natural conditions. This study deciphered the biosynthetic pathways of the floral fragrance developed by JA and SA in *J. sambac*, and is an important step forward to utilize this natural product. *J. sambac* is an economically important economically horticultural species with high value in industrial applications, gardens, and landscaping (Supplementary Fig. S25).

Supplementary data

The following supplementary data are available at *JXB* online.

Fig. S1. Phenotypic characteristics of *J. sambac* and workflow for genome assembly, error correction, phasing and anchoring.

Fig. S2. Flow cytometry analysis to estimate the genome size of *J. sambac*.

Fig. S3. Genome size estimation of *J. sambac* by GenomeScope1.0 via different k-mers (k17, k19, k21, k27, k47).

Fig. S4. Genome size estimation of *J. sambac* by GenomeScope2.0 via different k-mers (k17, k19, k21, k27, k47).

Fig. S5. BUSCO assessments of *J. sambac* assembly and annotation.

Fig. S6. Interaction frequency distribution of Hi-C links among chromosomes in the HA and HB haplotypes of *J. sambac*.

Fig. S7. Biallelic differentiation of *J. sambac*.

Fig. S8. Structure variant (SV) types and counts in HB versus HA (HA was set as a reference).

Fig. S9. Numbers of OGs among the *Oleaceae* species.

Fig. S10. The *J. sambac* evolutionary process.

Fig. S11. Expression of MADS-box genes associated with MP formation.

Fig. S12. The location of *TPS* genes in the pseudochromosomes of HA and HB haplotypes.

Fig. S13. The duplicated *TPS* sequence alignment comparison in HA and HB.

Fig. S14. The *BEAT* genes involved in benzyl acetate biosynthesis.

Fig. S15. Heatmap and correlation analysis of hormone metabolites in buds, leaves, and flowers of *J. sambac*.

Fig. S16. The phylogenetic tree of methyl salicylate esterase (*MES*) genes.

Fig. S17. Prediction of *cis*-acting elements in promoters of MYC transcription factors.

Fig. S18. Comparison of candidate gene expression involved in JA biosynthesis and consumption in SP, DP, and MP *J. sambac*.

Fig. S19. Comparison of candidate gene expression involved in JA biosynthesis and consumption in *J. sambac* (A), *O. fragrans* (B), and *A. thaliana* (C).

Fig. S20. The WGT event contributes to JA accumulation in the *Oleaceae*.

Fig. S21. SA and MeSA biosynthesis in *J. sambac*.

Fig. S22. Phylogeny tree and structure prediction of the *PBS3* gene family identified in *Oleaceae* and six other sequenced plants.

Fig. S23. The phylogenetic tree of SA 3-hydroxylase and SA 5-hydroxylase genes.

Fig. S24. The phylogenetic tree of SA receptor NPR family genes.

Fig. S25. Abundant JA and aromatic volatile biosynthesis in *J. sambac*.

Table S1. The RNA-seq information of SP, DP, and MP jasmines.

Table S2. The estimated results by flow cytometry analysis.

Table S3. The estimated results by different genome size estimated tools.

Table S4. BUSCO results of genome assembly and annotation.

Table S5. Chromosome length and the numbers of coding-protein genes in the HA and HB haplotype, and the Jas consensus genome.

Table S6. Comparisons of genome assemblies and annotations of *J. sambac*.

Table S7. Genes associated with aromatic volatile accumulation in Jas-vs-SP and Jas-vs-MP.

Table S8. Concentrations of JA and SA in sesame, ash, and *J. sambac*.

Acknowledgements

We thank Yimin Dang for photos of flowers and also thank Deshu Lin and Haifeng Wang for their suggestions supporting this project.

Author contributions

LZ conceived the study and revised the manuscript. XZ generated the phased haplotype of *J. sambac*. MX performed the data analysis and drafted the manuscript. XZ and QG assembled and annotated the consensus

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

genome of *J. sambac*. CJ, JW, FL, MJ, WW, JH, XC, and DL participated in data analysis and discussion. YJ, FC, CL, HH, and FC supervised the work and provided scientific advice. RNT and YL edited the manuscript. All authors contributed to writing and editing the final manuscript.

Conflict of interest

The authors declare no conflict of interest.

Funding

LZ thanks the National Natural Science Foundation of China (32272750). FC acknowledges funding from the National Natural Science Foundation of China (31801898).

