



OPEN ACCESS

EDITED BY

Shaojun Dai,
Shanghai Normal University, China

REVIEWED BY

Liangsheng Zhang,
Zhejiang University, China
Junhua Peng,
Huazhi Bio-Tech Co., Ltd., China

*CORRESPONDENCE

Yunwei Zhou
zhouyunwei1970@163.com
Miao He
hemiao@nefu.edu.cn

[†]These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION

This article was submitted to Functional and Applied Plant Genomics, a section of the journal Frontiers in Plant Science

RECEIVED 24 October 2022

ACCEPTED 24 November 2022

PUBLISHED 21 December 2022

CITATION

Chen L, Xia B, Li Z, Liu X, Bai Y, Yang Y, Gao W, Meng Q, Xu N, Sun Y, Li Q, Yue L, He M and Zhou Y (2022) *Syringa oblata* genome provides new insights into molecular mechanism of flower color differences among individuals and biosynthesis of its flower volatiles. *Front. Plant Sci.* 13:1078677. doi: 10.3389/fpls.2022.1078677

COPYRIGHT

© 2022 Chen, Xia, Li, Liu, Bai, Yang, Gao, Meng, Xu, Sun, Li, Yue, He and Zhou. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Syringa oblata genome provides new insights into molecular mechanism of flower color differences among individuals and biosynthesis of its flower volatiles

Lifei Chen^{1†}, Bin Xia^{2†}, Ziwei Li², Xiaowei Liu², Yun Bai¹, Yujia Yang², Wenjie Gao³, Qingran Meng⁴, Ning Xu⁵, Ying Sun², Qiang Li², Liran Yue², Miao He^{2*} and Yunwei Zhou^{1*}

¹College of Horticulture, Jilin Agricultural University, Changchun, China, ²College of Landscape Architecture, Northeast Forestry University, Harbin, China, ³School of Ecological Technology and Engineering, Shanghai Institute of Technology University, Shanghai, China, ⁴School of Perfume and Aroma Technology, Shanghai Institute of Technology University, Shanghai, China, ⁵School of Forestry, Northeast Forestry University, Harbin, China

Syringa oblata is a high ornamental value tree owing to its elegant colors, unique aromas and wide adaptability, however, studies on the molecular mechanism underlying the formation of its ornamental traits are still lacking. Here, we presented a chromosome-scale genome assembly of *S. oblata* and the final genome size was 1.11 Gb with a contig N50 of 4.75 Mb, anchored on 23 chromosomes and was a better reference for *S. oblata* transcriptome assembly. Further by integrating transcriptomic and metabolic data, it was concluded that *F3H*, *F3'H*, *4CL* and *PAL*, especially the *F3'H*, were important candidates involved in the formation of floral color differences among *S. oblata* individuals. Genome-wide identification and analysis revealed that the TPS-b subfamily was the most abundant subfamily of TPS family in *S. oblata*, which together with the CYP76 family genes determined the formation of the major floral volatiles of *S. oblata*. Overall, our results provide an important reference for mechanistic studies on the main ornamental traits and molecular breeding in *S. oblata*.

KEYWORDS

genome assembly, flower color formation, individuals, volatiles biosynthesis, *Syringa oblata*

Introduction

Syringa oblata is among the high ornamental value trees in the middle latitudes of Eurasia and North America and it has been cultivated in China for more than 1,000 years owing to its elegant colors, unique aromas and wide adaptability (Li et al., 2006). In previous studies, the majority of research on *S. oblata* has focused on the determination of volatile components, identification of active substances and transcriptomics. For example, previous studies of transcriptomic analysis in *S. oblata* revealed that molecular mechanism of the accumulation of flavonoids in response to different light intensity. cDNA libraries from three inflorescence developmental stages were sequenced and assembled to obtain differentially expressed genes involved in floral pigment biosynthesis and fragrance metabolism in *S. oblata*. (Zheng et al., 2015; Cui et al., 2019; Liu et al., 2019; Gegen et al., 2022; Tai et al., 2022). Despite the high economic and ornamental values of *S. oblata*, its genetic and breeding research still lag behind other flowering plants.

S. oblata flowers usually show varying degrees of purple color on different individuals (Zhang, 2011). The delphinidin-3-O-rutinoside and cyanidin-3-O-rutinoside are the main anthocyanins currently reported for *S. oblata* petal coloration (Zhang, 2011; Ma et al., 2022), and their syntheses are involved in the phenylpropanoid biosynthesis and anthocyanin biosynthesis pathways. The precursor of anthocyanin biosynthesis, Phenylalanine, forms anthocyanins after a series of catalysis by enzymes such as phenylalanine lyase (PAL), 4-coumarate-CoA ligase (4CL), etc., followed by the formation of glycosidic bonds under the action of glucosyltransferase and conversion to stable anthocyanins (Holton and Cornish, 1995; Zhang et al., 2014). These genes have been reported to be intimately associated with the formation of purple or blue flowers in plants. For example, naringenin 3-dioxygenase (*F3H*) mutant lines of *Petunia* obtained via Cas9-ribonucleoproteins delivery exhibit a paler purple (Yu et al., 2021), flower-specific expression of the *Phalaenopsis* flavonoid 3', 5'-hydroxylase (*F3'5'H*) modifies flower color pigmentation in *Petunia* and *Lilium* (Qi et al., 2013). Previous studies on the color of *S. oblata* flowers have elucidated some transcription factors associated with petal color changes during flower development (Ma et al., 2022), however, the molecular mechanism underlying the formation of differences in floral color among *S. oblata* individuals is still lacking, which is of great significance for its molecular breeding of flower color.

S. oblata has a strong floral fragrance, and previous studies have revealed that the main volatiles of its flowers are terpenoids such as lilac aldehyde, lilac alcohol, α -pinene, sabinene and β -pinene (Li et al., 2006; Gegen et al., 2022; Wang et al., 2022), while terpene synthase (TPS) is the pivotal gene family for the biosynthesis of plant terpenes (Chen et al., 2011). Plant secondary metabolites usually require post-modification processes to produce biologically active metabolites (Li et al., 2009) and many *S. oblata* flower volatiles undergo post-

modification processes, such as lilac aldehyde and lilac alcohol, which are oxygenated derivatives of linalool formed through the catalysis of cytochrome P450 (Raguso, 2016). Former studies on the volatiles of *S. oblata* have mainly focused on the determination of composition (Yao-zu et al., 1987; Li et al., 2006; Wu et al., 2021; Gegen et al., 2022), while little attention has been paid to the molecular mechanism of their synthesis and post-modification processes. Therefore, genome-wide identification and analysis of the TPS and post-modification genes of the volatiles of *S. oblata* are important references for revealing the molecular mechanism of its volatile biosynthesis.

As genome sequencing, assembly and annotation methods continue to evolve and progress, new plant reference genomes are constantly being published (Sun et al., 2022). In our study, we reported a chromosome-scale *S. oblata* genome with a longer contig and it was a better reference for its transcriptome assembly. Meanwhile, The combination of the metabolic results and transcriptomic data of *S. oblata* individuals with different colors were used to screen the key genes regulating the flower color. In addition, *S. oblata* TPS and related post-modification genes were genome-wide identified and analyzed to elucidate the molecular mechanism of its volatile biosynthesis. The genome sequence analysis, transcriptome and metabolic data from this study are significant references for studying the mechanism of important ornamental traits and breeding in *S. oblata*.

Materials and methods

Plant material and genome sequencing

The sequenced *S. oblata* was maintained at the Harbin Institute of Landscape Architecture and Greening Science in good growing condition, and its young leaves were collected for whole-genome and Hi-C sequencing. On the same day of the flowering period, petals at the expanding flower bud stage were collected from *S. oblata* individuals (Sob1, Sob2, Sob3) of different flower colors for RNA-seq, and their petal colors were compared using NCS color cards during the blooming period, and petals were also collected for anthocyanin determination. Plants were immediately stored in liquid nitrogen until they were shipped back to the laboratory and stored at -80°C .

