New Phytologist

Research

The chromosome-level genome assembly of *Cananga odorata* provides insights into its evolution and terpenoid biosynthesis

Yan Zheng[1,2]*, Danni Yang[1,2]*, Xin Yin[1,2], Xingyu Yang[1,2], Mingyue Chen[2,3], Xieshengyang Li[3], Tianyu Yang[2,4], Joeri Sergej Strijk[5], Damien Daniel Hinsinger[5,6], Yunqiang Yang[1,2], Xiangxiang Kong[1,2] and Yongping Yang[1,2,4]

[1]CAS Key Laboratory of Tropical Plant Resources and Sustainable Use, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Mengla, Yunnan, 666303, China; [2]Germplasm Bank of Wild Species, Yunnan Key Laboratory for Crop Wild Relatives Omics, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, 650201, China; [3]School of Agriculture, Yunnan University, Kunming, 650091, China; [4]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China; [5]Alliance for Conservation Tree Genomics, Pha Tad Ke Botanical Garden, PO Box 959, Luang Prabang, 06000, Lao PDR; [6]Université Paris-Saclay, Centre INRAE Île-de-France Versailles-Saclay, EPGV, Evry, F-91057, France

Authors for correspondence:
Yongping Yang
Email: yangyp@xtbg.ac.cn

Xiangxiang Kong
Email: kongxiangxiang@mail.kib.ac.cn

Yunqiang Yang
Email: yangyunqiang@xtbg.ac.cn

Received: *17 November 2023*
Accepted: *4 July 2024*

New Phytologist (2024) **243:** 2279–2294
doi: 10.1111/nph.19977

Key words: *Cananga odorata*, genome assembly, magnoliids evolution, sesquiterpenoid biosynthesis, *TPS*.

Summary

- *Cananga odorata* is known as a natural perfume tree of the Annonaceae family in Magnoliales. However, its phylogenetic position and the molecular mechanisms involved in the biosynthesis of the floral volatile organic compounds (VOCs) remain unclear.
- Here, by combining a variety of sequencing platforms, we present a telomere-to-telomere (T2T) genome of *C. odorata* with 735.83 Mb, which represents the highest integrity and assembly quality of genome in magnoliid plants reported to date. Phylogenetic analysis based on multiple datasets and approaches showed that *C. odorata*, as a member of magnoliids, is sister to eudicots, after their divergence from monocots.
- Metabolomic of VOCs in the essential oil and flowers scent showed that sesquiterpenes, especially β-caryophyllene, were the major compounds. Two *CoTPS21* homologues derived from tandem duplication events were highly expressed during flower development and were identified as the key sesquiterpene synthases for the production of β-caryophyllene. In addition, CoSPL3 and CoSPL9 were considered as potential transcription factors for activating the expression of *CoTPS21* homologues.
- Our results shed light on the molecular mechanisms underlying the biosynthesis of the unique floral fragrance in *C. odorata* and provide new insights into the phylogenetic position of magnoliids.

Introduction

Cananga odorata, commonly known as ylang–ylang, is a perennial tropical evergreen tree of the Annonaceae family, native to Indonesia and widely cultivated in tropical regions of Southeast Asia. As an important resource plant, *C. odorata* is popularly known as the 'king' of floral scent and a natural perfume tree, with important economic value and commercial prospects (Benini *et al.*, 2012; Tan *et al.*, 2015). Essential oil extracted from the fresh and mature flowers of *C. odorata* is a precious natural flavor essence that is widely used as the most important raw material in the cosmetics and perfume industries (Qin *et al.*, 2014). Previous studies have shown that the volatile organic compounds (VOCs) in *C. odorata* mainly consist of volatile sesquiterpenes, monoterpenes, aromatic benzenoids, ester and ether (Benini *et al.*, 2012; Chakira *et al.*, 2022). The major volatile components are β-caryophyllene, D-germacrene, (E,E)-α-farnesene, p-cresyl methyl ether, benzyl acetate, methyl benzoate, and geranyl acetate (Qin *et al.*, 2014; Chakira *et al.*, 2022). However, the floral metabolic production and the molecular mechanisms involved in the regulation of biosynthesis process of the floral VOCs of *C. odorata* are not well understood.

Cananga odorata belongs to Magnoliales, which together with Canellales, Laurales and Piperales constitute the magnoliid clades (Massoni *et al.*, 2014). The magnoliids is very ancient in plant evolution, and their phylogenetic position facilitates the resolution of the evolutionary process of existing flowering plants (Soltis *et al.*, 2015). However, the phylogenetic position of magnoliids with respect to monocots and eudicots still remains confusing despite previous valuable attempts (Chen *et al.*, 2019; Hu *et al.*, 2019; Lv *et al.*, 2020; Y. C. Chen *et al.*, 2020; Qin *et al.*, 2021). Genome data has been a powerful means for

*These authors contributed equally to this work.

resolving these uncertainties, with recent successive publications of the genomes of magnoliid plants. For instance, (1) genome analysis of some Lauraceae plants (Chaw *et al.*, 2019; Shen *et al.*, 2022; Y. C. Chen *et al.*, 2020) supports magnoliids as a sister group to eudicots; (2) analysis of the Aristolochiaceae plant *Aristolochia fimbriata* suggests the placement of magnoliids as sister to monocots (Qin *et al.*, 2021); and (3) the genomes of *Magnolia biondii* (Magnoliaceae) (Dong *et al.*, 2021), *Liriodendron chinense* (Magnoliaceae) (Chen *et al.*, 2019) and *Persea americana* (Lauraceae) (Rendón-Anaya *et al.*, 2019) suggest that magnoliids are sister to both eudicots and monocots. Therefore, the genome evolution among magnoliids is an important topic that is currently popularly studied but is still not fully resolved. As the second largest family in Magnoliales and one of the most species-rich pantropical plant families, Annonaceae has only two represented draft genomes so far (Strijk *et al.*, 2021; Talavera *et al.*, 2023). Genomic resources with representativeness and high quality for Annonaceae plants would facilitate the phylogenetic studies on angiosperm relationships and evolution.

In this study, we report a T2T genome of *C. odorata*. Comparative genomic and phylogenetic analyses with other angiosperms have contributed to figure out the phylogenetic position of magnoliids. We determined the major components of floral volatile compounds and their biosynthetic pathways and deciphered the molecular genetics underlying the biosynthesis of the main fragrant volatile compound products.

Materials and Methods

Plant material

Fresh leaves and flowers of ylang–ylang (*Cananga odorata* (Lamk.) Hook. f. & Thomson.) were collected from an individual plant cultivated at Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences (21°55′11″N, 101°15′27″E) in April 2021 and May 2022. Fresh leaves were used to extract DNA for further sequencing. The fresh flowers were used to extract essential oil and extract RNA for further transcriptome analysis.

Library construction and sequencing

For Oxford Nanopore Technologies (ONT) ultralong sequencing, genomic DNA (gDNA) was isolated using the CTAB approach from leaves of *C. odorata*. DNA purity and quality were measured using a NanoDrop™ One UV–V is spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Approximately 8–10 μg of size-selected (> 50 kb) gDNA fragments were constructed and sequenced on the PromethION platform (Oxford Nanopore Technologies, Oxford, UK) at the Genome Center of Grandomics (Wuhan, China) (Wang *et al.*, 2021). For HiFi sequencing, *c.* 8 μg of gDNA was used for library preparations followed by size screening. A PacBio Sequel II instrument was employed for sequencing. For NGS, libraries were constructed by using the MGIEasy Universal DNA Library Prep Kit v.1.0 protocol with the extracted gDNA. For Hi-C sequencing, fresh leaves were cross-linked with formaldehyde (2%) and digested with the restriction endonuclease DpnII for the Hi-C library construction (Rao *et al.*, 2014).

Genome assembly and gap filling

The first 65× reads from ONT ultralong sequencing were used to perform the draft assembly with NextDenovo v.2.4.0 (input type = raw; read type = ont; read cut-off = 1 k; seed cut-off = 99 365) (Hu *et al.*, 2023). Initially assembled reference genomes were calibrated with HiFi reads and Illumina reads via NextPolish (task = 55 512 121 212; lgs options = -min_read_len 1 k -max_depth 60; lgs_minimap2_options = -x map-ont; sgs_options = -max_depth 100 -bwa) to obtain the assembled genomes (Hu *et al.*, 2020). Purge_dups v.1.2.5 was used to remove redundant contigs, producing relatively complete scaffolds (Guan *et al.*, 2020). For pseudomolecule assembly, raw Hi-C reads were subjected to adapter removal and trimmed for low-quality bases using fastp v.0.20.0 with default parameters (Chen *et al.*, 2018). Clean reads were mapped to the contig assembly using Bwa-Mem2 v.2.2 (Li & Durbin, 2009). The deduplicated list of alignments of Hi-C reads to the contig assembly was generated using juicer v.1.6 (Durand *et al.*, 2016a). The contig assembly was then fixed to the chromosome length scaffold using 3D-DNA v.201008 (Dudchenko *et al.*, 2017). Heatmaps of Hi-C interactions were performed by the 3D-DNA visualization module and manually adjusted for inversion and misalignment assembly errors with Juicebox v.1.11.08 to finally obtain the chromosome-level genomes of *C. odorata* (Durand *et al.*, 2016b).

Genome annotation

De novo repeat identification was conducted using Repeat Modeler. The following tools were used for various repetitive sequences: LTR_FINDER and LTR harvest for LTR retrotransposons, Mite Hunter for mite repeats, HelitronScanner for helitrons, SINE-Finder for short interspersed nuclear elements (SINEs), and TARGeT for terminal inverted repeats (Price *et al.*, 2005).

Structural annotation of protein-coding genes was performed using *de novo* prediction, homology-based prediction, and transcriptome-based prediction in Braker v.2.1.6 with the soft-masked genome (Brůna *et al.*, 2021). Query protein sets were obtained from OrthoDB (v.10), which included sets from *Aristolochia fimbriata* Cham. (Aristolochiaceae), *Liriodendron chinense* (Hemsl.) Sarg. (Magnoliaceae), and *Magnolia officinalis* Rehder & E. H. Wilson (Magnoliaceae) (Kriventseva *et al.*, 2019). RNA-Seq reads were mapped to the genome assembly using Hisat2 (v.2.2.1) (Kim *et al.*, 2019) and minimap2 (v.2.22-r1101) (Li, 2018). ProtHint (v.2.5.0) was used to obtain protein hints (Brůna *et al.*, 2020). Finally, gene prediction was conducted in Augustus (v.3.3.3) and GeneMark-ET (v.4.38) (Brůna *et al.*, 2020).

For functional annotation, BLAST (blastX and blastn; e-value < 1e−10) analysis was performed to search against

multiple databases including Clusters of Orthologous Genes (COG), Protein family (Pfam), Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Swiss-Prot, and Nonredundant (NR) databases. Blast2GO was used for GO annotation. The completeness of the assembly and annotation was evaluated using Benchmarking Universal Single-Copy Orthologs (BUSCO, v.4.0) analysis (Simão *et al.*, 2015). Furthermore, for Biosynthetic Gene Clusters (BGCs) identification, the genome data and genome annotation files of *C. odorata* was submitter to PlantiSMATH website (http://plantismash.secondarymetabolites.org/).

Centromere and telomere characterization

Satellite and transposable element positions were extracted from the output of RepeatMasker. TRF v.4.09.1 (2 7 7 80 10 50 2000 -h) was used to identify tandem repeat sequences in centromere regions. VSEARCH v.2.22.1 (--clusterout_sort --clusterout_id --fasta_width 0 --id 0.8 --cluster_size) was used to cluster homologous CRC sequences on chromosomes. LASTZ ((multiple) --format = general) was used to compare similarities between candidate sequences and to locate CRCs and centromeric singletons on chromosomes. We considered a LASTZ comparison to have high confidence if both the coverage and recognition rates were *c.* 80% in that comparison. The positional schemtic was generated by PYGENOMETRACKS v.3.8. RECTCHR v.1.35 was used to display the columnar homology blocks detected between homologous chromosomes by MUMMER v.4.0.0rc1.

