1 **The chromosome-scale genome assembly of *Jasminum*** 2 ***sambac* var. *unifoliatum* provide insights into the formation** 3 **of floral fragrance**

4 Chengzhe Zhou^{1,2,3}, Chen Zhu^{1,2,3}, Caiyun Tian^{1,3}, Siyi Xie^{1,3}, Kai Xu^{1,3}, Linjie
5 Huang^{1,3}, Shengjing Wen^{1,3}, Cheng Zhang^{1,3}, Zhongxiong Lai^{1,2}, Yuling Lin^{1,2,*},
6 Yuqiong Guo^{1,3,*}

7 *Corresponding author

8 ¹College of Horticulture, Fujian Agriculture and Forestry University, Fuzhou, 350002,
9 China

10 ²Institute of Horticultural Biotechnology, Fujian Agriculture and Forestry University,
11 Fuzhou, 350002, China

12 ³Tea Industry Research Institute, Fujian Agriculture and Forestry University, Fuzhou,
13 350002, China

14

15 **Abstract**

16 The high-quality genome assembly of the JSU-FSP provides valuable information on
17 how jasmine produces the abundant VTs and VPBs that synergistically contribute to
18 the pleasant aroma of jasmine flower.

19

20 **Introduction**

21 Jasmine (*Jasminum sambac*) is a tropical or subtropical plant that is notable for the
22 sweet, heady fragrance of its flowers. The high accumulation of volatile terpenes (VTs)
23 and volatile phenylpropanoid/benzenoids (VPBs) in jasmine flowers creates a natural
24 source of essential oil to meet consumer demand all over the world. Additionally,
25 jasmine is also used to produce scented tea, which is reprocessed from finished tea by
26 absorbing aroma volatiles of the fresh flowers and has been favoured since antiquity
27 [1]. Cultivar jasmine varieties mainly belong to three major groups: *J. sambac* var.
28 *unifoliatum* (JSU, single-petal type), *J. sambac* var. *bifoliatum* (JSB, double-petal
29 type), and *J. sambac* var. *trifoliatum* (JST, multi-petal type). The flowers of JSU, with
30 the best aroma quality, are considered as the optimal materials for the production of

high-grade jasmine tea [2]. However, the lack of a reference genome limits the studies of basic and applied biology on the jasmine. The genetic bases for the rich production of VTs and VPBs remain largely unclear. While recently the reference genomes of JSB [3] and JST [4] cultivars were *de novo* sequenced and assembled, no JSU genome has yet been reported.

Results and discussions

Here, we assembled a reference genome for the autotriploid JSU cultivar ‘Fuzhou single-petal’ (JSU-FSP). According to the 21-mer analysis of 45.14 Gb Illumina reads and flow cytometry, we estimated that the haploid genome size of JSU-FSP is ~471.23 Mb and ~498.35 Mb, respectively. We initially assembled the genome by hifiasm v0.16 with 85.89 Gb PacBio HiFi reads, resulting in a 495.60 Mb sequence assembly with a contig N50 of 16.88 Mb. The BUSCO and CEGMA analyses showed that the completeness of the JSU-FSP genome was 97.83% and 98.91%, respectively, which indicated that the excellent continuity and completeness of the assembly. Then, all the contigs were corrected and orientated by Hi-C onto 13 pseudo-chromosomes, generating chromosome-scale sequences of 476.92 Mb (Fig. 1A and 1B).

The annotation results showed that there were 264.98 Mb repetitive sequences in JSU-FSP genome, of which long terminal repeat-retrotransposons (LTR-RTs) accounted for 66.33% (175.77 Mb). The Ty3-gypsy were the predominant LTR-RTs (58.92 Mb), followed by Ty1-copia (36.68 Mb). We identified 30,989 protein-coding genes through integrated *ab initio* modeling, homology-based searches, and transcriptome analysis (Fig. 1A). The BUSCO analysis showed that 96.16% of the core plant genes were present in identified genes, indicating an excellent completeness of genes prediction.

Through comparative genomics analysis between the JSU-FSP genome and the genomes of other 11 representative plant species, we observed that JSU-FSP genome went through an ancient WGD event by estimating *Ks* distributions (Fig. 1C). The phylogenetic tree based on 448 single-copy orthologous genes illustrated that *J. sambac* diverged from the ancestors of other four species of Oleaceae plants about

54~58 million years ago (Fig. 1D). Gene family analysis revealed that 26,638 genes in JSU-FSP were clustered into 17,446 families (Fig. 1E). The five Oleaceae plants shared 11,333 common gene families, while 1,221 gene families were unique in the JSU-FSP genome (Fig. 1F). The unique genes of JSU-FSP were clustered into “Terpenoid backbone biosynthesis” and “Sesquiterpenoid and triterpenoid biosynthesis” of the KEGG enrichment pathways, endowing jasmine with the genetic basis of heady fragrance (Fig. 1G).