Genomic DNA was extracted from leaves of *S. oblata* using the DNasecure Plant Kit (TIANGEN, China). DNA sequencing libraries were constructed based on the Illumina library preparation protocols. Paired-end (PE150) library with insert size of 350 bp was constructed according to the manufacturer's instructions and sequencing using the Illumina Hiseq X-ten (Illumina, USA). Then the raw reads were filtered out the adapter sequences and the low-quality and duplicated reads to obtain clean reads. For Pacbio libraries, SMRT Bell libraries with an insert size of 60 kb were constructed and then sequenced on the PacBio Sequel platform (Pacific Biosciences, USA) using the

P6 polymerase/C4 chemistry combination, based on the manufacturer's procedure. A Hi-C library was generated following the approach described by Lieberman-Aiden et al. (Lieberman-Aiden et al., 2009). Briefly, purified DNA was obtained by formaldehyde fixation of chromatin, cell lysis, Hind III nuclease digestion, recovery, ligation and protein removal by protease. The purified DNA was sheared into 350-bp fragments and ligated to adaptors. The fragments labeled with biotin were extracted using streptavidin beads and after PCR enrichment, the libraries were sequenced on Illumina HiSeq X instrument (Illumina, USA).

Genome survey and assembly

Iteratively selected 17 bp base sequences (K -value=17) were used for K -mer analysis. K -mer frequency distributions were counted and K -mer depth distribution curves were calculated and then GenomeScope (Vurture et al., 2017) was used to assess genome size, percentage of repetitive sequences, and heterozygosity ratio.

De novo assembly of the long reads from the PacBio SMRT Sequencer was performed using wtdbg2 (Luo et al., 2012; Md et al., 2019). The Fuzzy Bruijn Graph algorithm was used to assemble and integrate 1,024bp sequences from the reads into vertex sequences, and then based on the position of the vertex sequences on the reads, the vertex sequences were concatenated to obtain the genome sequences. Subsequently, the Hi-C sequencing data were aligned to the assembled scaffolds by BWA-mem (Md et al., 2019) and the scaffolds were clustered onto chromosomes with LACHESIS (Burton et al., 2013). A BUSCO analysis was conducted to determine genome completeness using BUSCO (Simão et al., 2015) together with the embryophyta odb10 database.

RNA-seq and assembly

Total RNA was extracted using RNAPrep Pure Plant Kit (TIANGEN, China). The library insert size was 300 bp and more than 6 Gb of 150-bp pair-end clean reads were obtained using the Illumina HiSeq novoseq 6000 platform (Illumina, USA). Transcriptome public data of *S. oblata* was downloaded from SRA database. Trinity (Grabherr et al., 2011) was used to assemble the transcriptome without reference and HISAT2 (Kim et al., 2019) was used to assemble the transcriptome with reference, selecting the default parameters.

Genome annotation

A combined strategy based on homology comparison and ab initio search was used for repeat sequence annotation. Tandem repeats were extracted by ab initio prediction using TRF

(Benson, 1999). The homolog prediction used Repbase database (Jurka et al., 2005) employing RepeatMasker (Chen, 2004) to extracted repeat regions. And ab initio prediction built *de novo* repetitive elements database by LTR_FINDER (Xu and Wang, 2007), RepeatScout (Price et al., 2005), and RepeatModeler2 (Flynn et al., 2020), then was supplied to RepeatMasker for DNA-level repeat identification.

Gene structural annotation was based on ab initio prediction, homology-based prediction and RNA-Seq assisted prediction. For Ab initio prediction, Augustus (Stanke and Morgenstern, 2005) AND GlimmerHMM (Majoros et al., 2004) were used. Proteins sequences of *Antirrhinum majus*, *Arabidopsis thaliana*, *Fraxinus excelsior*, *Olea europaea*, *Sesamum indicum* and *Solanum tuberosum* were aligned using TblastN (E-value $\leq 1e-5$) (Camacho et al., 2009) with the genome of *S. oblata* and then the matching proteins were aligned to the homologous genome sequences for accurate spliced alignments with GeneWise (Madeira et al., 2022). Transcriptome reads assemblies were generated with Trinity for the genome annotation. The non-redundant reference gene set was generated by merging genes predicted by three methods with EvidenceModeler (Haas et al., 2008). The protein-coding genes functions were assigned using BLAST against public protein databases, including NR (<https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/>), SwissProt (Boeckmann et al., 2003), KEGG (Kanehisa and Goto, 2000), InterPro (Jones et al., 2014), GO (Ashburner et al., 2000) and Pfam (Finn et al., 2013).

The tRNAs were predicted using tRNAscan-SE (Lowe and Eddy, 1997). The rRNAs were predicted by using BLAST against rRNA sequences of relative species. Other ncRNAs, including miRNAs, snRNAs were identified by searching against the Rfam (Griffiths-Jones et al., 2003) using the infernal (Nawrocki and Eddy, 2013).

Comparative genomic and phylogenetic analysis

Orthologous relationships between genes of *S. oblata* and 13 other species including *Oryza sativa* (Yu et al., 2002), *O. europaea* (Unver et al., 2017), *F. excelsior* (Sollars et al., 2017), *Vitis vinifera* (Jaillon et al., 2007), *S. indicum* (Wang et al., 2014), *A. majus* (Li et al., 2019), *A. thaliana* (The Arabidopsis Genome, 2000), *Brassica napus* (Chalhoub et al., 2014), *Populus trichocarpa* (Tuskan et al., 2006), *Solanum lycopersicum* (Sato et al., 2012), *P. axillaris*, *P. inflata* (Bombarely et al., 2016) and *S. tuberosum* (Xu et al., 2011) were inferred through all-against-all protein sequence similarity searches with OthoMCL (Li et al., 2003) and retained only the longest predicted transcript per locus. Single-copy gene families were subjected to MUSCLE (Edgar, 2004) multiple sequence alignment and ambiguously aligned positions were trimmed using Gblocks (Dereeper et al., 2008) and the tree was inferred using RAxML (Stamatakis, 2014). Divergence time estimation was performed using MCMCTree (<http://abacus.gene.ucl.ac.uk/>

software/paml.html) in the PAML package (Yang, 1997) with time correction points for *P. trichocarpa* – *A. thaliana* (97–109 Mya), *P. trichocarpa* – *V. vinifera* (107–135 Mya), *A. majus* – *A. thaliana* (111–131 Mya), and *B. napus* – *O. sativa* (148–173 Mya), respectively. The likelihood model in the software package Café' (De Bie et al., 2006) was used to perform gene families expansion and contraction analysis.

Genome synteny and whole-genome duplication analysis

BLASTP (E-value $\leq 1e-5$) analysis of proteins was performed between and within *Jasminum sambac* (Xu et al., 2022), *S. oblata* and *O. europaea* and *V. vinifera*, and then based on the location of genes and blast results using MCscanX (Wang et al., 2012), searching for co-linear regions within or between genomes. The Ks values of ortholog pairs or paralog pairs were plotted using the WGDI (Sun et al., 2021), and plotted the frequency distribution of Ks versus gene pairs. Synteny dot plots were plotted by dot-plotter in McscanX.

Identification of anthocyanin biosynthesis genes, TPS genes and CYP450 genes

Genes potentially involved in anthocyanin biosynthesis (ko00940, ko00942) were identified in *S. oblata* and other representative plants based on the KAAS web service in KEGG (Moriya et al., 2007). The PF01397 and PF03936 files were used to Hmsearch for *S. oblata* and other related species protein data, respectively ($P < 10.0$), and the two datasets were combined and de-redundant to obtain the Hmsearch dataset (Gbenro et al., 2020). To make the dataset more complete, a blastp search of the *S. oblata* and other related species protein libraries was performed using the *A. thaliana* TPS protein sequences (Aubourg et al., 2002) to obtain the homologous sequence dataset, which was further combined and de-redundant with the Hmsearch dataset, and the dataset was structurally verified using the Pfam to obtain structurally complete TPS genes. Hmsearch (PF00067) and homology comparisons based on *A. thaliana* CYP450 proteins (Paquette et al., 2000) were used to identify CYP450 gene family members of *S. oblata*, and further Pfam was used to verify the structural completeness of the identification results. TBtools was used to Chromosome localization and mapping (Chen et al., 2020).