Genome structural comparison analysis

JCVI software was used to identify the shared chromosomal rearrangements among *C. odorata* and other species (Tang *et al.*, 2008) The comparative synteny module was used to generate a collinearity point map and to identify syntenic blocks with the maximum gap between anchor genes set to 30. Then, the genome rearrangement regions and the connections between these regions in different species were determined, and the ancestral connection mode of the main branches of angiosperms was constructed to compare with the prediction mode of the common ancestor of the existing angiosperms.

Phylogenetic analysis

The phylogenetic analyses were conducted using different datasets and methods to verify the evolutionary relationships of magnoliids with 23 plants, including 5 eudicots, 5 monocots, 6 magnoliids, 3 amborellales nymphaeales austrobaileyales (ANA)-clade members, 1 Ceratophyllale, 1 Chloranthale, *Selaginella moellendorffii* Hieron. and an outgroup gymnosperm. Two methods were used to screen orthologous groups based on single-copy gene (SSCG) families and mostly single-copy gene (MSCG) families. To reconstruct the phylogeny, we aligned the protein sequences of each gene family using MUSCLE v.3.8.31 (Edgar, 2004), and then forced the nucleotide sequences onto the amino acid alignment using PAL2NAL v.14 (Suyama *et al.*, 2006). We also forced the nucleotide sequences onto the amino acid alignment to obtain codon preserving nucleotide sequence alignment results. Finally, five different alignments including amino acid, nucleotide and codon1/2/3 alignments for each gene family were performed for phylogenetic analyses. The concatenation-based analyses and coalescent-based analyses were performed as previously described (Qin *et al.*, 2021).

Genomic synteny analysis and gene duplication identification

We analyzed the Ks distributions for whole genome duplication (WGD) events in *Cinnamomum micranthum* (Hayata) Hayata (Lauraceae), *C. odorata*, *Chimonanthus salicifolius* H. H. Hu (Calycantaceae), *Amborella trichopoda* Baill. (Amborellaceae) and *Vitis vinifera* L. (Vitaceae). KaKs_calculator was used to calculate Ks values (Zhang *et al.*, 2006). The covariance between *C. micranyhum*, *C. salicifolius*, *A. trichopoda* and *C. odorata* was determined using WGDI after accurate calculation of the longest protein sequence of each gene (P. Sun *et al.*, 2022). DupGene_finder was used to further classify the duplicate gene pairs in the *C. odorata* into different duplication patterns (Qiao *et al.*, 2019). Calculate_4DTv_correction.pl was used to calculate fourfold degenerate synonymous site (4DTv) values for orthologous and paralogous genes between *C. odorata* and other indicated species (https://github.com/JinfengChen/Scripts/blob/master/FFgenome/03.evolution/distance_kaks_4dtv/bin/calculate_4DTV_correction.pl).

Karyotype analysis

The AAK and AEK were downloaded from public websites (https://github.com/SunPengChuan/Angiosperm-karyotype-evolution/tree/master/Karyotype). WGDI (https://github.com/SunPengChuan/wgdi) was used to conduct the collinearity relationship between the ancestral karyotype and species with the -icl and mg = 25,25 parameters (P. Sun *et al.*, 2022). Then, collinear fragments were obtained by -bi, filtered by -c. -km, used to obtain the distribution of relevant ancestral karyotypes in the species, and visualized them using the -ak parameter with default parameters (Murat *et al.*, 2017).

Extraction and analysis of essential oil

Flowers at five flower developmental stages, including S1 (Bud stage), S2 (display petal stage), S3 (initial flowering stage), S4 (full flowering stage) and S5 (end flowering stage), were harvested. The essential oil was extracted using a hydrodistillation approach as previously described (Chakira *et al.*, 2022) with improvement. Briefly, *c.* 200 g of frozen fresh flowers was ground in liquid nitrogen and then dissolved in 2000 ml of distilled water. The mixture was kept boiling for distillation for 4 h. The refrigeration tube was cooled to receive the hot hydrosol vapor and oil. Then, the harvested cooled oil was further extracted with 800 µl of hexane, filtered, and stored in an amber bottle at 4°C until GC-MS analysis.

Flower volatile compound analysis

The volatile compounds of flowers at the five developmental stages were analyzed by headspace solid-phase microextraction coupled to gas chromatography mass spectrometry (HS-SPME-GC-MS) as previously described (J. G. Li *et al.*, 2021; Xia *et al.*, 2021), with improvement. Briefly, each sample of fresh flowers (*c.* 3–5 flowers) was quickly put into a wet glass headspace bottle (50 ml) and sealed with paraffin film. The SPME fiber (50/30 μm DVB/CAR/PDMS) was placed upon the exposed flowers to extract the volatile components for 30 min at 25°C ± 3°C. Then, the fiber was inserted into the heated injector port of the GC instrument and desorbed at 250°C for 1 min for further GC-MS analysis.

GC-MS analysis

The volatile compounds were further analyzed using an Agilent 7890 gas chromatograph/5975 mass selective detector with a 30 m DB-5MS capillary column. Helium (He) was used as the carrier gas at a constant flow of 1.2 ml min^{-1}. GC-MS was performed with the following program: initial temperature of 80°C, 4°C min^{-1} to 180°C, and 12°C min^{-1} to 200°C for a 2 min bakeout. The inlet temperature was increased and kept at 250°C, and the MS transfer line was set at 290°C. The MS acquisition parameters included scanning from *m/z* 50 to 600 in electron impact (EI) mode for routine analysis. An empty sample and a n-hexane (solvent) were run as the control to exclude instrumental and solvent background interferences, and 'Subtract Background (BSB)' were selected to exclude background interferences. The NIST98 database was used to auto-match the mass spectra of the volatile compounds with retention index through Chem-Station (Agilent). All peak identities at each retention index were verified using EI mass spectral auto-matches to the authentic reference standards in NIST library combing match index. The Kovats retention index (KRI or RI) of each compound was calculated as previously described (Lu *et al.*, 2022), and then we compared the RI values via the known RI values (https://webbook.nist.gov/chemistry/name-ser/). For relative content analysis, the normalized peak area measurements were used to calculate the percentage of each flower volatile component. For absolute content analysis, n-pentadecane was served as an internal standard, and the peak areas of the compounds were corrected by that of the internal standard.

Transcriptome sequencing and analysis

For transcriptome sequencing, samples with each 2 g from three fresh leaves and 2 g from three flowers were used for total RNA isolation by the Eastep® Super Total RNA Extraction Kit (Promega). For each sample, three biological replicates were harvested at the same time. The qualified RNA was used to construct an RNA-Seq library according to the library construction protocol of the VAHTS Universal V6 RNA-Seq Library Prep Kit for Illumina® (NR604-01/02). A NovaSeq 6000 S4 platform by Benagen Co. (Wuhan, China) was used for sequencing. The RNA-Seq clean reads were mapped to the reference genome by HISAT2 (--very-sensitive-dta) (Kim *et al.*, 2019). The raw count of each transcript was tallied using featureCounts v.2.0.3 and normalized using Transcripts per kilobase of exon model per million mapped reads (TPM) values using an R script (Liao *et al.*, 2019). The heatmap was made with TBtools based on the gene transcript levels at different flower developmental stages (C. J. Chen *et al.*, 2020).

Candidate gene identification

The terpene and phenylpropanoid biosynthesis pathway genes in *C. odorata* and 22 other species were identified using protein sequence similarity by BLASTP with the known homologues in *Arabidopsis* from TAIR (https://www.arabidopsis.org/). To identify the putative *TPS* gene families, two Pfam domains (PF01397 and PF03036) were used to search against the proteome using HMMER in TBtools (C. J. Chen *et al.*, 2020), with an E-value cut-off of 10^{-5}.

The sequence alignment of *TPS* gene families from *C. odorata* and other representative species was performed using Mafft v.7 software online (https://mafft.cbrc.jp/alignment/software/) (Katoh *et al.*, 2002). The phylogenetic tree was constructed by Mega X software using the maximum-likelihood method and bootstrapping with 1000 replicates. Further annotation of the tree was performed using the online tool iTOL (https://itol.embl.de/).

Expression of *CoTPS21* homologues in *E. coli* and *in vitro* enzymatic assay

The CDS of *CoTPS21.2* (*Codo51071*) and *CoTPS21.3* (*Codo51075*) were cloned and inserted into the *pET28a* vector after the codon optimization and then transformed into *E. coli* Rosetta strain (DE3; Tsingke, Beijing, China). Single right colonies with OD_{600} of 0.6–0.8 were induced with 0.1 mM Isopropyl-β-D-Galactopyranoside (IPTG) at 28°C for 3 h. Then, the crude recombinant protein was purified using Ni-NTA 6FF Sefinose (TM) Resin Kit (C600332, BBI) following the product protocol. In the enzyme assay, total of 10 ng recombinant protein was incubated with 10 ng FPP as substrate in reaction buffer at 30°C for 2 h. The volatiles products of enzyme assay were analyzed using SPME-GC-MS. The primers used are listed in Supporting Information Table S1.

Expression of *CoTPS21* homologues in *Nicotiana benthamiana* for functional analysis

The coding sequence (CDS) of *CoTPS21.1*, *CoTPS21.2*, *CoTPS21.3*, *AtHMGR1* (*AT1G76490*) and *AtFPPS1* (*AT3G14530*) was cloned and inserted into the *PRI101-Flag* vector to generate overexpression vectors. These recombinant constructs were transferred into *Agrobacterium tumefaciens* strain LBA4404. The functional confirmation of *CoTPS21* genes by transient expression in *N. benthamiana* was conducted as previously described (Forestier *et al.*, 2021). The primers used are listed in Table S1.

Prediction of promoter *cis*-elements and transcription factors of *CoTPS21* homologues

The *cis*-elements of *CoTPS21* promoters were predicted using the PlantCARE database (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/). The upstream transcription factors of *CoTPS21* promoters were predicted using the PlantPAN2.0 database (http://plantpan2.itps.ncku.edu.tw/TFsearch.php).

Transcriptional activity assays

Transcriptional activity assays were conducted by luciferase luminescence as previously described (Zheng *et al.*, 2021). Three 2000-bp *CoTPS21* promoters were cloned and fused with the *PRI101-LUC* vector as reporters. The CDS of *TF* was cloned and inserted into the *PRI101-Flag* vector as effectors. Equal amounts of *A. tumefaciens* cells containing the reporter and effector were mixed and injected into tobacco leaves for fluorescence signal detection. The fluorescence intensity was analyzed using ImageJ. The primers used are listed in Table S1.

Yeast One-Hybrid assay

The CDS of *CoSPL3.1*, *CoSPL9.1* and *CoSPL9.2* was cloned and inserted into the *pGADT7-Rec2* vector, and the promoter of *CoTPS21.1*, *CoTPS21.2* and *CoTPS21.3* was cloned and inserted into the *pHIS2* vector. Then the AD and pHIS2 recombinant plasmids were co-transformed into the Y187 yeast strain for interaction validation following the product protocol Y187-pHis2 Yeast One-Hybrid interaction proving kit (YH1011-10 T; Coolaber, Beijing, China). The primers used are listed in Table S1.