Data availability

The data generated and datasets analysed to support the results of this study are available from the corresponding authors upon request. The raw data, assembly, and annotated data have been deposited in the National Genomics Data Center (PRJCA006075).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology **215**, 403–410.

Bera P, Mukherjee C, Mitra A. 2017. Enzymatic production and emission of floral scent volatiles in *Jasminum sambac*. Plant Science **256**, 25–38.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**, 2114–2120.

Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. 2013. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. Nature Biotechnology **31**, 1119–1125.

Chen N. 2004. Using RepeatMasker to identify repetitive elements in genomic sequences. In: Chapter 4, Unit 4.10, Current Protocols Bioinformatics. USA: John Wiley & Sons.

Chen Y, Ulyshen MD, Poland TM. 2016. Abundance of volatile organic compounds in white ash phloem and emerald ash borer larval frass does not attract *Tetrastichus planipennisi* in a Y-tube olfactometer. Insect Science **23**, 712–719.

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Molecular Plant **13**, 1194–1202.

Chen XJ, Wang P, Gu M, Hou B, Zhang C, Zheng Y, Yun Sun Y, Jin S, Ye N. 2022. Identification of PAL genes related to anthocyanin synthesis in tea plants and its correlation with anthocyanin content. Horticultural Plant Journal **8,** 381**–**394.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. Nature Methods **18**, 170–175.

Chini A, Fonseca S, Fernández G, *et al*. 2007. The JAZ family of repressors is the missing link in jasmonate signalling. Nature **448**, 666–671.

Demole E, Lederer E, Mercier D. 1962. Isolement et détermination de la structure du jasmonate de méthyle, constituant odorant caractéristique de l'essence de jasmin. Helvetica Chimica Acta **45**, 675–685.

Dudareva N, D'Auria JC, Nam KH, Raguso RA, Pichersky E. 1998. Acetyl-CoA:benzylalcohol acetyltransferase - an enzyme involved in floral scent production in *Clarkia breweri*. The Plant Journal **14**, 297–304.

Dudchenko O, Batra SS, Omer AD, *et al*. 2017. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science **356**, 92–95.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology **16**, 157.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology **20**, 238.

Fu ZQ, Dong X. 2013. Systemic acquired resistance: turning local infection into global defense. Annual Review of Plant Biology **64**, 839–863.

Garg S, Fungtammasan A, Carroll A, *et al*. 2020. Chromosome-scale, haplotype-resolved assembly of human genomes. Nature Biotechnology **3**, 309–312.

Gfeller A, Dubugnon L, Liechti R, Farmer EE. 2010. Jasmonate biochemical pathway. Science Signaling **3**, cm3.

Haas BJ, Delcher AL, Mount SM, *et al*. 2003. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Research **31**, 5654–5666.

Hui S, Hao M, Liu H, Xiao J, Li X, Yuan M, Wang S. 2019. The group I GH3 family genes encoding JA-Ile synthetase act as positive regulator in the resistance of rice to *Xanthomonas oryzae* pv. *oryzae*. Biochemical and Biophysical Research Communications **508**, 1062–1066.

Kamali S, Singh A. 2022. Jasmonates as emerging regulators of plants response to variable nutrient environment. Critical Reviews in Plant Sciences **41**, 271–285.

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

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology **14**, R36.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biology **5**, R12.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics **26**, 589–595.

Li Q, Zheng J, Li S, Huang G, Skilling SJ, Wang L, Li L, Li M, Yuan L, Liu P. 2017. Transporter-mediated nuclear entry of jasmonoyl-isoleucine is essential for jasmonate signaling. Molecular Plant **10**, 695–708.

Li S, Cao L, Chen X, Liu Y, Persson S, Hu J, Chen M, Chen Z, Zhang D, Yuan Z. 2021. Synthetic biosensor for mapping dynamic responses and spatio-temporal distribution of jasmonate in rice. Plant Biotechnology Journal **19**, 2392–2394.

Lomsadze A, Burns PD, Borodovsky M. 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Research **42**, e119.

Malamy J, Carr JP, Klessig DF, Raskin I. 1990. Salicylic acid: a likely endogenous signal in the resistance response of tobacco to viral infection. Science **250**, 1002–1004.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics **27**, 764–770.

Mehmood M, Pérez-Llorca M, Casadesús A, Farrakh S, Munné-Bosch S. 2021. Leaf size modulation by cytokinins in sesame plants. Plant Physiology and Biochemistry **167**, 763–770.