Determination of *S. oblata* flower anthocyanin and volatiles

Samples for the determination of anthocyanins from *S. oblata* petals were prepared with reference to the method in Zhang's

published paper (Zhang, 2011). The chromatographic conditions were as follows: the chromatographic instrument was a WATERS ACQUITY UPLC (Waters, USA), the column was BEH C18 (2.1X150mm 1.7 μ m), the mobile phase was acetonitrile (A) - 0.1% formic acid (B), gradient elution, the flow rate was 0.3ml/min, the column temperature was 45°C, the injection volume was 5 μ l, the wavelength was 200–500nm, the mass spectrometer was WATERS MALDI SYNAPT Q-TOF MS (Waters, USA), ion source was ESI, ion scan mode was ESI+, scan range was 20–2000 m/z, detection voltage was 1800 V, Desolvation Gas Flow: 700 lit/hr, Cone Gas Flow (L/Hr): 50 lit/hr. The standard curve was prepared using anthocyanin standards for quantitative analysis.

The quick-frozen flowers were ground to a powder in liquid nitrogen and 1.0 g was weighed into a headspace flask and held at 40°C for 30 min for equilibration. Fifteen microliters of 2-octanol (CAS: 123-96-6) was added as an internal standard. Headspace extraction was performed using divinylbenzene/carboxy/polydimethylsiloxane fibers (MilliporeSigma, USA) for 30 min, followed by desorption at 250°C for 3 min using a Model 7890A-5975C equipped with an HP-INNOWAX column (Agilent, USA) with nitrogen as the carrier gas at a flow rate of 3 ml/min without splitting. The temperature setting of the GC was: from 40°C to 120°C at a rate of 4°C/min; to 180°C at a rate of 2°C/min; then to 260°C at a rate of 20°C and held for 10 minutes. Scans were performed from m/z 35 to 550 in electron impact mode at 70 eV. All analyses were performed in triplicate.

Result

Genome sequencing, assembly and annotation

The genome size of *S. oblata* was estimated to be 1,232.83 Mb with heterozygosity of 1.32% and a repeat content of 66.56% on the basis of k-mer statistics (Supplementary Table 1 and Supplementary Figure 1). In total, 146.00 Gb of PacBio long reads (~118.43X coverage of the genome) and 53.00 Gb of Illumina clean reads (~42.99X coverage of the genome) were generated (Supplementary Table 2). The total length of the final assembly was 1,112.64 Mb with 2,417 contigs and a contig N50 of 4.75 Mb (Table 1). The genome size was close to the results based on genome surveys. The assembly was further improved with Hi-C data, of the total scaffold sequences, 98.60% was anchored to 23 chromosome-level pseudomolecules (Supplementary Table 3 and Supplementary Figure 2). The assessment of *S. oblata* genome completeness with BUSCO indicated that 92.00% of the 1,440 BUSCO groups were present in the genome assembly (Supplementary Table 4), and CEGMA results indicated 96.37% coverage of the conserved core eukaryotic genes (Supplementary Table 5). Additionally, 98.75% of the Illumina short reads were mapped to the genome

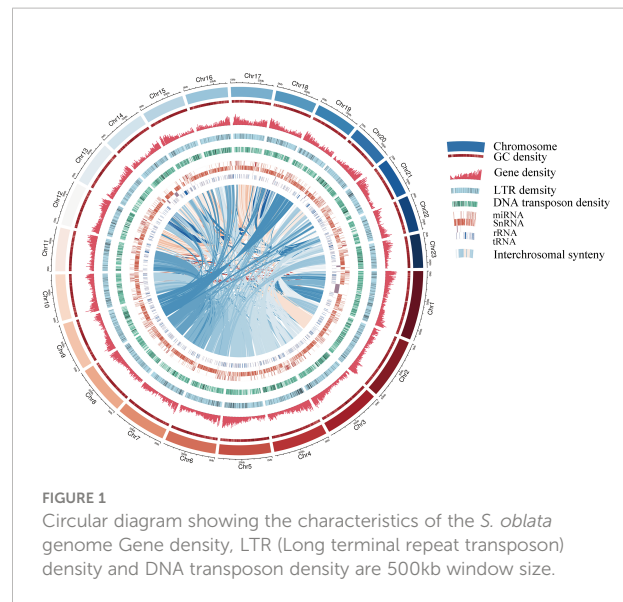
Table 1 The major characteristics of *S. oblata* genome.

Parameter	Number/Size
Estimated genome size (Gb)	1.23
Assembled genome size (Gb)	1.11
Chromosome-anchored scaffolds (Gb)	1.10
Karyotype (chromosomes, 2n)	46
N50 of contigs (Mb)	4.75
Number of contigs	2,417
Longest contigs (Mb)	21.49
N50 of scaffolds (Mp)	44.71
Longest scaffolds (Mb)	70.97
GC content (%)	34.67
Repeat sequences (%)	56.64
Number of protein-coding genes	42,531
Number of non-coding RNAs	6,427

assembly, and 97.93% of genome was covered by reads (Supplementary Table 6), further supporting a high base accuracy of the *S. oblata* genome assembly.

A pipeline combining *de novo* predictions, homology-based predictions, and RNA sequencing data were used to construct gene models for the *S. oblata* genome. A total of 42,531 genes were annotated this way, with an average length of 4,103 bp and an average coding sequence length of 1,082 bp, similar to those of other reported plants (Supplementary Table 7 and Supplementary Table 8). The resulting protein models were then compared with protein sequences in six protein databases; NR, SwissProt, KEGG, InterPro, GO and Pfam. We found that 40,621 (95.50%) gene products could be annotated by at least one of the databases (Supplementary Table 9). The assembled draft *S. oblata* genome contained 56.64% repetitive sequences. Long terminal repeats (LTRs) of retroelements were the most abundant interspersed repeat, occupying 52.07% of the genome, followed by Tandem repeats at 5.30% (Supplementary Table 10). Genes annotated as encoding non-coding RNAs (ncRNAs) in the current genome included 873 microRNAs (miRNAs), 732 transfer RNAs (tRNAs), 295 ribosomal RNAs (rRNAs), and 4,527 small nuclear RNAs (snRNAs) (Supplementary Table 11). These results further supported the completeness of *S. oblata* genome sequence and a schematic representation of the genome was given in Figure 1.

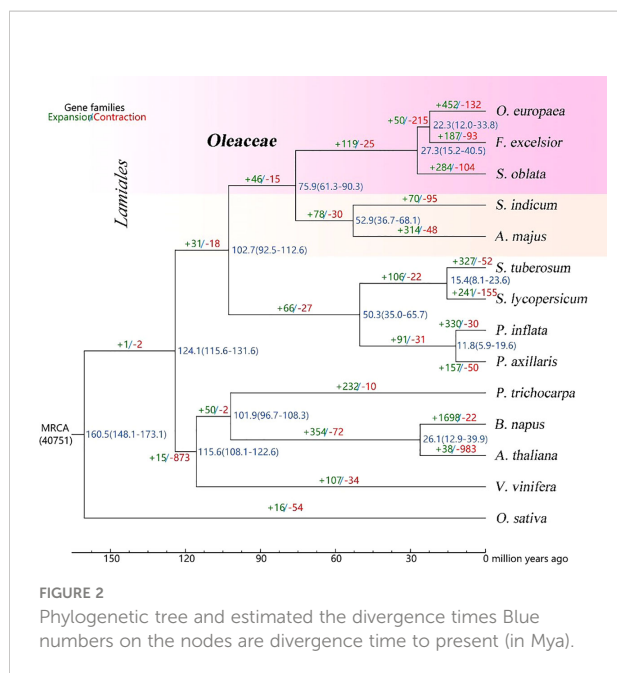
Compared with other published *S. oblata* genomes, our genome had a higher Contig N50 with its longest contig reaching 21.49 Mb, which was superior to other *S. oblata* genomes (Supplementary Table 12). To further validate the quality of our genome assembly and annotation, the published *S. oblata* transcriptome data were assembled with our and other published genomes as reference genome. The results showed that the transcriptome assembled with our genome as the reference genome had a higher alignment rate, especially a higher aligned one-time rate (Supplementary Table 13).