Results

A near-complete T2T reference genome for *C. odorata*

Ylang–ylang planted in the Xishuangbanna Tropical Botanical Garden was selected for genome assembly. The genome size was estimated to be 776.04 Mb, with a heterozygosity rate of 0.49% based on *K-mer* analysis ($K = 21$) (Supporting Information Fig. S1; Table S2). Multiple sequencing platforms were used for the high-quality genome assembly of *C. odorata*. We first generated 45.45 Gb ONT ultralong reads (65× coverage) using the PromethION platform and 25.28 Gb (37 × coverage) HiFi reads using the PacBio Sequel II platform (Table S3). After assembly polishing and redundancy elimination, the final genome size was 735.83 Mb, with 19 contigs, an N50 length of 70.99 Mb and a GC content of 33.16% (Fig. 1a; Tables 1, S4). We further used the Hi-C-Pro program to produce chromosomal interaction maps with Hi-C reads mapped to the preliminary assembly, and 19 contigs anchored to 8 pseudochromosomes, which covered 95.66% of the Illumina short reads and 99.43% of the Hi-C clean reads (Figs 1b, S2; Table S5). Finally, four gapless chromosomes of *C. odorata* were assembled, and the other chromosomes had only five gaps (Table S6).

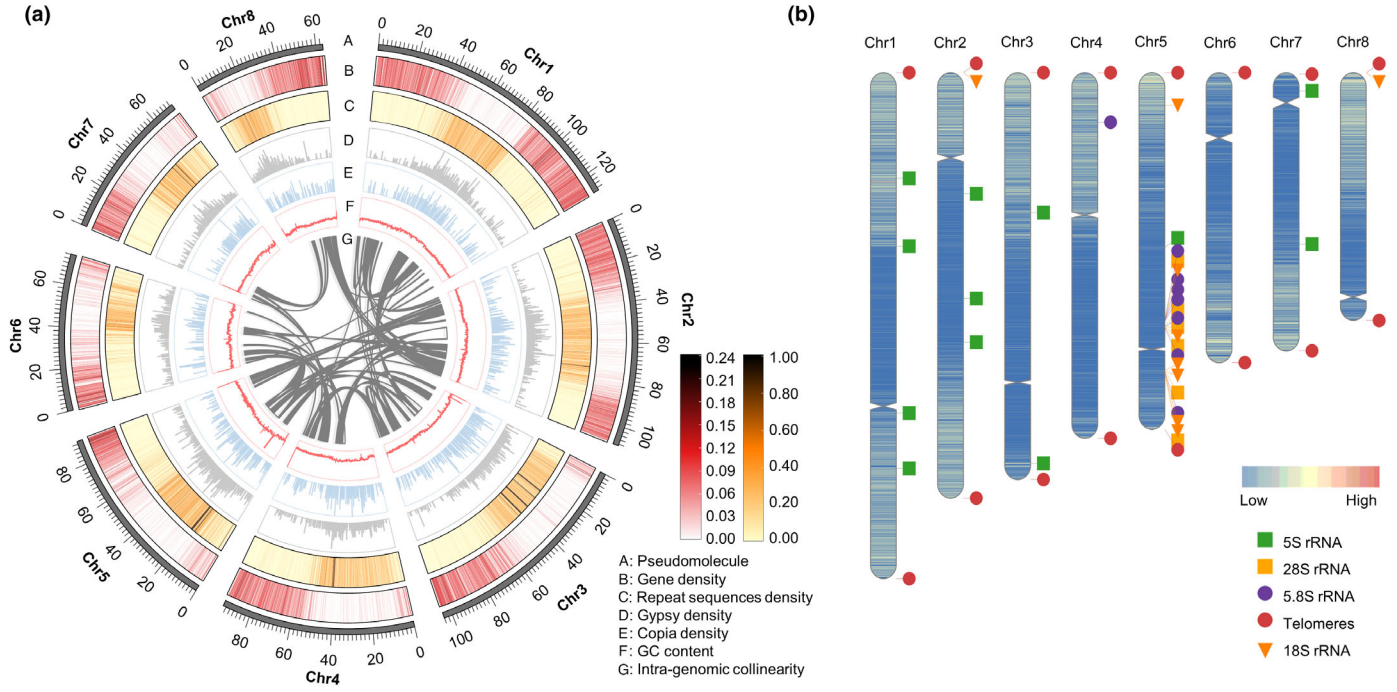

Fig. 1 Telomere-to-telomere (T2T) genome assembly of *Cananga odorata*. (a) Genome information, comprising chromosome pseudomolecule information (A), gene density (B), repeat sequence density (C), Gypsy density (D), Copia density (E), GC content (F), and intragenomic collinearity (G). (b) Telomere and centromere detection map. The red circles represent telomeres of the assembled chromosomes. The color gradient indicates the gene density. Chr, Chromosome; rRNA, ribosomal RNA.

Table 1 Genomic feature of *Cananga odorata*.

Total size of assembled contigs (Mb)	735.83
Number of contigs (gaps)	8 (5)
GC content (%)	33.16
Contig N50 (Mb)	93.52
Number of transcripts	119 566
Number of non-transposable element transcripts loci	744 132
Total size of TEs (Mb)	454.08
Busco (%)	98.30%

Based on the occurrence of the characteristic telomere repeats (CCCTAAA at the 5′ end or TTTAGGG at the 3′ end) (Fu *et al.*, 2023), we identified 16 telomeres at the terminus of 8 pseudochromosome (Table S7). The centromeric regions were predicted with the length ranging from 1.00 to 2.86 Mb on each 8 pseudochromosome (Table S8). We further assessed the completeness and accuracy of the assembly using various approaches. Analysis with Busco showed that a total of 1586 complete conserved genes were detected, with a high completeness rate of 98.3% (Fig. S3; Table S9). Furthermore, *c.* 97.34% RNA-Seq data mapped to the assembled genome (Table S10). These results suggested that the genome assembly of *C. odorata* was nearly complete, and this is the first T2T genome in magnoliids.

Genome annotation

Based on homology prediction, *de novo* prediction and transcriptome prediction approaches, a total of 26 449 protein-coding genes were identified in *C. odorata* genome with an average gene length of 4756.32 bp and an average of 4.52 exons per gene (Tables S11, S12). The number of genes was smaller than that in *Magnolia biondii* (Magnoliaceae), larger than that in *Aristolochia fimbriata* (Aristolochiaceae), and similar to that in *Cinnamomum kanehirae* (Lauraceae) (Table S12). Functional annotation was performed using COG, Pfam, GO, KEGG, Swiss-Prot, and NR databases, and 78.88% (20 863) of the predicted protein-coding genes had been annotated (Table S13). We also identified the noncoding RNAs in the *C. odorata* genome, including 356 tRNAs, 133 microRNAs and 705 rRNAs (Table S11). Meanwhile, the repeated sequences, accounting for 61.71% of the genome, were further identified, including 45.72% of long terminal repeats (LTRs), which consisted of 29.10% Gypsy, 12.58% Copia, and 4.04% other unknown sequences (Table S14). We further performed a genome-wide investigation of BGCs in *C. odorata* using the plantiSMASH genome mining algorithm. A total of 18 BGCs were identified in *C. odorata* genome, including terpene, putative, saccharide, alkaloid, lignan, saccharide-polyketide and saccharide-terpene biosynthesis pathway (Table S15).

The phylogenetic position of magnoliids within angiosperms

Numerous phylogenetic and genome sequencing studies have proposed discordant relationships for the magnoliid, monocot and eudicot groups of angiosperms (Y. C. Chen *et al.*, 2020; H. T. Li *et al.*, 2021; Qin *et al.*, 2021). To explore these phylogenetic discrepancies, we first analyzed the chromosomal rearrangements shared among some major groups including *C. odorata* and other angiosperms (2 ANA-clade, 8 Magnoliids, 5 monocots and 6 eudicots). A collinearity point map among these genomes was constructed and was related to chromosomes 1, 2, 4 and 5 of *C. odorata*, containing the A1-A2 (chr1), B1-B2 (chr2), C1-C2 (chr4), D1-D2 (chr5) collinearity modules (Fig. S4). We found that A1 and A2 modules was fused in eudicots, but there is no consistency in magnoliids and monocots, similar results appeared in C1 and C2 module (Fig. S5). It seemed that the phylogenetic relationships between magnoliids, eudicots, and monocots were still unresolved by genome structural comparisons analysis.

We further performed phylogenetic analyses based on different sampling datasets and used OrthoMCL to verify the evolutionary relationships of magnoliids with 23 plants, including 5 eudicots, 5 monocots, 6 magnoliids, 3 ANA-clade members, 1 Ceratophyllale, 1 Chloranthale, *Selaginella moellendorffii* (Selaginellaceae) and an outgroup gymnosperm (*Ginkgo biloba* L.) (Table S16). A total of 45 SSCG families and 267 MSCG families were identified among the 23 representative species (Table S17). We first constructed a multi-species maximum-likelihood (ML) phylogenomic tree by the concatenation method with nucleotides, amino acids and three codon position sequences. The concatenation-based phylogenetic trees of MSCGs supported the T2 hypothesis (magnoliids and eudicots are sister clades after their divergence from monocots) (Figs 2a–c, S6). The concatenation-based phylogenetic trees of SSCGs also supported the T2 hypothesis (Figs S7, S8).

Furthermore, we constructed gene trees using a coalescent-based method, and the trees were then input into ASTRAL for species tree inference. The phylogenetic trees based on MSCGs showed that *C. odorata* closely clustered with the other two magnoliids, and the phylogenetic trees strongly supported the T2 hypothesis that magnoliids are a sister clade to eudicots (q2: 50.98% of nucleotides, 56.79% of proteins, 47.06% of codon3) (Figs 2d, S6). In addition, the phylogenetic trees based on SSCGs also significantly supported T2 as the main topology (Figs S7, S8). Together, our estimated gene trees of both concatenation and coalescent analyses produced an identical strongly supported topology (Tables S17, S18), suggesting that the magnoliids are likely to be sister to eudicots rather than sister to monocots or sister to the clade of eudicots and monocots.

We also constructed a phylogenetic tree using 1393 single-copy gene families and calibrated it by fossil evidence. The results showed that the divergence time of Annonaceae and Magnoliaceae species was *c.* 72.11 (50.71–94.02) million years ago (Ma) and *C. odorata* and *Annona muricata* diverged *c.* 27.54 (12.11–45.02) Ma (Fig. S9).

Comparative genomic analysis, WGD and karyotype analysis

To analyze gene family evolution, we compared the *C. odorata* genome with 8 other plant genomes using OrthoFinder (Emms & Kelly, 2015) and identified 751 expanded and 935