Mousavi SA, Chauvin A, Pascaud F, Kellenberger S, Farmer EE. 2013. GLUTAMATE RECEPTOR-LIKE genes mediate leaf-to-leaf wound signalling. Nature **500**, 422–426.

Nobuta K, Okrent RA, Stoutemyer M, Rodibaugh N, Kempema L, Wildermuth MC, Innes RW. 2007. The GH3 acyl adenylase family member PBS3 regulates salicylic acid-dependent defense responses in Arabidopsis. Plant Physiology **144**, 1144–1156.

Ou S, Chen J, Jiang N. 2018. Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Research **46**, e126.

Pauwels L, Barbero GF, Geerinck J, *et al*. 2010. NINJA connects the co-repressor TOPLESS to jasmonate signalling. Nature **464**, 788–791.

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology **33**, 290–295.

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023

Raskin I. 1992. Salicylate, a new plant hormone. Plant Physiology **99**, 799–803.

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen CJ, Vert JP, Heard E, Dekker J, Barillot E. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biology **16**, 259.

Shah J. 2003. The salicylic acid loop in plant defense. Current Opinion in Plant Biology **6**, 365–371.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**, 3210–3212.

Sollars ES, Harper AL, Kelly LJ, *et al*. 2017. Genome sequence and genetic diversity of European ash trees. Nature **541**, 212–216.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics **24**, 637–644.

Thines B, Katsir L, Melotto M, Niu Y, Mandaokar A, Liu G, Nomura K, He SY, Howe GA, Browse J. 2007. JAZ repressor proteins are targets of the SCF(COI1) complex during jasmonate signalling. Nature **448**, 661–665.

Torrens-Spence MP, Bobokalonova A, Carballo V, Glinkerman CM, Pluskal T, Shen A, Weng JK. 2019. PBS3 and EPS1 complete salicylic acid biosynthesis from isochorismate in Arabidopsis. Molecular Plant **12**, 1577–1586.

Unver T, Wu Z, Sterck L, *et al*. 2017. Genome of wild olive and the evolution of oil biosynthesis. Proceedings of the National Academy of Science, USA **114**, E9413–E9422.

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. 2017. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics **33**, 2202–2204.

Wang P, Fang J, Lin H, *et al*. 2022. Genomes of single- and double-petal jasmines (*Jasminum sambac*) provide insights into their divergence time and structural variations. Plant Biotechnology Journal **20**, 1232–1234.

Wang Y, Tang H, Debarry JD, *et al*. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Research **40**, e49.

Warren RF, Merritt PM, Holub E, Innes RW. 1999. Identification of three putative signal transduction genes involved in R gene-specified disease resistance in Arabidopsis. Genetics **152**, 401–412.

Wheeler TJ, Eddy SR. 2013. nhmmer: DNA homology search with profile HMMs. Bioinformatics **29**, 2487–2489.

Xu S, Ding Y, Sun J, Zhang Z, Wu Z, Yang T, Shen F, Xue G. 2022. A high-quality genome assembly of *Jasminum sambac* provides insight into floral trait formation and Oleaceae genome evolution. Molecular Ecology Resources **22**, 724–739.

Yang X, Yue Y, Li H, Ding W, Chen G, Shi T, Chen J, Park MS, Chen F, Wang L. 2018. The chromosome-level quality genome provides insights into the evolution of the biosynthesis genes for aroma compounds of *Osmanthus fragrans*. Horticulture Research **5**, 72.

Yang Y, Yang XH, Guo X, Hu X, Dong DF, Li GG, Xiong XY. 2022. Exogenously applied methyl jasmonate induces early defense related genes in response to *Phytophthora infestans* infection in potato plants. Horticultural Plant Journal **8**, 511–526.

Zeng LH, Hu M, Yan YM, Lu Q, Cheng YX. 2012. Compounds from the roots of *Jasminum sambac*. Journal of Asian Natural Products Research **14**, 1180–1185.

Zhang LS, Chen F, Zhang XT, *et al*. 2020. The water lily genome and the early evolution of flowering plants. Nature **577**, 79–84.

Zheng Z, Qualley A, Fan B, Dudareva N, Chen Z. 2009. An important role of a BAHD acyl transferase-like protein in plant innate immunity. The Plant Journal **57**, 1040–1053.

Zhu Z, Napier R. 2017. Jasmonate - a blooming decade. Journal of Experimental Botany **68**, 1299–1302.

Downloaded from https://academic.oup.com/jxb/article/74/4/1275/6847707 by National Science & Technology Library user on 29 March 2023