Comparative genomic and phylogenomic analyses

The gene families were analyzed for the genomes of *S. oblata* and 13 other representative species. Based on the analysis of gene family clustering, a total of 40,768 gene families were identified, of which 7,023 were shared by all 14 species, and 41 of these shared families were single-copy gene families (Supplementary Figure 3). Further, phylogenetic tree and estimating divergence times analysis were performed based on 615 single-copy genes. (Figure 2). Phylogenomic analysis showed that *S. oblata* was most related to the ancestor of *O. europaea* and *F. excelsior*, with an estimated divergence time of 27.3 million years ago (MYA). The expanded and contracted gene families of the 14 plant species were compared with their most recent common ancestor (MRCA), *S. oblata* has gained more gene families (284) than it lost (104), among these genes. KEGG studies based on the 284 expanded gene families showed enrichment of genes including those encoding Phenylpropanoid biosynthesis, Sesquiterpenoid and triterpenoid biosynthesis and Biosynthesis of secondary metabolites were enriched (Supplementary Table 14), suggesting the secondary metabolism of *S. oblata* was enhanced in evolution. Patterns of gene-family sharing among Lamiales species were shown in Supplementary Figure 4 for 16,525 families, including 33,022 proteins, of which 338 families were *S. oblata*-specific, containing 3,635 proteins. Interestingly, KEGG analyses found these *S. oblata*-specific genes were particularly enriched in the terms Phenylpropanoid biosynthesis, Sesquiterpenoid and triterpenoid biosynthesis, Degradation of aromatic compounds (Supplementary Table 15).

To resolve the genome-wide duplication events of *S. oblata*, three Oleaceae species (*J. sambac*, *S. oblata* and *O. europaea*) and *V. vinifera* were selected for comparative genomic analysis.

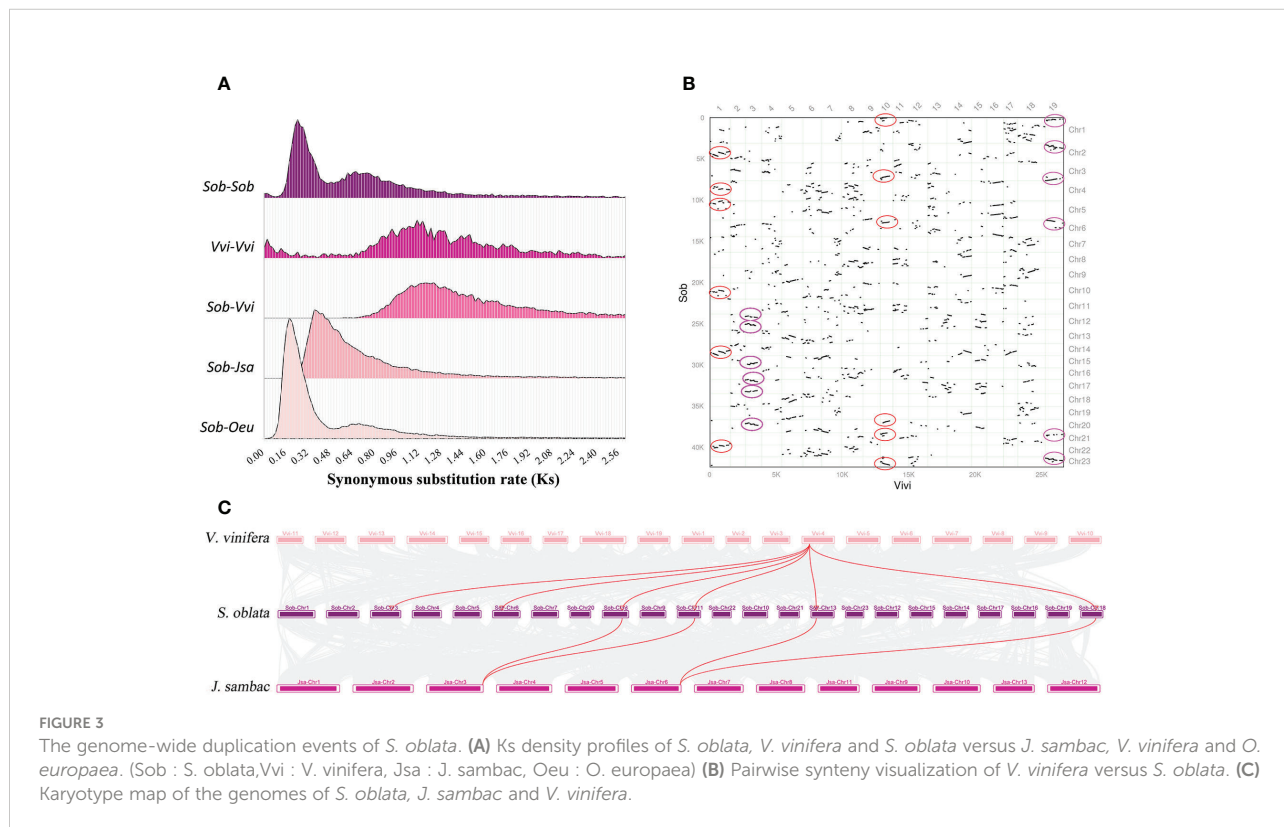


Based on the calculated Ks distribution, two characteristic peaks appeared in the Ks distribution of *S. oblata*, which was smaller than the Ks peaks of *V. vinifera* and *S. oblata* versus *V. vinifera*, indicating that two independent genome-wide duplication events occurred in *S. oblata* after divergence from *V. vinifera*

(Figure 3A). By comparing the Ks value profiles of *S. oblata* and *S. oblata* versus *J. sambac*, *O. europaea*, it was concluded that the first whole-genome duplication event of *S. oblata* occurred before divergence from *J. sambac* and the other shared with *O. europaea* occurred after divergence from *J. sambac*. Further analysis of homologous gene synteny between *S. oblata* and *V. vinifera*, *S. oblata* and *J. sambac* revealed six homologous blocks were found in some regions of the *V. vinifera* genome and a 2:1 relationship between *S. oblata* and *J. sambac*. (Figures 3B, C). In summary, after divergence from *V. vinifera*, *S. oblata* and *J. sambac* shared a common whole-genome triplication (WGT) event before divergence, and after divergence from *J. sambac*, *S. oblata* and *O. europaea* shared a whole-genome duplication (WGD) event.

Analysis of molecular mechanism of floral color differences among *S. oblata* individuals

In our study, the contents of delphinidin-3-O-rutinoside and cyanidin-3-O-rutinoside in the petals of *S. oblata* individuals were quantified and their contents were found to be significantly and positively correlated with the purple-deepening (Figure 4A). By homology matching, 50 genes from 8 gene families involved in the biosynthesis of cyanidin-3-O-rutinoside and delphinidin-