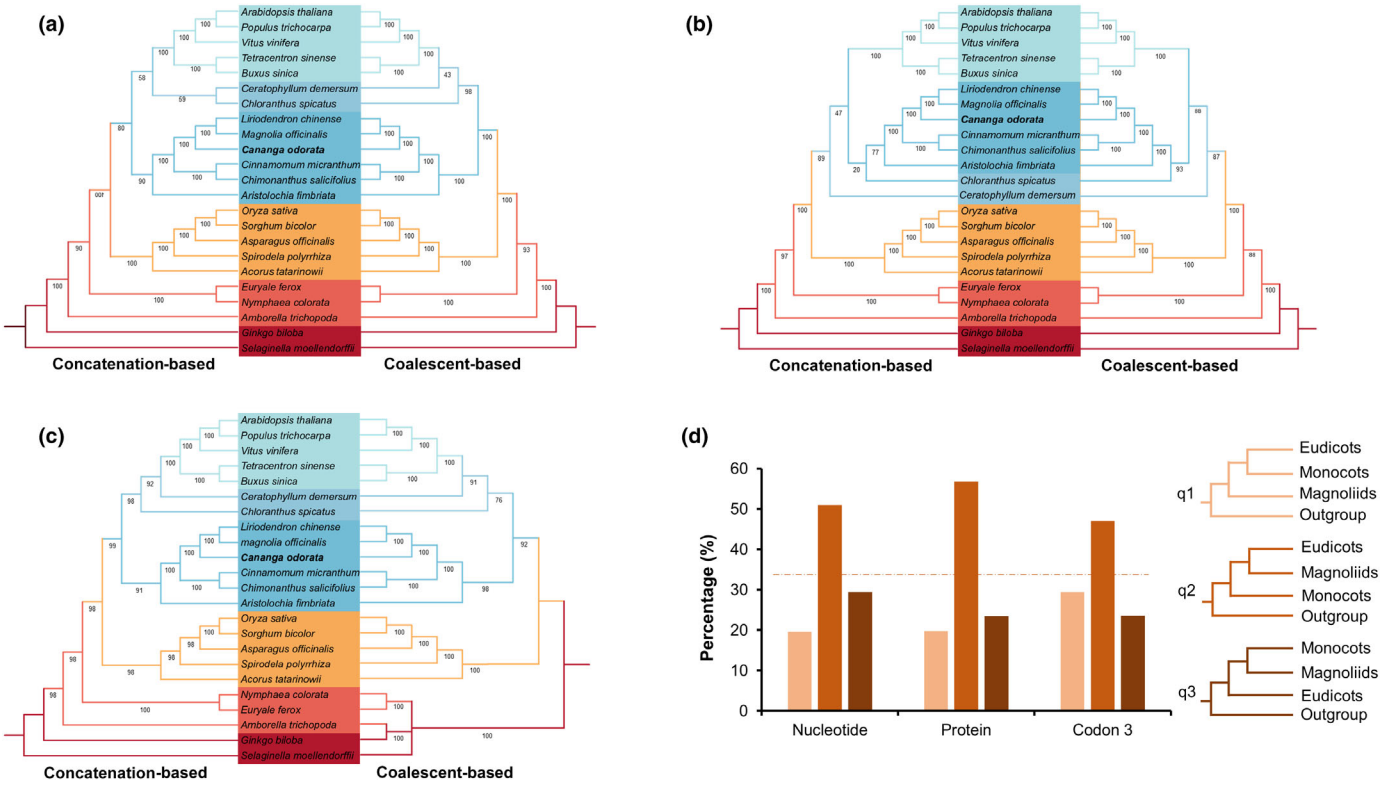

Fig. 2 Phylogenetic relationships between *Cananga odorata* and other angiosperms. (a–c) Phylogenetic tree based on nucleotide sequences (a), amino acid sequences (b) and codon 3 sequences (c) of 267 mostly single-copy gene (MSCG) families obtained using concatenation (left) and coalescent (right) methods. Different colors represent different plant groups, from upper panel to lower panel are eudicots clade (*Arabidopsis thaliana*, *Populus trichocarpa*, *Vitis vinifera*, *Tetracentron sinense* and *Buxus sinica*), Chloranthales and Ceratophyllales clade (*Ceratophyllum demersum*, *Chloranthus spicatus*), magnoliids clade (*Liriodendron chinense*, *Magnolia officinalis*, *Cananga odorata*, *Cinnamomum micranthum*, *Chimonanthus salicifolius* and *Aristolochia fimbriata*), monocots clade (*Oryza sativa*, *Sorghum bicolor*, *Asparagus officinalis*, *Spirodela polyrhiza* and *Acorus tatarinowii*), amborellales nymphaeales austrobaileyales clade (*Euryale ferox*, *Nymphaea colorata* and *Amborella trichopoda*) and outgroup clade (*Ginkgo biloba*). The number appeared at branches indicates the supporting rate of the phylogenetic tree. (d) Proportions of the 267 MSCG gene trees with different topologies. The dashed line refers to the mean proportion of 33%. q1, magnoliids as sister to both eudicots and monocots (T1 hypothesis); q2, magnoliids as a sister group to eudicots (T2 hypothesis); q3, magnoliids as sister to monocots (T3 hypothesis).

contracted gene families in the *C. odorata* genome (Fig. 3a; Table S19). Expanded gene families might facilitate better adaptation to the environment in plants due to their new functions acquired during the evolutionary process (Moore & Puruggannan, 2005; Cui *et al.*, 2022; Li *et al.*, 2022). Further GO enrichment analysis revealed that the expanded gene families were mainly enriched in the pathways of catalytic activity, secondary metabolite biosynthetic process, farnesyl diphosphate metabolic process, and sesquiterpene synthase activity (Fig. 3b; Table S20). Notably, the significant enrichment of sesquiterpene biosynthesis-related pathways suggested that the expansion of these gene families possibly played roles in the synthesis of fragrant substances in *C. odorata*.

Further comparative analysis of the distribution of pairwise synonymous substitutions (Ks) values of paralogs of a range of species with *C. odorata* showed that *C. odorata* was absent of a lineage-specific WGD but did experience the ancestral WGD event (Ks, *c.* 1.78) commonly shared among magnoliids (Fig. 3c). In addition, we performed the Fourfold synonymous third-codon transversion position (4dTV) analysis, and the results was consistent with those of Ks analysis (Fig. S10). Collinearity and synteny analyses revealed that the syntenic depth ratio between *C. odorata* and *Aristolochia fimbriata* and between *C. odorata* and *Cinnamomum micranthum* were 2 : 1 and 1 : 2, respectively (Figs 3c, S11; Table S21). These results confirmed that only single ancient WGD event occurred in *C. odorata*. We also inferred the chromosome evolution of *C. odorata*, an evaluation of the ancestral eudicot karyotype (AEK) and ancestral angiosperm karyotype (AAK) was used to examine how the ancestral chromosomes have undergone rearrangements to generate the current *C. odorata* genome. The results showed that all 8 chromosomes of *C. odorata* underwent rearrangement events, and 31.85% and 24.68% of all the genes in *C. odorata* corresponded to AEK genes and AAK genes, respectively (Figs 3d, S12; Table S22).

Floral metabolic profiling during flower development in *C. odorata*

To identify the main compounds of floral VOCs in the essential oil and floral scent of *C. odorata*, we performed GC-MS analysis on the floral VOCs at five flower developmental stages, from

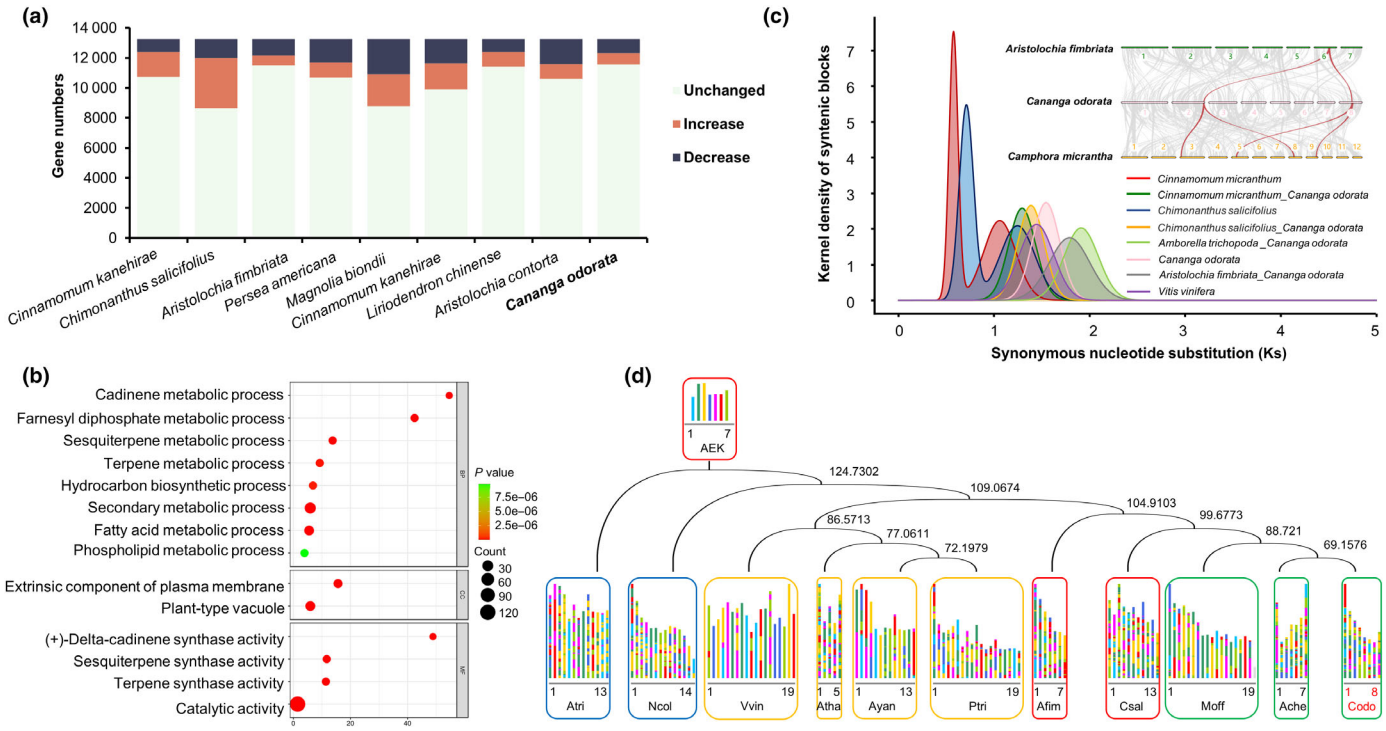

Fig. 3 Comparative genomic analysis and genome evolution of *Cananga odorata*. (a) The expanded and contracted gene families in *C. odorata* and 8 other species. Included are the numbers of expanded gene families (red color), contracted gene families (blue color) and unchanged gene families (green color). (b) Gene ontology enrichment analysis of the most significantly expanded gene families in *C. odorata*. (c) Ks values for orthologous and paralogous genes between *C. odorata* and other indicated species and Chromosome syntenic comparison between *C. odorata* and *Cinnamomum micranthum* and between *C. odorata* and *Aristolochia fimbriata*. The gray lines indicate the syntenic regions spanning the genomes, and the red lines indicate the examples. (d) Karyotype analysis of *C. odorata* ($2n = 16$) based on the Ancestral Eudicot Karyotype (AEK) of modern eudicots. The number in the tree indicates the divergence time with unit of Million years ago (Ma). Atri, *Amborella trichopoda*; Ncol, *Nymphaea colorata*; Vvin, *Vitis vinifera*; Atha, *Arabidopsis thaliana*; Ayan, *Acer yangbiense*; Ptri, *Populus trichocarpa*; Afim, *Aristolochia fimbriata*; Csal, *Chimonanthus salicifolius*; Moff, *Magnolia officinalis*; Ache, *Annona cherimola*; Codo, *Cananga odorata*.

green flower buds to fully mature yellow flowers (Fig. 4a). A total of 44 and 75 volatile compounds in the essential oil and floral scent were identified, respectively.

In the essential oil, 9, 25, 30, 23 and 20 volatile compounds in essential oil were detected in stages 1–5, respectively (Figs 4b,c, S13). The content of dominant VOCs at different developmental stages were different. By analyzing the relative and absolute content of the VOCs, we found that stages 1 and 2 were characterized mainly by sesquiterpenes including β-caryophyllene, α-caryophyllene and β-cubebene. Stage 3 was dominant in sesquiterpenes with little aromatic ester, which belong to the benzenoid/phenylpropanoid group, such as geranyl acetate and benzyl benzoate. The last two mature stages, stages 4 and 5, during which the flower gradually turns yellow and emits a pleasing aroma, were dominated by sesquiterpenes and more aromatic ester (Tables S23, S24). Principal component analysis (PCA) performed using the absolute content of volatile compounds in essential oil at stages 1–5 also showed that the volatile compounds with high positive scores on both PC1 and PC2 included β-caryophyllene, α-caryophyllene and β-cubebene (Fig. 4d; Table S25). Notably, β-caryophyllene was the most abundant component in essential oil at the mature stages of flowering (Fig. 4b).