As aroma volatiles of jasmine are synthesized and emitted in the opening process [5], flower buds or flowers at five different stages were selected for transcriptome sequencing (Fig. 1H). The expression of key genes was investigated to unveil the mechanism of VTs and VPBs biosynthesis. We identified that among the 73 structural genes involved in VTs and VPBs metabolic pathways, the expression of 43 genes in opening flowers was significantly higher than that in buds. This indicated the differential expression of these related genes in opening flowers was responsible for the aroma formation. In addition to *de novo* synthesis, plant volatiles are present in glycosidically bound forms and can be hydrolyzed to free volatiles by GH1, GH3, and GH5 families of β -glucosidase (BGLU) [6]. Phylogenetic analysis revealed that a total of 81 BGLU genes in JSU-FSP genome were clustered into GH1 (47), GH3 (20), and GH5 (14) subfamilies (Fig. 1J). Among which, 11 BGLU genes were highly expressed in opening flowers, implying that jasmine aroma is also affected by the expression of BGLU genes [7].

Acknowledgements

This work was supported by the Construction of Plateau Discipline of Fujian Province (102/71201801101) and the Construction Project for Technological Innovation and Service System of Tea Industry Chain of Fujian Agriculture and Forestry University (K1520005A01). We thank Xin Huang from the School of Foreign Languages of Fuzhou University for her linguistic assistance; Ying Yu from Biomarker Technologies for assistance with sequencing; Prof. Yong Li from Henan Agricultural University for providing the genomic data of *Forsythia suspensa*; and Tianlong Fu

99

100 **Figure 1.** Landscape of the genome of *J. sambac* var. *unifoliatum* cv. Fuzhou
 101 single-petal (JSU-FSP). A, statistics for the assembly and annotation of the JSU-FSP
 102 genome. B, the circulars map shows chromosome ideograms (a), transposable element
 103 density (b), gene density (c), GC content (d), and syntenic blocks (colored lines in the
 104 center). C, density distribution of *Ks* within *J. sambac*, *O. europaea*, and *V. vinifera*. D,
 105 phylogenetic tree with estimated divergence time and gene family
 106 expansion/contraction. E, statistics for the gene family. F, venn diagram of
 107 orthogroups within five Oleaceae plants. G, KEGG enrichment analysis for the unique
 108 paralogs in JSU-FSP genome. H, images of five flowering stages of JSU-FSP, bar = 1
 109 cm. I, differentially expressed genes involved in terpenes and
 110 phenylpropanoid/benzenoids metabolic pathways. J, phylogenetic tree of
 111 β -glucosidase (BGLU) proteins from JSU-FSP, *A. thaliana*, and *O. sativa*. K, the
 112 expression profile of BGLU genes in different flowering stages of JSU-FSP.

113

114 References

- 115 1. Zhou HC, Hou ZW, Wang DX et al. Large scale preparation, stress analysis, and storage of
 116 headspace volatile condensates from *Jasminum sambac* flowers. Food Chemistry. 2019; **286**:
 117 170-178.
- 118 2. Deng YM, Sun XB, Gu CS et al. Identification of pre-fertilization reproductive barriers and
 119 the underlying cytological mechanism in crosses among three petal-types of *Jasminum*
 120 *sambac* and their relevance to phylogenetic relationships. PLOS ONE. 2017; **12**: e0176026.
- 121 3. Chen G, Mostafa S, Lu Z et al. The jasmine (*Jasminum sambac*) genome and flower
 122 fragrances. bioRxiv. 2021.
- 123 4. Xu SX, Ding YL, Sun JT et al. A high-quality genome assembly of *Jasminum sambac*
 124 provides insight into floral trait formation and Oleaceae genome evolution. Molecular
 125 Ecology Resources. 2021; **00**: 1-16.
- 126 5. Zhou CZ, Zhu C, Tian CY et al. Integrated volatile metabolome, multi-flux full-length
 127 sequencing, and transcriptome analyses provide insights into the aroma formation of
 128 postharvest jasmine (*Jasminum sambac*) during flowering. Postharvest Biology and
 129 Technology. 2022; **183**: 111726.
- 130 6. Zhou Y, Zeng LT, Gui JD et al. Functional characterizations of β -glucosidases involved in
 131 aroma compound formation in tea (*Camellia sinensis*). Food Research International. 2017; **96**:
 132 206-214.
- 133 7. Moon J, Watanabe N, Sakata K et al. Linalyl β -d-glucopyranoside and its 6'-O-malonate as
 134 aroma precursors from *Jasminum sambac*. Phytochemistry. 1994; **36**: 1435-1437.

135