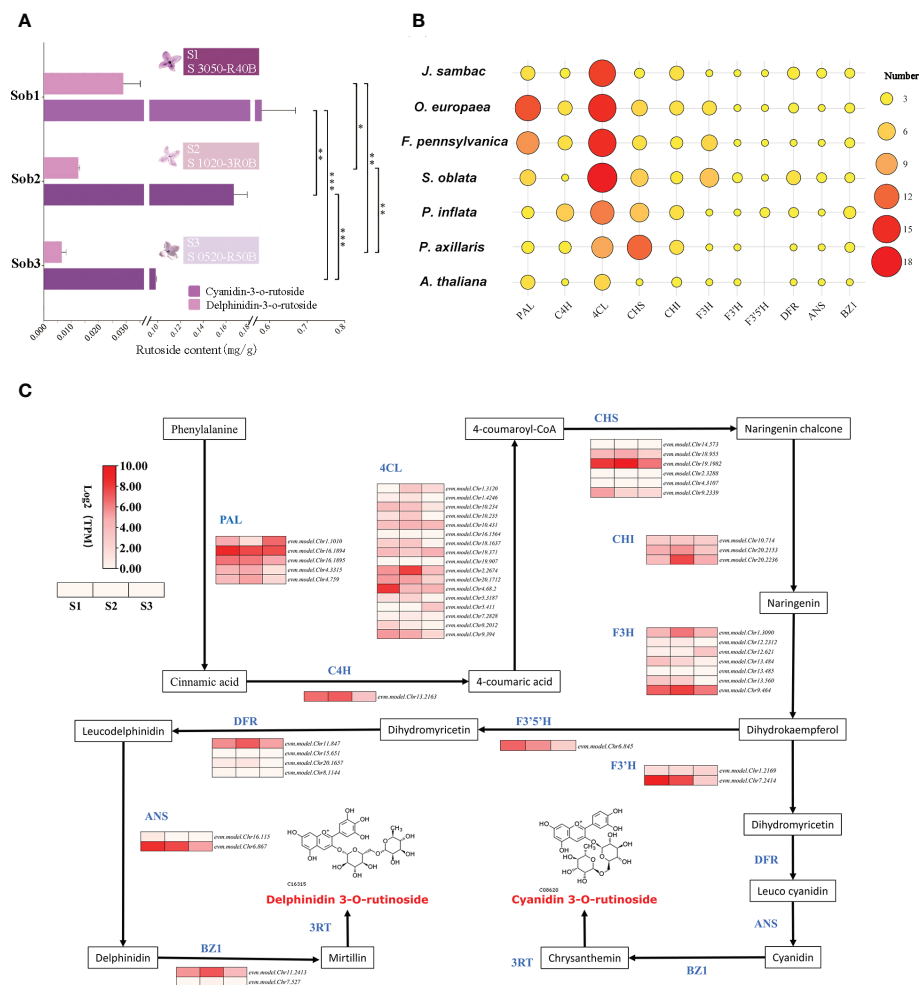


FIGURE 4

Analysis of molecular mechanism of floral color differences among *S. oblata* individuals. (A) The content of delphinidin-3-O-rutinoside and cyanidin-3-O-rutinoside of different *S. oblata* individuals. (B) Heat map of the number of delphinidin-3-O-rutinoside and cyanidin-3-O-rutinoside biosynthesis-related genes in different species. (C) The biosynthesis pathway and the expression profile for delphinidin-3-O-rutinoside and cyanidin-3-O-rutinoside in *S. oblata* individuals. T test (*: $P \leq 0.05$, **: $P \leq 0.01$, ***: $P \leq 0.001$) was used for statistical analyses; mean \pm sd.

3-O-rutinoside were annotated and four gene families (*4CL*, *F3H*, *F3'H*, *DFR*) were increased in *S. oblata* compared to other species (Figure 4B; Supplementary Table 16). By analyzing the co-expression patterns of each synthase genes of *S. oblata* individuals, nine candidate genes affecting flower color differences among individuals of *S. oblata* were predicted (Figure 4C; Supplementary Table 16, higher expression in S1 than in S2, S3). Further correlation analysis of the contents of delphinidin-3-O-rutinoside and cyanidin-3-O-rutinoside with the expression pattern of synthase genes resulted in *evm.model.Chr13.484* (*F3H*), *evm.model.Chr7.2414* (*F3'H*), *evm.model.Chr8.2012* (*4CL*) and *evm.model.Chr16.1894* (*PAL*) showed significant ($p < 0.05$) positive correlations with their contents (Supplementary Figure 5). In addition, the expression pattern of synthase was analyzed by clustering analysis, and

evm.model.Chr7.2414 (*F3'H*) showed the greatest distance from all other genes and the greatest variation among different individuals (Supplementary Figure 6). The above results suggested that *F3H*, *F3'H*, *4CL* and *PAL*, especially the *F3'H*, were important candidates for regulating flower color differences among *S. oblata* individuals.

Analysis of molecular mechanism of biosynthesis of *S. oblata* flower volatiles

The flower of *S. oblata* had a fragrant aroma. In this study, the volatile components of *S. oblata* were determined and terpenes were found to be the most abundant volatiles (Figure 5B). Based on homologous sequence comparison and

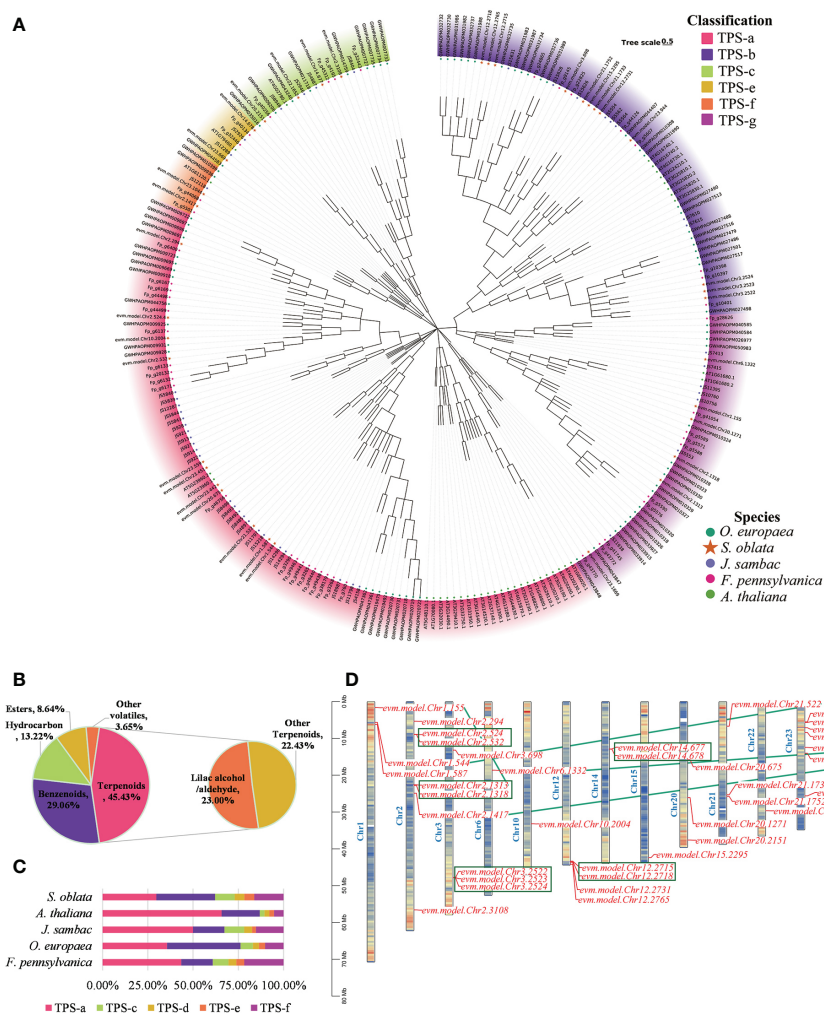


FIGURE 5

Analysis of molecular mechanism of biosynthesis of *S. oblata* flower volatiles. (A) Evolutionary trees of *A. thaliana*, *J. sambac*, *O. europaea* and *F. pennsylvanica* TPS genes. (B) The volatile components of *S. oblata*. (C) Percentage of TPS subgroups in different species. (D) Chromosome location map of TPS genes in *S. oblata*.

Pfam search, 37 TPS genes in *S. oblata* were identified, and evolutionary trees were further constructed with TPS genes of *A. thaliana*, *J. sambac*, *O. europaea* and *F. pennsylvanica* (Figure 5A; Supplementary Table 17). TPS genes of *S. oblata* were classified into six subgroups, TPS-a, TPS-b, TPS-c, TPS-e, TPS-f and TPS-g. The TPS-b subgroup had the highest percentage of genes of total TPS genes in *S. oblata* compared to other species, accounting for 32.43% (*A. thaliana*: 21.05%, *J. sambac*: 17.39%, *O. europaea*: 32.00%, and *F. pennsylvanica*: 17.39%) (Figure 5C). Nineteen *S. oblata* TPS genes (51.35% of all genes) underwent duplication, 10 of these genes underwent recent tandem duplication events, forming five gene clusters (green box markers) on chromosomes 2, 3, 12 and 14, and 10 genes underwent chromosomal duplication events, with tandem

and chromosomal duplication resulting in TPS genes expansion in the *S. oblata* genome (Figure 5D).