In the VOCs of flower scent, sesquiterpenes and benzenoids/phenylpropanoids (aromatic esters and ether-oxides) were also the main compounds observed. The highest contents were sesquiterpenes (64.19 and 58.14%) and esters (23.45 and 32.33%) at the last two mature stages (Fig. S14; Table S26). β-caryophyllene with significantly high yield was detected at the last two maturity stages (28.76 and 25.13%), which was consistent with that in the essential oil. Our results showed that sesquiterpenes and benzenoids/phenylpropanoids are the two most abundant groups of chemicals in *C. odorata* flowers, and the last two developmental stages of flowers that are most suitable for *C. odorata* flower harvest were characterized by most abundant sesquiterpenes and aromatic esters.

Terpenoid and benzenoid/phenylpropanoid biosynthesis

To understand the genetic basis of terpenoid biosynthesis in *C. odorata* flowers, a total of 34 genes related to the terpene skeleton synthesis of two canonical terpene biosynthesis pathways (methylerythritol phosphate pathway and mevalonate pathway) were identified (Fig. 5a; Table S27). Based on our BGCs analysis, these genes didn't form BGCs (Table S15). Further comparative genomics analysis among 23 species showed that most of the

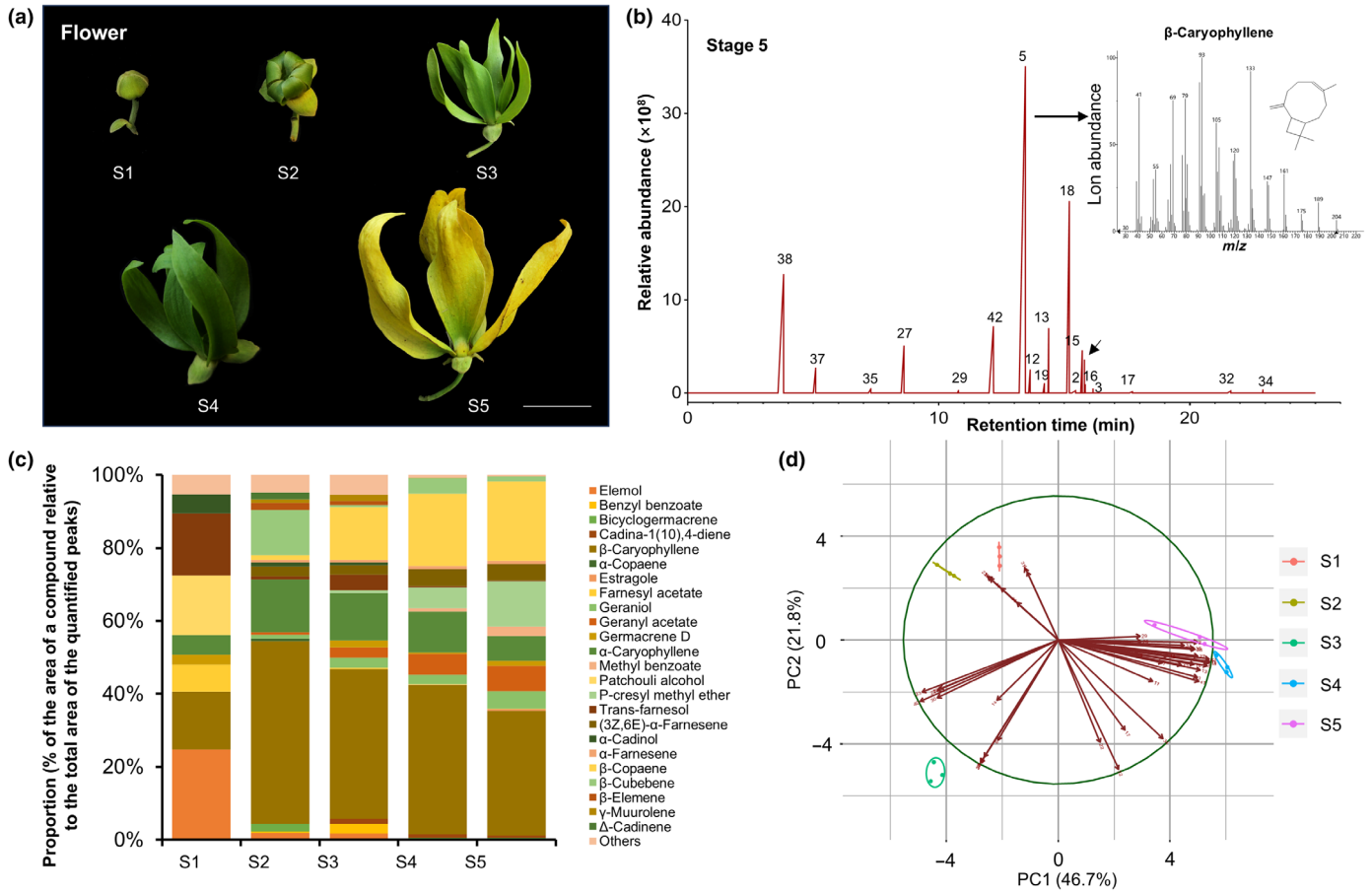

Fig. 4 Metabolic analysis of volatile organic compounds (VOCs) in the essential oil of *Cananga odorata* flowers. (a) The morphological characteristics of *C. odorata* flowers at five different developmental stages. (S1) Bud stage; (S2) display petal stage; (S3) initial flowering stage; (S4) full flowering stage; (S5) end flowering stage. Bar, 1.5 cm. (b) GC traces of floral VOCs in essential oils extracted from flowers at S5. The arrow indicates internal standard (n-pentadecane) used in the assay. The number beside the peak indicates the compounds which were identical to that in Supporting Information Table S23. EI mass spectrum of β-caryophyllene is shown upper right. (c) Relative amounts of VOCs in the essential oil of *C. odorata* flowers. The measurements were independently conducted at 3 times with similar results. (d) PCA of the VOCs in the essential oil of *C. odorata* flowers at stages 1–5. The PCA biplot shows the loadings of variables and the Principal Component (PC) scores of samples. The compounds of number are listed in Table S25.

genes involved in the first stage were present as one copy (Table S28), while the *TPS* genes, which are the synthases responsible for the last catalytic reaction to generate terpenoid compounds (Chen *et al.*, 2011; Jiang *et al.*, 2019), were found to expand in the *C. odorata* genome. We identified 56 putative *TPS* genes in *C. odorata*, and phylogenetic analysis further divided them into 6 subfamilies of the *TPS-a*, *TPS-b*, *TPS-c*, *TPS-e/f* and *TPS-g* clades (Figs 5b, S15; Table S29). Comparative genomics analysis revealed that the *TPS-a* and *TPS-b* genes significantly expanded in *C. odorata*, compared to that in the other 5 magnoliid plants (Table S30). We found that proximal duplication events (involving 20 *TPS* homologues) and tandem duplication events (involving 22 *TPS* homologues) majorly impacted *TPS* gene expansion, which possibly contributed to the abundance and diversity of volatile terpenoids in *C. odorata* (Fig. 5c; Table S31). We also identified 75 genes encoding the key enzymes in the two main pathways (Shikimate pathway and phenylalanine pathway) of benzenoid/phenylpropanoid biosynthesis in the *C. odorata* genome (Fig. S16; Tables S32, S33), and the tandem duplication events and proximal duplication events (involving 39 genes) significantly affected both the upstream and downstream genes involved in the biosynthesis of some specific benzenoids (Table S34).

Further transcriptome analysis showed that the expression pattern of some genes involved in benzenoid/phenylpropanoid biosynthesis was consistent the emission pattern of metabolism (Table S35). Furthermore, in the mevalonate pathway, which is responsible for sesquiterpenoid biosynthesis, the expression levels of the key genes in the first step of sesquiterpene biosynthesis were abundant in the early stages of flower development. The genes involved in the second step were mainly highly expressed in stages 3 and 4 (Fig. 5a). As for *TPS* homologues, notably, three *TPS-a* genes *TPS21.1* (*Codo51068*), *TPS21.2* (*Codo51071*) and *TPS21.3* (*Codo51075*), which originated from tandem duplication events and were the homologues of *TPS21* in *Arabidopsis*, were specifically activated in flowers at stage 2 and remained at significantly high levels into the mature stages (Fig. 5c). These results were consistent with the emission trends of the main

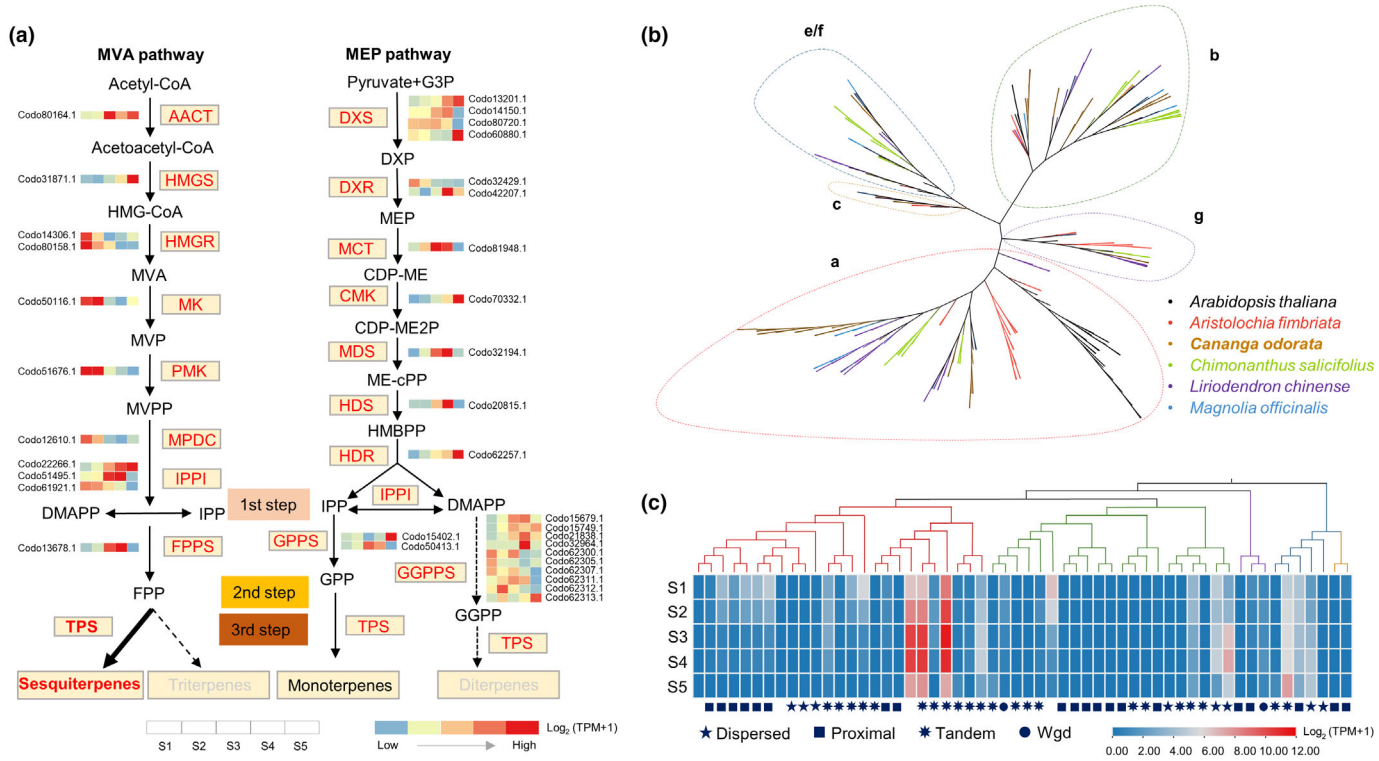

Fig. 5 Terpene biosynthesis pathway in *Cananga odorata*. (a) Diagram of the terpene biosynthetic pathways and expression profiles of genes encoding enzymes involved in the mevalonate and methylerythritol phosphate pathways in *C. odorata*. The key synthases and intermediates are shown in red and black. The five different colored boxes next to each gene indicated the gene expression levels at the five developmental stages. (b) Phylogenetic tree of TPSs in *C. odorata* and other representative species. TPSs from different species are indicated with different colors. (c) Gene expression levels of *TPS* genes from flowers at different stages in *C. odorata* by RNA-Seq. The upper panel is the phylogenetic tree of TPS protein; the red color indicates *TPS-a* subfamilies, the green color indicates *TPS-b* subfamilies, the orange color indicates *TPS-c* subfamilies, the blue color indicates *TPS-e/f* subfamilies and the purple color indicates *TPS-g* subfamilies. Different shapes in the lower panel indicate different duplication event origins of the *TPS* genes. Expression values were scaled by $\log_2$ (TPM + 1). S, Stage of *C. odorata* flower. TPM, Transcripts per kilobase per million mapped reads in RNA-Seq.

sesquiterpenes in the floral VOCs of *C. odorata*, indicating that the high levels of expression of these three *TPS-a* genes might play important roles in sesquiterpene biosynthesis.

The genetic basis of sesquiterpenoid biosynthesis

To better determine the roles of *TPS* genes in sesquiterpene biosynthesis in *C. odorata*, we cloned three *CoTPS-a* genes (*TPS21.1*, *TPS21.2* and *TPS21.3*) with specific high expression to determine their functions. These *TPS21* homologues were inserted into the *pET28a* vector and then transformed into the *E. coli* Rosetta strain for prokaryotic expression and further *in vitro* enzyme assays. The results of SPME-GC-MS showed that purified *CoTPS21.2* and *CoTPS21.3* recombinant protein specifically catalyzed the biosynthesis of β-caryophyllene, which is the main sesquiterpene of the floral VOCs in *C. odorata*, using FPP as a substrate (Figs 6a–d, S17). Furthermore, the results of transient expression analysis in *N. benthamiana* showed that when co-expressed *CoTPS21.2* or *CoTPS21.3* with *AtHMGR1* and *AtFPPS1* genes, which were previously identified to increase heterologous sesquiterpene production in plants (Green *et al.*, 2012; Song *et al.*, 2012; Jin *et al.*, 2015), one specific peak was clearly detected and identified as β-caryophyllene by GC-MS analysis (Fig. S18). These results suggested that CoTPS21.2 and CoTPS21.3 were the key synthases that catalyze the biosynthesis of β-caryophyllene in *C. odorata*.

In addition, our results showed that the *Codo13515* and *Codo21750* genes, which belong to the *TPS-e/f* subfamily and were significantly activated at stage 4–5, catalyzed the biosynthesis of α-Farnesene when co-expressing them with *AtHMGR1* and *AtFPPS1* in *N. benthamiana* (Fig. S19). α-Farnesene is also a kind of sesquiterpene widely used in perfume industry and is rich in the last two stage of floral development in *C. odorata*. Together, these results suggested that after the expansion events of the *TPS* gene family in *C. odorata*, the specially-activated *TPS* genes during mid- to late-flower development play important roles in the sesquiterpenes biosynthesis, which are the major compounds of floral VOCs in *C. odorata*.

The transcriptional control of genes encoding synthases for terpene biosynthesis is related to the *cis*-elements distributed in their promoter regions (M. Sun *et al.*, 2022). To investigate the potential TF to activate the transcription of the *TPS21* genes, we further screened the *cis*-elements in the 2000-bp regions upstream of the coding region of the *TPS21* genes (Fig. S20). We found that SBP and MYB-binding elements were the potential upstream TFs of these three *TPS21* genes. To further verify

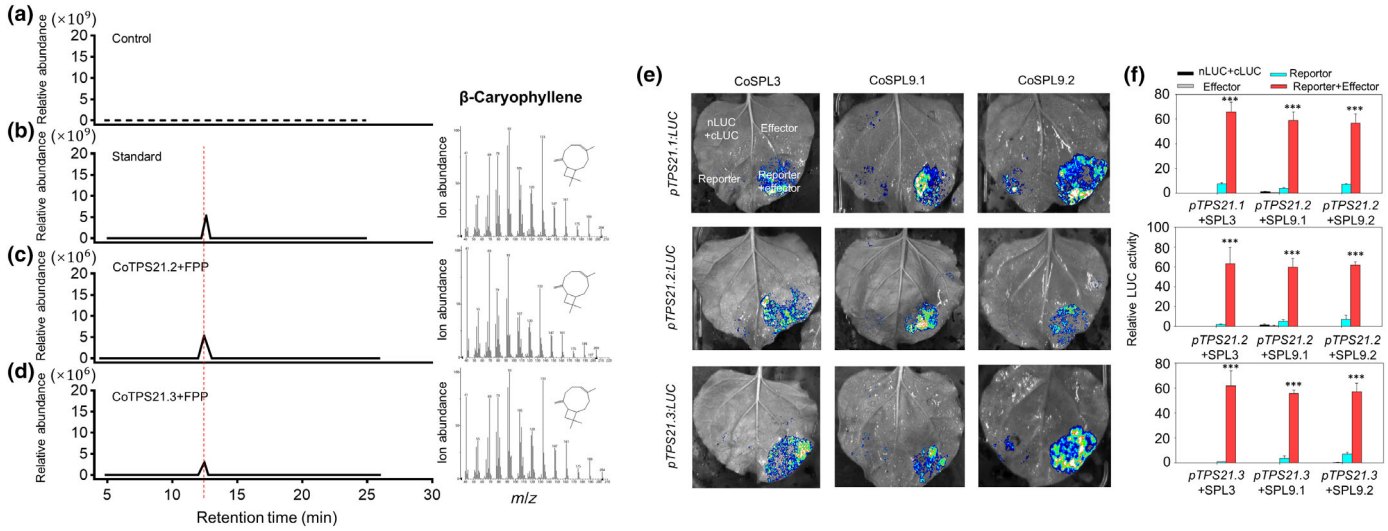

Fig. 6 Functional characterization of the *CoTPS21* homologues of *Cananga odorata* in sesquiterpene biosynthesis. (a) Total ion chromatogram (TIC) diagram of reaction products of control with FPP. (b) TIC diagram of β-caryophyllene standard. (c, d) TIC diagram of volatile terpenes products of overexpressing *CoTPS21.2* (c) and *CoTPS21.3* (d) genes with FPP as a substrate obtained from *in vitro* enzyme assay. Recombinant proteins were purified using Ni-NTA 6FF Sefinose™ Resin Kit (C600332, BBI). The volatile terpenes products were analyzed using solid-phase microextraction (SPME)-GC-MS. (e) Transcription activity analysis showing the activation activity of three effectors (CoSPL3, CoSPL9.1 and CoSPL9.2) at the reporters containing the *CoTPS21* promoters. (f) Relative LUC activity of effector with corresponding reporter construct. Data are mean ± SD, $n = 5$. ***, $P < 0.01$ compared with the control (Student's *t*-test). *m/z*, mass charge ratio. The red dashed line refers to the retention time of β-caryophyllene production.

whether these TFs can activate the *TPS21* promoters, we performed a transcriptional activity assay by co-infiltrating the reporter (*pTPS21:LUC* series) and the effector (*35S:TF* series) into tobacco leaves. Our results showed that a strong LUC signal was detected when co-transformation of *pTPS-21:LUC* with *CoSPL3* (*Codo33137*), *CoSPL9.1* (Codo13415) and *CoSPL9.2* (*Codo81486*), which belong to the SBP elements (Fig. 6e,f). Further results of yeast one-hybrid assays showed that CoSPL3, CoSPL9.1 and CoSPL9.1 could bind to the *TPS21* promoters (Fig. S21). However, we did not detect the activation signal when co-transformation of some TFs belonging to the MYB-binding elements, such as *CoMYC2.1* (*Codo12352*) and *MYC2.2* (*Codo31189*) (Fig. S22). Taken together, these results indicated that CoSPL3 and CoSPL9 homologues may be the key upstream TFs activating *CoTPS21* genes *in C. odorata*.

Discussion

Cananga odorata is not only one of the most popular raw materials for fragrances world-wide but also considered a potentially useful plant in the agriculture and medicine industries (Burdock & Carabin, 2008). *C. odorata* belongs to the Annonaceae family Magnoliids, and the flower of *C. odorata* is the main organ that produces aromatic essential oil (Nurhayani *et al.*, 2019). High-quality and representative genomic resources will facilitate floral fragrance biosynthesis and evolutionary studies of *C. odorata* as well as other members of Annonaceae. Here, we generated a high-quality T2T genome of *C. odorata* by combining a variety of sequencing platforms. This nearly complete genome assembly of *C. odorata* will undoubtedly be regarded as a benchmark for genetic research on *C. odorata* and Annonaceae plants. The *C. odorata* genome will contribute to figuring out the phylogenetic placement of magnoliids through a phylogenomic approach and shed light on the genome evolution relationships of angiosperms.

The phylogenetic position of magnoliids is still unclear, although numerous attempts at phylogenomic analysis according to *c.* 20 genomes of magnoliid plants have been performed (Dong *et al.*, 2021; Qin *et al.*, 2021; Shen *et al.*, 2022). This phylogenetic discordance in the position of magnoliids is probably due to ILS, ancient hybridization, the use of different and sparse taxon samples, and parallel evolution that occurred during plant evolution (Leebens-Mack *et al.*, 2005; Rendón-Anaya *et al.*, 2019; Yang *et al.*, 2020; Qin *et al.*, 2021). In this study, we used two complementary tools to extract the nucleotide sequences, amino acids and partitioned codons of SSCGs and MSCGs to reconstruct the phylogeny with concatenation- and coalescent-based analysis. In addition, we also improved taxon sampling by selecting key lineages in the magnoliids, monocots and eudicots, ANA-clade members, and additional lineages to cover key representative clades (Table S16). Finally, all the phylogenetic analyses performed most parsimoniously implied that *C. odorata* was more closely related to eudicots than to monocots, suggesting that magnoliids is sister to eudicots after the common ancestor of magnoliids and eudicots diverged from monocots.

Cananga odorata is widely used in the perfume and essential oil industries due to its unique floral scent. The qualities and quantities of VOCs in essential oils and floral scents depend on genetic background, geographical climate environment and cultivation methods during adaptation and domestication of *C. odorata* (Benini *et al.*, 2012; Qin *et al.*, 2014). The VOCs of essential oils of *C. odorata* populations from different region are quite different (Benini *et al.*, 2012). Some populations have higher ester

compounds, while some populations have more than half of the terpene compounds (Jin *et al.*, 2015). Our results showed that the complex, rich aroma of *C. odorata* flowers is a mixture of chemical compounds, mainly sesquiterpenes and aromatic esters, especially β-caryophyllene, indicating that these main compounds potentially contributed to the distinct floral scent in *C. odorata*.

Previous studies have demonstrated that the functional divergence of *TPS* genes caused by expansion or duplication events could contribute to the differences in terpene concentration and production among different species (H. Chen *et al.*, 2020; Zhang *et al.*, 2020; Chen *et al.*, 2023). In this study, we found remarkable duplication and expansion of *TPS-a* genes, which are involved in the production of major sesquiterpene components. Some TPSs can catalyze several products, while most of TPSs can synthesize one or two compounds dominantly as the major products (Shimada *et al.*, 2004; Fähnrich *et al.*, 2011; D. S. Li *et al.*, 2021). In this study, functional experiments showed that *CoTPS21.2* and *CoTPS21.3*, which likely encode a kind of sesqui-TPSs, can specially generate only one major product, β-caryophyllene, which is the main compound of sesquiterpenes in *C. odorata* flowers. Some *TPS* genes show high overlapping functions in regulating floral fragrance (Wang *et al.*, 2023). Our results indicated that these two *CoTPS21* genes, which are homologous genes and have similar functions, may have co-evolved and both have been retained to contribute to the floral metabolism of *C. odorata*. Previous studies have shown that the transcription factors ERF, NAC, MYB, and bHLH are participated in regulating the expression of genes involved in the floral volatiles synthesis in *Petunia hybrida* and *Cymbidium goeringii* (Spitzer-Rimon *et al.*, 2012; Liu *et al.*, 2017; Ramya *et al.*, 2019). In this study, we identified CoSPL3 and CoSPL9 belonging to SBP families were the potential upstream TFs for the *CoTPS21* genes expression.