The most abundant of the *S. oblata* terpene volatiles were Lilac aldehyde and lilac alcohol (Figure 5B), which were derivatives of the monoterpene linalool. Linalool was catalyzed by successive oxygenation of P450 superfamily genes to Lilac aldehyde and lilac alcohol. A total of 287 P450 genes were identified in *S. oblata* genome, and by further constructing an evolutionary tree with the TPS genes of *A. thaliana*, the P450 genes of *S. oblata* were classified into 9 classes and 50 families (Supplementary Figure 7; Supplementary Table 18). Genes of the CYP76 family were important genes for catalyzing the formation of Lilac aldehyde and lilac alcohol from linalool, and 14 CYP76 family members, of which 8 genes underwent tandem

duplication and formed three gene clusters on chromosomes 4, 10 and 12 (red gene markers), and tandem duplication resulted in the amplification of *S. oblata* CYP76 family members (Supplementary Figure 8).

Discussion

In view of the importance of *S. oblata* in horticulture and the need for further research, the genome of *S. oblata* will provide an important reference for elucidating the molecular mechanisms of its major ornamental traits and molecular breeding. Our study, we sequenced, assembled and annotated the genome of *S. oblata*. The combination of Second- and Third-generation sequencing and Hi-C technology has shown high efficiency in assembling complex plant genomes (Kersey, 2019). The genome of *S. oblata* has high heterozygosity (1.32%) and highly repetitive sequences (56.64%). The *S. oblata* genome was assembled in our study by combining the Illumina short reads and the PacBio long reads sequenced with a contig N50 of 4.75 Mb and a largest contig of 21.49 Mb. Its contig N50 was better than those genomes of *O. europaea*, *F. excelsior* and *O. fragrans* and other published *S. oblata* genomes. Further, those sequences were scaffolded into 23 chromosomes using Hi-C. The reference-based transcriptome assembly strategy requires a reference genome and the quality of the genome will directly influence the accuracy of the assembled transcripts (Martin and Wang, 2011). Compared with other published *S. oblata* genomes, the transcriptome assembly using our genome as reference possessed a higher alignment rate, indicating that our genome was a better reference for assembling *S. oblata* transcriptome.

Whole-genome duplication provides species with abundant evolutionary material that enhances species diversity and environmental adaptability (Sémon and Wolfe, 2007). Recent study has shown that a recent whole-genome triplication event occurred in *J. sambac* (Xu et al., 2022). In our study, we identified that the *S. oblata* has experienced one WGT event and one WGD event after the WGT- γ using Ks value profiles and synteny analysis, and among them, one was shared with *J. sambac* (WGT) and another occurred divergence with *J. sambac* (WGD).

The flowers of *S. oblata* had an elegant purple color and anthocyanins mainly contribute to its coloration, while delphinidin-3-O-rutinoside and cyanidin-3-O-rutinoside were the most abundant anthocyanins reported in *S. oblata* currently (Zhang, 2011; Ma et al., 2022). In this study, by transcriptome sequencing and quantitative analysis of anthocyanins in the petals of different *S. oblata* individuals, *F3H*, *F3'H*, *4CL* and *PAL* were screened as important candidates for the regulation of anthocyanin contents, of which *PAL* and *4CL* acted as key rate-

limiting enzymes in the phenylpropane biosynthesis pathway and functioned in flower or fruit color formation in many species, such as the different deletion mutation in the *PAL* gene affected the peel color of *Mangifera indica* (Zhao et al., 2022), *4CL* transcript levels affected metabolite flux of anthocyanin in *Narcissus tazetta* Flower (Yang et al., 2021), while *F3H* has been reported to regulate the purple color of petals in *Petunia* (Yu et al., 2021). Notably, further clustering analysis showed that the expression pattern of *F3'H* differed particularly in different *S. oblata* individuals, and *F3'H* played an influential role in regulating purple traits in flowers or fruits of different individuals of the same species, such as the red mutant of the dark purple *Petunia* *Ipomoea tricolor* lacked the *F3'H* gene (Hoshino et al., 2003), and the increased dose of the recessive allele in the *F3'H* gene shifted the major anthocyanin in the storage roots of *Ipomoea batatas* from purplish cyanidin derivatives to reddish pelargonidin derivatives (Tanaka et al., 2019). As a result, we speculated that the differences in flower color among *S. oblata* individuals was regulated by a combination of *F3H*, *F3'H*, *4CL* and *PAL* genes, emphatically the *F3'H* gene may play a key role.

Depending on the structure, plant floral volatiles can be divided into three main groups: terpenoids, phenylpropanoids/benzenoids and fatty acids (Muhlemann et al., 2014), and previous studies indicated that terpenes were the main components of *S. oblata* volatiles (Li et al., 2006; Gegen et al., 2022; Wang et al., 2022). Lilac aldehyde and lilac alcohol were the main volatiles of the *S. oblata* flowers found in our study which were oxygenated derivatives of the monoterpenoid linalool (Kreck et al., 2003; Boachon et al., 2015). Whereas the TPS-b subfamily is responsible for the synthesis of monoterpenes in angiosperms (Jia et al., 2022), members of the CYP76 family were shown to be involved in the sequential oxygenation-catalyzed biosynthesis of linalool to lilac aldehyde and lilac alcohol (Boachon et al., 2015). Yan et al. collected and analyzed a new monoterpene synthase gene, named *SoLIM*, which was classified as TPS-b subfamily member from the flowers of *Syringa oblata* and *S. oblata* var. *alba* to regulate D-limonene emissions during the flowering phase (Yan et al., 2019). Some cytochrome P450 genes were reported to be involved in metabolism of terpenoid volatiles, such as homologous with *V. vinifera* cytochrome P45076A2 and CYP76A2 in *Syringa oblata* (Zheng et al., 2015). In our study, the *S. oblata* TPS-b subfamily had the most members and its CYP76 family genes were amplified by tandem repeats forming gene clusters. Additional studies have shown that lilac aldehyde and lilac alcohol can enhance the attractiveness of plants to pollinating insects (Plepyš et al., 2002; Dötterl et al., 2007; Jürgens et al., 2014), so the amplification of TPS-b subfamily and CYP76 family genes of *S. oblata* may improve competitiveness to attract insect pollination.

In summary, we published a chromosome-scale *S. oblata* genome with a longer contig and it was a better reference for transcriptome assembly. Genome-wide annotation and transcriptomic and metabolic data integration improved our understanding of the molecular mechanism underlying the formation of floral color differences among *S. oblata* individuals. In addition, the genome-wide identification and analysis of TPS and CYP450 genes in volatiles of *S. oblata* provided a basis for the elucidation of the mechanism of volatile compounds biosynthesis mechanism and further research. These results are important references for the resolution of molecular mechanism of horticultural traits and molecular breeding in *S. oblata*.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: CNGDB (<https://db.cngb.org/>), CNP0003698.

Author contributions

YZ, MH, BX, and LC conceived and designed the study. ZL and YY prepared the materials. BX, ZL, YY, WG, QM, and XL performed data analyses. LC, BX, YB, YS, QL, and LY wrote the manuscript. All authors contributed to the article and approved the submitted version.