Taken together, in this study, we generated a high-quality chromosome-level assembly for *C. odorata*. This is the first T2T genome in magnoliids and provides a valuable genomic basis for obtaining insights into the phylogenetic position of magnoliids. In this study, we identified the components of VOCs in *C. odorata* and revealed the molecular relationship between the fragrance of *C. odorata* and *TPS* genes, which provide important genetic information for enhancing the yield potential of sesquiterpenoids in *C. odorata*. The high-quality reference genome of *C. odorata* would contribute to identify potential genes involved in key agronomic trait, such as higher and better essential oil production, which might facilitate for the potential commercial values of *C. odorata*.

Acknowledgements

The work was supported by the (nos. 32200306, 32170385, 32070362 and 32100315), the Second Tibetan Plateau Scientific Expedition and Research (STEP) program (no. 2019QZKK0502), the Postdoctoral Directional Training Foundation of Yunnan Province (to Yan Zheng) and the Strategic Priority Research Program of the Chinese Academy of Sciences, Pan-Third Pole Environment Study for a Green Silk Road (Pan-TPE) (Grant no. XDA2004010306). We are grateful to Dr Fei Li (Service center for experimental biotechnology, Kunming Institute of Botany, Chinese Academy of Sciences) for assistances in GC-MS analysis.

Competing interests

None declared.

Author contributions

Yongping Yang, XK and Yunqiang Yang conceived and designed the experiments. YZ and DY performed the experiments and analyzed the sequencing data. YZ and DY wrote the manuscript. X Yin, X Yang, MC, XL, TY, SJ and HD participated in the sample collecting, data analysis, and interpretation of data. All authors read and approved the final manuscript. YZ and DY contributed equally to this work.

ORCID

Xiangxiang Kong https://orcid.org/0000-0001-7096-5404
Danni Yang https://orcid.org/0000-0003-0786-0524
Yongping Yang https://orcid.org/0000-0002-0327-2664
Yunqiang Yang https://orcid.org/0000-0002-8109-8315
Yan Zheng https://orcid.org/0000-0002-2810-7382

Data availability

All the raw sequencing data for the Illumina short reads, PacBio HiFi reads and RNA sequencing reads have been deposited in the Science Data Bank (doi: 10.57760/sciencedb.13236). The genome assembly and genome annotation files are available at the Science Data Bank (doi: 10.57760/sciencedb.13195).

References

Benini C, Mahy G, Bizoux JP, Wathelet JP, du Jardin P, Brostaux Y, Fauconnier ML. 2012. Comparative chemical and molecular variability of *Cananga odorata* (Lam.) Hook.f. & Thomson forma genuina (ylang–ylang) in the Western Indian Ocean Islands: implication for valorization. *Chemistry Biodiversity* **9**: 1389–1402.

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics* **3**: lqaa108.

Brůna T, Lomsadze A, Borodovsky M. 2020. GENEMARK-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics and Bioinformatics* **2**: lqaa026.

Burdock GA, Carabin IG. 2008. Safety assessment of ylang–ylang (*Cananga* spp.) as a food ingredient. *Food and Chemical Toxicology* **46**: 433–445.

Chakira A, Garcia C, Soria C, Minier J, Chillet M. 2022. Effect of flower development stages on the dynamics of volatile compounds in ylang–ylang (*Cananga odorata*) essential oil. *Horticulturae* **8**: 986.

Chaw SM, Liu YC, Wu YW, Wang HY, Lin CI, Wu CS, Ke HM, Chang LY, Hsu CY, Yang HT *et al.* 2019. *Stout camphor* tree genome fills gaps in understanding of flowering plant genome evolution. *Nature Plants* **5**: 63–73.

Chen CJ, 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 F, Tholl D, Bohlmann J, Pichersky E. 2011. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *The Plant Journal* **66**: 212–229.

Chen H, Guo M, Dong S, Wu X, Zhang G, He L, Jiao Y, Chen S, Li L, Luo H. 2023. A chromosome-scale genome assembly of *Artemisia argyi* reveals unbiased subgenome evolution and key contributions of gene duplication to volatile terpenoid diversity. *Plant Communications* **4**: 100516.

Chen H, Kollner TG, Li G, Wei G, Chen X, Zeng D, Qian Q, Chen F. 2020. Combinatorial evolution of a terpene synthase gene cluster explains terpene variations in *Oryza*. *Plant Physiology* **182**: 480–492.

Chen J, Hao Z, Guang X, Zhao C, Wang P, Xue L, Zhu Q, Yang L, Sheng Y, Zhou Y *et al.* 2019. Liriodendron genome sheds light on angiosperm phylogeny and species-pair differentiation. *Nature Plants* **5**: 18–25.

Chen S, Zhou Y, Chen Y, Gu J. 2018. FASTP: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890.

Chen YC, Li Z, Zhao YX, Gao M, Wang JY, Liu KW, Wang X, Wu LW, Jiao YL, Xu ZL *et al.* 2020. The *Litsea* genome and the evolution of the laurel family. *Nature Communications* **11**: 1675.

Cui F, Ye X, Li X, Yang Y, Hu Z, Overmyer K, Brosché M, Yu H, Salojärvi J. 2022. Chromosome-level genome assembly of the diploid blueberry *Vaccinium darrowii* provides insights into its subtropical adaptation and cuticle synthesis. *Plant Communications* **3**: 100307.

Dong S, Liu M, Liu Y, Chen F, Yang T, Chen L, Zhang X, Guo X, Fang D, Li L *et al.* 2021. The genome of *Magnolia biondii* Pamp. provides insights into the evolution of Magnoliales and biosynthesis of terpenoids. *Horticulture Research* **8**: 38.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP *et al.* 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**: 92–95.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016a. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Systems* **3**: 99–101.

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016b. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems* **3**: 95–98.

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

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

Fähnrich A, Krause K, Piechulla B. 2011. Product variability of the 'cineole cassette' monoterpene synthases of related *Nicotiana* species. *Molecular Plant* **4**: 965–984.

Forestier ECF, Czechowski T, Cording AC, Gilday AD, King AJ, Brown GD, Graham IA. 2021. Developing a *Nicotiana benthamiana* transgenic platform for high-value diterpene production and candidate gene evaluation. *Plant Biotechnology Journal* **19**: 1614–1623.

Fu AZ, Zheng YY, Guo J, Grierson D, Zhao XY, Wen CL, Liu Y, Li J, Zhang XW, Yu Y *et al.* 2023. Telomere-to-telomere genome assembly of bitter melon (*Momordica charantia* L. var. *abbreviata* Ser.) reveals fruit development, composition and ripening genetic characteristics. *Horticulture Research* **10**: uhac228.

Green SA, Chen X, Nieuwenhuizen NJ, Matich AJ, Wang MY, Bunn BJ, Yauk YK, Atkinson RG. 2012. Identification, functional characterization, and regulation of the enzyme responsible for floral (E)-nerolidol biosynthesis in kiwifruit (*Actinidia chinensis*). *Journal of Experimental Botany* **63**: 1951–1967.

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**: 2896–2898.

Hu J, Fan J, Sun Z, Liu S. 2020. NEXTPOLISH: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**: 2253–2255.

Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, Li J, Sandoval JR, Cooper DN, Ye K *et al.* 2023. An efficient error correction and accurate assembly tool for noisy long reads. *bioRxiv*. doi: 10.1101/2023.03.09.531669.

Hu LS, Xu ZP, Wang MJ, Fan R, Yuan DJ, Wu BD, Wu HS, Qin XW, Yan L, Tan LH *et al.* 2019. The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. *Nature Communications* **10**: 4702.

Jiang SY, Jin J, Sarojam R, Ramachandran S. 2019. A comprehensive survey on the terpene synthase gene family provides new insight into its evolutionary patterns. *Genome Biology and Evolution* **11**: 2078–2098.

Jin J, Kim MJ, Dhandapani S, Tjhang JG, Yin JL, Wong L, Sarojam R, Chua NH, Jang IC. 2015. The floral transcriptome of ylang–ylang (*Cananga odorata* var. *fruticosa*) uncovers biosynthetic pathways for volatile organic compounds and a multifunctional and novel sesquiterpene synthase. *Journal of Experimental Botany* **66**: 3959–3975.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**: 3059–3066.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**: 907–915.

Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. 2019. ORTHODB v.10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research* **47**: D807–D811.

Leebens-Mack J, Raubeson LA, Cui L, Kuehl JV, Fourcade MH, Chumley TW, Boore JL, Jansen RK, Depamphilis RK. 2005. Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. *Molecular Biology and Evolution* **22**: 1948–1963.

Li DS, Hua J, Luo SH, Liu YC, Chen YG, Ling Y, Guo K, Liu Y, Li SH. 2021. An extremely promiscuous terpenoid synthase from the Lamiaceae plant *Colquhounia coccinea* var. *mollis* catalyzes the formation of sester−/di−/sesqui−/mono-terpenoids. *Plant Communications* **2**: 100233.

Li H. 2018. MINIMAP2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li HT, Luo Y, Gan L, Ma PF, Gao LM, Yang JB, Cai J, Gitzendanner MA, Fritsch PW, Zhang T *et al.* 2021. Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biology* **19**: 232.

Li JG, Wang YM, Dong YM, Zhang WY, Wang D, Bai HT, Li K, Li H, Shi L. 2021. The chromosome-based lavender genome provides new insights into Lamiaceae evolution and terpenoid biosynthesis. *Horticulture Research* **8**: 53.

Li Z, Wang J, Zhang X, Zhu G, Fu Y, Jing Y, Huang B, Wang X, Meng C, Yang Q *et al.* 2022. The genome of *Aechmea fasciata* provides insights into the evolution of tank epiphytic habits and ethylene-induced flowering. *Communications Biology* **5**: 920.

Liao Y, Smyth GK, Shi W. 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research* **47**: e47.

Liu F, Xiao Z, Yang L, Chen Q, Shao L, Liu J, Yu Y. 2017. PhERF6, interacting with EOBI, negatively regulates fragrance biosynthesis in petunia flowers. *New Phytologist* **215**: 1490–1502.

Lu B, Zhang FM, Yu FQ, Rinaldi AC. 2022. Ethnobiological notes and volatile profiles of two rare Chinese desert truffles. *Mycology* **13**: 177–184.

Lv Q, Qiu J, Liu J, Li Z, Zhang W, Wang Q, Fang J, Pan J, Chen Z, Cheng W *et al.* 2020. The *Chimonanthus salicifolius* genome provides insight into magnoliid evolution and flavonoid biosynthesis. *The Plant Journal* **103**: 1910–1923.

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

Moore RC, Purugganan MD. 2005. The evolutionary dynamics of plant duplicate genes. *Current Opinion in Plant Biology* **8**: 122–128.

Murat F, Armero A, Pont C, Klopp C, Salse J. 2017. Reconstructing the genome of the most recent common ancestor of flowering plants. *Nature Genetics* **49**: 490–496.

Nurhayani FO, Wulandari AS, Suharsi TK. 2019. The floral morphology and anatomy of kenanga (*Cananga odorata* (Lam.) Hook.f. & Thomson). *IOP Conference Series: Earth and Environmental Science* **394**: 12034.

Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. *Bioinformatics* **21**: i351–i358.

Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH. 2019. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biology* **20**: 38.

Qin L, Hu Y, Wang J, Wang X, Zhao R, Shan H, Li K, Xu P, Wu H, Yan X *et al.* 2021. Insights into angiosperm evolution, floral development and chemical biosynthesis from the *Aristolochia fimbriata* genome. *Nature Plants* **7**: 1239–1253.

Qin XW, Hao CY, He SZ, Wu G, Tan LH, Xu F, Hu RS. 2014. Volatile organic compound emissions from different stages of *Cananga odorata* flower development. *Molecules* **19**: 8965–8980.

Ramya M, Park PH, Chuang Y-C, Kwon OK, An HR, Park PM, Baek YS, Kang B-C, Tsai W-C, Chen H-H. 2019. RNA sequencing analysis of *Cymbidium goeringii* identifies floral scent biosynthesis related genes. *BMC Plant Biology* **19**: 337.

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES *et al.* 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**: 1665–1680.

Rendón-Anaya M, Ibarra-Laclette E, Méndez-Bravo A, Lan T, Zheng C, Carretero-Paulet L, Perez-Torres CA, Chacón-López A, Hernandez-Guzmán G, Chang TH *et al.* 2019. The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen-influenced gene space adaptation. *Proceedings of the National Academy of Sciences, USA* **116**: 17081–17089.

Shen T, Qi H, Luan X, Xu W, Yu F, Zhong Y, Xu M. 2022. The chromosome-level genome sequence of the camphor tree provides insights into Lauraceae evolution and terpene biosynthesis. *Plant Biotechnology Journal* **20**: 244–246.

Shimada T, Endo T, Fujii H, Hara M, Ueda T, Kita M, Omura M. 2004. Molecular cloning and functional characterization of four monoterpene synthase genes from *Citrus unshiu* Marc. *Plant Science* **166**: 49–58.

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.

Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. 2015. Polyploidy and genome evolution in plants. *Current Opinion in Genetics and Development* **35**: 119–125.

Song AA, Abdullah JO, Abdullah MP, Shafee N, Othman R, Tan EF, Noor NM, Raha AR. 2012. Overexpressing 3-hydroxy-3-methylglutaryl coenzyme a reductase (HMGR) in the lactococcal mevalonate pathway for heterologous plant sesquiterpene production. *PLoS ONE* **7**: e52444.

Spitzer-Rimon B, Farhi M, Albo B, Cna'ani A, Zvi MMB, Masci T, Edelbaum O, Yu Y, Shklarman E, Ovadis M *et al.* 2012. The R2R3-MYB-like regulatory factor EOBI, acting downstream of EOBII, regulates scent production by activating *ODO1* and structural scent-related genes in petunia. *Plant Cell* **24**: 5089–5105.

Strijk JS, Hinsinger DD, Roeder MM, Chatrou LW, Couvreur TLP, Erkens RHJ, Sauquet H, Pirie MD, Thomas DC, Cao K. 2021. Chromosome-level reference genome of the soursop (*Annona muricata*): A new resource for Magnoliid research and tropical pomology. *Molecular Ecology Resources* **21**: 1608–1619.

Sun M, Zhang Y, Zhu L, Liu N, Bai H, Sun G, Zhang J, Shi L. 2022. Chromosome-level assembly and analysis of the *Thymus* genome provide insights into glandular secretory trichome formation and monoterpenoid biosynthesis in thyme. *Plant Communications* **3**: 100413.

Sun P, Jiao B, Yang Y, Shan L, Li T, Li X, Xi Z, Wang X, Liu J. 2022. WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Molecular Plant* **15**: 1841–1851.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* **34**: W609–W612.

Talavera A, Fernández-Pozo N, Matas AJ, Hormaza JI, Bombarely A. 2023. Genomics in neglected and underutilized fruit crops: a chromosome-scale genome sequence of cherimoya (*Annona cherimola*). *Plants, People, Planet* **5**: 408–423.

Tan LT, Lee LH, Yin WF, Chan CK, Abdul Kadir H, Chan KG, Goh BH. 2015. Traditional uses, phytochemistry, and bioactivities of *Cananga odorata* (ylang–ylang). *Evidence-based Complementary and Alternative Medicine* **2015**: 896314.

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. 2008. Synteny and collinearity in plant genomes. *Science* **320**: 486–488.

Wang S, Xiao Y, Zhou Z-W, Yuan J, Guo H, Yang Z, Yang J, Sun P, Sun L, Deng Y *et al.* 2021. High-quality reference genome sequences of two coconut cultivars provide insights into evolution of monocot chromosomes and differentiation of fiber content and plant height. *Genome Biology* **22**: 304.

Wang W, Wang MY, Zeng Y, Chen X, Wang X, Barrington AM, Tao J, Askinson RG, Nieuwenhuizen NJ. 2023. The terpene synthase (*TPS*) gene family in kiwifruit shows high functional redundancy and a subset of TPS likely fulfl overlapping functions in fruit favour, foral bouquet and defence. *Molecular Horticulture* **3**: 9.

Xia ZQ, Huang DM, Zhang SK, Wang WQ, Ma FN, Wu B, Xu Y, Xu BQ, Chen D, Zou ZL *et al.* 2021. Chromosome-scale genome assembly provides insights into the evolution and flavor synthesis of passion fruit (*Passiflora edulis Sims*). *Horticulture Research* **8**: 14.

Yang Y, Sun P, Lv L, Wang D, Ru D, Li Y, Ma T, Zhang L, Shen X, Meng F *et al.* 2020. Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nature Plants* **6**: 215–222.

Zhang L, Chen F, Zhang X, Li Z, Zhao Y, Lohaus R, Chang X, Dong W, Ho SYW, Liu X *et al.* 2020. The water lily genome and the early evolution of flowering plants. *Nature* **577**: 79–84.

Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. 2006. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics, Proteomics & Bioinformatics* **4**: 259–263.

Zheng Y, Gao Z, Luo L, Wang Y, Chen Q, Yang Y, Kong X, Yang Y. 2021. Divergence of the genetic contribution of FRIGIDA homologues in regulating the flowering time in *Brassica rapa* ssp. *rapa*. *Gene* **796–797**: 145790.

Supporting Information

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

Fig. S1 Estimate of genome size of *Cananga odorata* based on *K-mer* analysis.

Fig. S2 Hi-C heat map of *Cananga odorata* chromosome interactions.

Fig. S3 BUSCO analysis of genome assembly of *Cananga odorata*.

Fig. S4 Collinearity point map in genomic comparison between *Cananga odorata* and other representing angiosperms.

Fig. S5 The connection patterns of collinearity modules in the genome of *Cananga odorata* and other representative angiosperms.

Fig. S6 Phylogenetic relationships between *Cananga odorata* and other angiosperms based oncodon1/2 sequences of 267 MSCG families.

Fig. S7 Phylogenetic relationships between *Cananga odorata* and other angiosperms based on nucleotites, amino acids and codon1 sequences of 45 SSCG families.

Fig. S8 Phylogenetic relationships between *Cananga odorata* and other angiosperms based on codon2/3 sequences of 45 SSCG families.

Fig. S9 Phylogenetic tree showing the relationships among *Cananga odorata* and 8 other species with differentiation times.

Fig. S10 Distribution of 4DTv shown for the orthologous and paralogous genes between *Cananga odorata* and other indicated species.

Fig. S11 Comparative analysis of syntenic dotplot among genomics.

Fig. S12 Karyotype analysis of *Cananga odorata* ($2n = 16$) based on the Ancestral Angiosperm Karyotype of current angiosperms.

Fig. S13 GC traces of floral VOCs in essential oils extracted from flowers at stage1–4.

Fig. S14 Metabolic analysis of VOCs in the flower scents of *Cananga odorata*.

Fig. S15 Phylogenetic tree of TPS proteins identified in *Cananga odorata* genome and other five sequenced angiosperm genomes.

Fig. S16 Diagram of the benzenoid/phenylpropanoid biosynthesis pathways and the expression profiles of genes encoding synthases involved in these pathways in *Cananga odorata*.

Fig. S17 Identification of the recombinant CoTPS21.2 and CoTPS21.3 proteins.

Fig. S18 Functional characterization of three *CoTPS21* homologues in sesquiterpene biosynthesis in *Nicotiana benthamiana*.

Fig. S19 Functional characterization of *CoTPS-13 515* and *CoTPS-21 750* genes in sesquiterpene biosynthesis.

Fig. S20 Promoter *cis*-element analysis of three *CoTPS21* homologous promoters using the PlantCARE and PlantPAN database web servers.

Fig. S21 Yeast one-hybrid assays of the interaction between three effectors (CoSPL3, CoSPL9.1 and CoSPL9.2) and three CoTPS21 promoters.

Fig. S22 Transcription activity analysis showing the activation activity of two effectors (CoMYC2.1 and CoMYC2.2) at three reporters containing the *CoTPS21* promoters.

Table S1 List of primers used in this study.

Table S2 Genome size estimation based on different K-mer of *Cananga odorata* genome.

Table S3 Summary of the long reads sequencing of *Cananga odorata* genome.

Table S4 Statistics of assembly results of *Cananga odorata* genome using Nanopore sequencing data.

Table S5 Quantification of the mapped illumina and Hi-C reads of *Cananga odorata* genome.

Table S6 Genome information of *Cananga odorata* after Hi-C assembly.

Table S7 General statistics of the telomeres of *Cananga odorata* genome.

Table S8 Predicted centromeric regions of *Cananga odorata* genome.

Table S9 Genome integrity estimation of *Cananga odorata* genome.

Table S10 The mapping rates of transcriptome to the assembled genome of *Cananga odorata*.

Table S11 Genome annotation information of *Cananga odorata*.

Table S12 Comparison of *Cananga odorata* gene set with other species.

Table S13 Database and the corresponding annotated genome of *Cananga odorata*.

Table S14 Information of the repeat sequence in *Cananga odorata* genome.

Table S15 The BGCs analysis of the *Cananga odorata* genome.

Table S16 Summary of the selected 23 genomes used in this study.

Table S17 Summary of the resulted topology using different datasets and phylogenetic methods.

Table S18 Taxon sampling strategies, and phylogenetic software and parameters among different magnoliid genomes.

Table S19 Comparison of expansion and contraction genes in *Cananga odorata* genome with other species.

Table S20 GO analysis of expansion and contraction gene families in *Cananga odorata* genome.

Table S21 Integrated syntenic and phylogenomic dating of *Cananga odorata* intragenomic and *C. odorata* with *Aristolochia*

fimbriata and *C. micrantha* intragenomic syntenic anchor gene pairs.

Table S22 Comparison of karyotype genes in *Cananga odorata* genome with other species.

Table S23 Composition and relative content analysis of VOCs in essential oil of *Cananga odorata* at different floral development stages.

Table S24 Composition and absolute content analysis of VOCs in essential oil of *Cananga odorata* at different floral development stages.

Table S25 The volatile organic compounds compounds in essential oil analyzed by PCA.

Table S26 Composition analysis of volatile organic compounds in flowers of *Cananga odorata* at different development stages.

Table S27 The candidate genes involved in the terpene synthesis pathways in *Cananga odorata*.

Table S28 Comparison of the number of candidate genes in the terpene biosynthesis pathway in *Cananga odorata* and other species.

Table S29 The list of putative terpene synthase genes in *Cananga odorata*.

Table S30 Comparison of the number of *TPS* subfamily genes in *Cananga odorata* and other species.

Table S31 Duplication events of *TPS* gene family in *Cananga odorata*.

Table S32 The candidate genes involved in the benzenoid/phenylpropanoid biosynthesis in *Cananga odorata*.

Table S33 Comparison of the number of candidate genes involved in the benzenoid/phenylpropanoid biosynthesis in *Cananga odorata* and other species.

Table S34 Duplication events of candidate genes involved in the benzenoid/phenylpropanoid biosynthesis in *Cananga odorata*.

Table S35 Transcriptome data of flower fragrance metabolism related genes in *Cananga odorata*.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.