References

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25 (1), 25–29. doi: 10.1038/75556
- Aubourg, S., Lecharny, A., and Bohlmann, J. (2002). Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Mol. Genet. Genomics* 267 (6), 730–745. doi: 10.1007/s00438-002-0709-y
- Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* 27 (2), 573–580. doi: 10.1093/nar/27.2.573
- Boachon, B., Junker, R. R., Miesch, L., Bassard, J.-E., Höfer, R., Caillieudeaux, R., et al. (2015). CYP76C1 (Cytochrome P450)-mediated linalool metabolism and the formation of volatile and soluble linalool oxides in *Arabidopsis* flowers: A strategy for defense against floral antagonists. *Plant Cell* 27 (10), 2972–2990. doi: 10.1105/tpc.15.00399
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* 31 (1), 365–370. doi: 10.1093/nar/gkg095
- Bombarely, A., Moser, M., Amrad, A., Bao, M., Bapaume, L., Barry, C. S., et al. (2016). Insight into the evolution of the solanaceae from the parental genomes of *petunia hybrida*. *Nat. Plants* 2 (6), 16074. doi: 10.1038/nplants.2016.74
- Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., and Shendure, J. (2013). Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31 (12), 1119–1125. doi: 10.1038/nbt.2727
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and applications. *BMC Bioinf.* 10 (1), 421. doi: 10.1186/1471-2105-10-421
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I. A. P., Tang, H., Wang, X., et al. (2014). Early allopolyploid evolution in the post-neolithic *brassica napus* oilseed genome. *Science*. 345 (6199), 950–953. doi: 10.1126/science.1253435
- Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinf.* 5 (1), 4.10.11–14.10.14. doi: 10.1002/0471250953.bi0410s05
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13 (8), 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Chen, F., Tholl, D., Bohlmann, J., and 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, 1, 212–229. doi: 10.1111/j.1365-3113X.2011.04520.x
- Cui, L., Hu, M., Cao, P., Niu, Y., Li, C., Liu, Z., et al. (2019). Chemical constituents and coagulation activity of *Syringia oblata* lindl flowers. *BMC Chem.* 13 (1), 108. doi: 10.1186/s13065-019-0621-8
- De Bie, T., Cristianini, N., Demuth, J. P., and Hahn, M. W. (2006). CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* 22 (10), 1269–1271. doi: 10.1093/bioinformatics/btl097
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36 (suppl_2), W465–W469. doi: 10.1093/nar/gkn180
- Dötterl, S., Burkhardt, D., Jürgens, A., and Mosandl, A. (2007). Stereoisomeric pattern of lilac aldehyde in *Silene latifolia*, a plant involved in a nursery pollination system. *Phytochemistry* 68 (4), 499–504. doi: 10.1016/j.phytochem.2006.11.013

Funding

This study was supported by Scientific research start-up funds of Jilin Agricultural University (No. 202023298).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1078677/full#supplementary-material>

- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32 (5), 1792–1797. doi: 10.1093/nar/gkh340
- Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2013). Pfam: the protein families database. *Nucleic Acids Res.* 42 (D1), D222–D230. doi: 10.1093/nar/gkt1223
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* 117 (17), 9451–9457. doi: 10.1073/pnas.1921046117
- Gbenro, S., Hippe, K., and Cao, R. (2020). “HMMeta: Protein function prediction using hidden Markov models,” in *Proceedings of the 11th ACM international conference on bioinformatics, computational biology and health informatics* (Virtual Event, USA: Association for Computing Machinery).
- Gegen, Z.-I., Gao, Y., Site, G. -I., Alatan, C. -M., Tai, B. -L., and Tu, Y. (2022). Mechanism of *Syringa oblata* in treating angina pectoris based on GC-MS and network pharmacology. *China J. Chin. Materia Med.* 47 (3), 836–845. doi: 10.19540/j.cnki.cjcm.20210927.201
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29 (7), 644–652. doi: 10.1038/nbt.1883
- Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. R. (2003). Rfam: an RNA family database. *Nucleic Acids Res.* 31 (1), 439–441. doi: 10.1093/nar/gkg006
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. *Genome Biol.* 9 (1), R7. doi: 10.1186/gb-2008-9-1-r7
- Holton, T. A., and Cornish, E. C. (1995). Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* 7 (7), 1071–1083. doi: 10.1105/tpc.7.7.1071
- Hoshino, A., Morita, Y., Choi, J.-D., Saito, N., Toki, K., Tanaka, Y., et al. (2003). Spontaneous mutations of the flavonoid 3'-hydroxylase gene conferring reddish flowers in the three morning glory species. *Plant Cell Physiol.* 44 (10), 990–1001. doi: 10.1093/pcp/pcg143
- Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449 (7161), 463–467. doi: 10.1038/nature06148
- Jia, Q., Brown, R., Köllner, T. G., Fu, J., Chen, X., Wong, G. K.-S., et al. (2022). Origin and early evolution of the plant terpene synthase family. *Proceedings of the National Academy of Sciences.* 119, 15. doi: 10.1073/pnas.2100361119
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30 (9), 1236–1240. doi: 10.1093/bioinformatics/btu031
- Jürgens, A., Glück, U., Aas, G., and Dötterl, S. (2014). Diel fragrance pattern correlates with olfactory preferences of diurnal and nocturnal flower visitors in *Salix caprea* (Salicaceae). *Botanical J. Linn. Soc.* 175 (4), 624–640. doi: 10.1111/boj.12183
- Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., Walichiewicz, J. J. C., et al. (2005). Repbase update, a database of eukaryotic repetitive elements. *Cytogenetic and genome research.* 110, 1–4, 462–467. doi: 10.1159/000084979
- Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28 (1), 27–30. doi: 10.1093/nar/28.1.27
- Kersey, P. J. (2019). Plant genome sequences: past, present, future. *Curr. Opin. Plant Biol.* 48, 1–8. doi: 10.1016/j.pbi.2018.11.001
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37 (8), 907–915. doi: 10.1038/s41587-019-0201-4
- Kreck, M., Püschel, S., Wüst, M., and Mosandl, A. (2003). Biogenetic studies in *Syringa vulgaris* L.: synthesis and bioconversion of deuterium-labeled precursors into lilac aldehydes and lilac alcohols. *J. Agric. Food Chem.* 51 (2), 463–469. doi: 10.1021/jf020845p
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 326(5950), 289–293. doi: 10.1126/science.1181369
- Li, Z.-G., Lee, M.-R., and Shen, D.-L. (2006). Analysis of volatile compounds emitted from fresh *Syringa oblata* flowers in different florescence by headspace solid-phase microextraction-gas chromatography-mass spectrometry. *Analytica Chimica Acta* 576 (1), 43–49. doi: 10.1016/j.aca.2006.01.074
- Li, J. L., Luo, X. D., Zhao, P.-J., and Zeng, Y. (2009). Post-modification enzymes involved in the biosynthesis of plant terpenoids. *Acta Botanica Yunnanica.* 31, 05, 461–468. doi: 10.3724/sp.j.1143.2009.09108
- Li, L., Stoeckert, C. J., and Roos, D. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome research.* 13, 9, 2178–2189. doi: 10.1101/gr.1224503
- Liu, Y.-Y., Chen, X.-R., Wang, J.-P., Cui, W.-Q., Xing, X.-X., Chen, X.-Y., et al. (2019). Transcriptomic analysis reveals flavonoid biosynthesis of *Syringa oblata* Lindl. in response to different light intensity. *BMC Plant Biol.* 19 (1), 487. doi: 10.1186/s12870-019-2100-8
- Li, M., Zhang, D., Gao, Q., Luo, Y., Zhang, H., Ma, B., et al. (2019). Genome structure and evolution of. *Nat. Plants* 5 (2), 174–183. doi: 10.1038/s41477-018-0349-9
- Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25 (5), 955–964. doi: 10.1093/nar/25.5.955
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 1 (1), 2047–217X. doi: 10.1186/2047-217x-1-18
- Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., et al. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* 50 (W1), W276–W279. doi: 10.1093/nar/gkac240
- Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20 (16), 2878–2879. doi: 10.1093/bioinformatics/bth315
- Martin, J. A., and Wang, Z. (2011). Next-generation transcriptome assembly. *Nat. Rev. Genet.* 12 (10), 671–682. doi: 10.1038/nrg3068
- Ma, B., Wu, J., Shi, T.-L., Yang, Y.-Y., Wang, W.-B., Zheng, Y., et al. (2022). Lilac (*Syringa oblata*) genome provides insights into its evolution and molecular mechanism of petal color change. *Commun. Biol.* 5 (1), 686. doi: 10.1038/s42003-022-03646-9
- Md, V., Misra, S., Li, H., and Aluru, S. (2019). Efficient architecture-aware acceleration of bwa-mem for multicore systems. *arXiv:1907.12931*. doi: 10.1109/IPDPS.2019.00041
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35 (suppl_2), W182–W185. doi: 10.1093/nar/gkm321
- Muhlemann, J. K., Klempien, A., and Dudareva, N. (2014). Floral volatiles: biosynthesis to Funct.. *Plant, cell & environment.* 37 (8), 1936–1949. doi: 10.1111/pce.12314
- Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29 (22), 2933–2935. doi: 10.1093/bioinformatics/btt509
- Paquette, S. M., Bak, S., and Feyereisen, R. (2000). Intron-exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of arabisidopsis thaliana. *DNA Cell Biol.* 19 (5), 307–317. doi: 10.1089/10445490050021221
- Plepeys, D., Ibarra, F., and Löfstedt, C. (2002). Volatiles from flowers of *Platanthera bifolia* (Orchidaceae) attractive to the silver y moth, autographa gamma (Lepidoptera: Noctuidae). *Oikos.* 99, 1, 69–74. doi: 10.1034/j.1600-0706.2002.990107.x
- Price, A. L., Jones, N. C., and Pevzner, P. A. (2005). *De novo* identification of repeat families in large genomes. *Bioinformatics* 21 (suppl_1), i351–i358. doi: 10.1093/bioinformatics/btl1018
- Qi, Y., Lou, Q., Quan, Y., Liu, Y., and Wang, Y. (2013). Flower-specific expression of the phalaenopsis flavonoid 3', 5'-hydroxylase modifies flower color pigmentation in petunia and lily. *Plant Cell Tissue Organ Culture (PCTOC)* 115 (2), 263–273. doi: 10.1007/s11240-013-0359-2
- Raguso, R. A. (2016). More lessons from linalool: Insights gained from a ubiquitous floral volatile. *Curr. Opin. Plant Biol.* 32, 31–36. doi: 10.1016/j.pbi.2016.05.007
- Sato, S., Tabata, S., Hirakawa, H., Asamizu, E., Shirasawa, K., Isobe, S., et al. (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485 (7400), 635–641. doi: 10.1038/nature11119
- Sémon, M., and Wolfe, K. H. (2007). Consequences of genome duplication. *Curr. Opin. Genet. Dev.* 17 (6), 505–512. doi: 10.1016/j.gde.2007.09.007
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31 (19), 3210–3212. doi: 10.1093/bioinformatics/btv351
- Sollars, E. S. A., Harper, A. L., Kelly, L. J., Sambles, C. M., Ramirez-Gonzalez, R. H., Swarbreck, D., et al. (2017). Genome sequence and genetic diversity of European ash trees. *Nature* 541 (7636), 212–216. doi: 10.1038/nature20786
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30 (9), 1312–1313. doi: 10.1093/bioinformatics/btu033

- Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33 (suppl_2), W465–W467. doi: 10.1093/nar/gki458
- Sun, P., Jiao, B., Yang, Y., Shan, L., Li, T., Li, X., et al. (2021). WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *bioRxiv* 2021, 2004.2029.441969. doi: 10.1101/2021.04.29.441969
- Sun, Y., Shang, L., Zhu, Q.-H., Fan, L., and Guo, L. (2022). Twenty years of plant genome sequencing: achievements and challenges. *Trends Plant Sci.* 27 (4), 391–401. doi: 10.1016/j.tplants.2021.10.006
- Tai, B., Bai, L., Ji, R., Yu, M., NAla, Huang, L., et al. (2022). Phytochemical and pharmacological progress on syringa oblata, a traditional Mongolian medicine. *Chin. Herbal Medicines* 14 (3), 392–402. doi: 10.1016/j.chmed.2022.04.002
- Tanaka, M., Sakai, T., and Takahata, Y. (2019). Allele dosage-dependent selection of recessive F³H allele homozygote altered anthocyanin composition in sweetpotato. *Mol. Breed.* 39 (10), 152. doi: 10.1007/s11032-019-1062-8
- The Arabidopsis Genome, I. (2000). Analysis of the genome sequence of the flowering plant arabidopsis thaliana. *Nature* 408 (6814), 796–815. doi: 10.1038/35048692
- Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., et al. (2006). The genome of black cottonwood, populus trichocarpa (Torr. Gray). 313 (5793), 1596–1604. doi: 10.1126/science.1128691
- Unver, T., Wu, Z., Sterck, L., Turktaş, M., Lohaus, R., Li, Z., et al. (2017). Genome of wild olive and the evolution of oil biosynthesis. *Proceedings of the National Academy of Sciences*. 114, 44, E9413–E9422. doi: 10.1073/pnas.1708621114
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., et al. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33 (14), 2202–2204. doi: 10.1093/bioinformatics/btx153
- Wang, Y., Lu, L., Li, J., Li, H., You, Y., Zang, S., et al. (2022). A chromosome-level genome of syringa oblata provides new insights into chromosome formation in oleaceae and evolutionary history of lilacs. *The Plant Journal*. 111, 3, 836–848. doi: 10.1111/tpj.15858
- Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40 (7), e49–e49. doi: 10.1093/nar/gkr1293
- Wang, L., Yu, S., Tong, C., Zhao, Y., Liu, Y., Song, C., et al. (2014). Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. *Genome Biol.* 15 (2), R39. doi: 10.1186/gb-2014-15-2-r39
- Wu, J., Long, J., Liu, H., Sun, G., Li, J., Xu, L., et al. (2021). Biogenic volatile organic compounds from 14 landscape woody species: Tree species selection in the construction of urban greenspace with forest healthcare effects. *J. Environ. Manage.* 300, 113761. doi: 10.1016/j.jenvman.2021.113761
- Xu, S., Ding, Y., Sun, J., Zhang, Z., Wu, Z., Yang, T., et al. (2022). A high-quality genome assembly of jasminum sambac provides insight into floral trait formation and oleaceae genome evolution. *Molecular Ecology Resources*. 22, 2, 724–739. doi: 10.1111/1755-0998.13497
- Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., et al. (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475 (7355), 189–195. doi: 10.1038/nature10158
- Xu, Z., and Wang, H. (2007). LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35 (suppl_2), W265–W268. doi: 10.1093/nar/gkm286
- Yang, Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. *Computer applications in the biosciences*. 13, 5, 555–556. doi: 10.1093/bioinformatics/13.5.555
- Yang, J., Ren, Y., Zhang, D., Chen, X., Huang, J., Xu, Y., et al. (2021). Transcriptome-based WGCNA analysis reveals regulated metabolite fluxes between floral color and scent in narcissus tazetta flower. *Int. J. Mol. Sci.* 22 (15), 8249. doi: 10.3390/ijms22158249
- Yan, Z., Qin, Y., Zheng, J., Leng, P., and Hu, Z. (2019). Gene cloning and expression analysis of limonene synthase in syringa oblata and s. oblata var. alba. *J. Forestry Res.* 30 (4), 1301–1309. doi: 10.1007/s11676-018-0697-3
- Yao-zu, C., Zhao-lin, L., and Hai-quan, L. (1987). Analysis of fragrance volatiles of fresh flowers by pre-column absorption and GC/MS. *Chromatographia* 23 (7), 502–506. doi: 10.1007/BF02309419
- Yu, J., Hu, S., Wang, J., Wong, G. K.-S., Li, S., Liu, B., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* l. ssp. *indica*). 296 (5565), 79–92. doi: 10.1126/science.1068037
- Yu, J., Tu, L., Subburaj, S., Bae, S., and Lee, G.-J. (2021). Simultaneous targeting of duplicated genes in petunia protoplasts for flower color modification via CRISPR-Cas9 ribonucleoproteins. *Plant Cell Rep.* 40 (6), 1037–1045. doi: 10.1007/s00299-020-02593-1
- Zhang, J. (2011). “Analysis of anthocyanin and its flower color formation mechanism in four ornamental plants,” in *doctor's thesis*. [XiAn]: Northwest Agricultural and Forestry University.
- Zhang, Y., Butelli, E., and Martin, C. (2014). Engineering anthocyanin biosynthesis in plants. *Curr. Opin. Plant Biol.* 19, 81–90. doi: 10.1016/j.pbi.2014.05.011
- Zhao, Z., Gao, A., Luo, R., Liu, K., and Yu, K. (2022). The different deletion mutation in the phenylalanine ammonia-lyase (PAL) gene affects the peel color of mango (*Mangifera indica* l.). *Genet. Resour. Crop Evol.* 69 (7), 2301–2306. doi: 10.1007/s10722-022-01393-8
- Zheng, J., Hu, Z., Guan, X., Dou, D., Bai, G., Wang, Y., et al. (2015). Transcriptome analysis of syringa oblata lindl. inflorescence identifies genes associated with pigment biosynthesis and scent metabolism. *PLoS One* 10 (11), e0142542. doi: 10.1371/journal.pone.0142542