



OPEN ACCESS

EDITED BY

Yusuf Khan,
Inland Norway University of Applied
Sciences, Norway

REVIEWED BY

Mingzhuo Li,
North Carolina State University,
United States
Gagandeep Singh,
The University of Queensland, Australia

*CORRESPONDENCE

Fen Wang
fenmin521@163.com
Baoxing Song
baoxing.song@pku-iaas.edu.cn
Litang Lu
ltrlv@gzu.edu.cn

†These authors have contributed equally to
this work

SPECIALTY SECTION

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

RECEIVED 27 July 2022

ACCEPTED 25 August 2022

PUBLISHED 23 September 2022

CITATION

Wang F, Zhang B, Wen D, Liu R, Yao X,
Chen Z, Mu R, Pei H, Liu M, Song B and
Lu L (2022) Chromosome-scale genome
assembly of *Camellia sinensis* combined
with multi-omics provides insights into its
responses to infestation with green
leafhoppers.
Front. Plant Sci. 13:1004387.
doi: 10.3389/fpls.2022.1004387

COPYRIGHT

© 2022 Wang, Zhang, Wen, Liu, Yao, Chen,
Mu, Pei, Liu, Song and Lu. 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.

Chromosome-scale genome assembly of *Camellia sinensis* combined with multi-omics provides insights into its responses to infestation with green leafhoppers

Fen Wang^{1,2*†}, Baohui Zhang^{2,3†}, Di Wen^{1†}, Rong Liu¹,
Xinzhan Yao^{2,4}, Zhi Chen¹, Ren Mu¹, Huimin Pei¹, Min Liu⁵,
Baoxing Song^{1,6*} and Litang Lu^{2,4*}

¹The Department of Life Science and Agriculture, Qiannan Normal College for Nationalities, Duyun, China, ²The Key Laboratory of Plant Resources Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), Guiyang, China, ³Horticulture Institute (Guizhou Horticultural Engineering Technology Research Center), Guizhou Academy of Agricultural Sciences, Guiyang, China, ⁴College of Tea Science, Guizhou University, Guiyang, China, ⁵Biomarker Technologies Corporation, Beijing, China, ⁶Peking University Institute of Advanced Agricultural Sciences, Weifang, China

The tea plant (*Camellia sinensis*) is an important economic crop, which is becoming increasingly popular worldwide, and is now planted in more than 50 countries. Tea green leafhopper is one of the major pests in tea plantations, which can significantly reduce the yield and quality of tea during the growth of plant. In this study, we report a genome assembly for DuyunMaojian tea plants using a combination of Oxford Nanopore Technology PromethION™ with high-throughput chromosome conformation capture technology and used multi-omics to study how the tea plant responds to infestation with tea green leafhoppers. The final genome was 3.08Gb. A total of 2.97Gb of the genome was mapped to 15 pseudo-chromosomes, and 2.79Gb of them could confirm the order and direction. The contig N50, scaffold N50 and GC content were 723.7kb, 207.72Mb and 38.54%, respectively. There were 2.67Gb (86.77%) repetitive sequences, 34,896 protein-coding genes, 104 miRNAs, 261 rRNA, 669 tRNA, and 6,502 pseudogenes. A comparative genomics analysis showed that DuyunMaojian was the most closely related to Shuchazao and Yunkang 10, followed by DASZ and tea-oil tree. The multi-omics results indicated that phenylpropanoid biosynthesis, α -linolenic acid metabolism, flavonoid biosynthesis and 50 differentially expressed genes, particularly peroxidase, played important roles in response to infestation with tea green leafhoppers (*Empoasca vitis* Göthe). This study on the tea tree is highly significant for its role in illustrating the evolution of its genome and discovering how the tea plant responds to infestation with tea green leafhoppers will contribute to a theoretical foundation to breed tea plants resistant to insects that will ultimately result in an increase in the yield and quality of tea.

KEYWORDS

DuyunMaojian, genome, chromatin interaction mapping, multi-omics, tea green leafhoppers, resistant to insects

Introduction

The tea tree, which has been grown in China for approximately 2,000 years (Kingdom-Ward, 1950), is a very important beverage crop with substantial economic value (Fang et al., 2022). The tea has many functions and effects, such as engendering liquid and allaying thirst, improving eyesight, preventing cancer and other healthy values (Ross and Kasum, 2002; Khan et al., 2008; Tounekti et al., 2013). China and India are the primary producers of tea (Bhattacharyya et al., 2022; Na Nagara et al., 2022; Nyhus Dhillon et al., 2022), and the tea industry in Sri Lanka is the second largest source of foreign exchange (Munasinghe et al., 2017). Tea is becoming increasingly popular worldwide and is now planted in more than 50 countries (Xia et al., 2020b; Yue et al., 2022) because of intriguing flavors (Han et al., 2016; Tan et al., 2019) and health benefits (Zhang et al., 2021b; Ahammed and Li, 2022). DuyunMaojian tea in Guizhou Province is a famous variety that won the Panama World's Fair Award in 1915 and the cultivation is thought to date back to 1368–1644 AD. The shape of DuyunMaojian tea is like a fishhook and is full of pekoe, which produces high-quality taste. Tea plantations in Guizhou Province occupy 470,000 ha, and ranked first in China by 2022. Duyun Maojian tea plants usually grow in a severe environment at 30° north latitude and an elevation of 800–1,200 m with little sunshine and much cloud and fog. It has a soft aroma, fresh and thick smell, and a sweet aftertaste. Thus, DuyunMaojian tea plant is a good material for breeding. However, tea green leafhoppers (*Empoasca vitis* Göthe) can significantly reduce the yield and quality of tea and cause losses in yield of 11–55% in the absence of effective means of prevention and treatment (Hazarika et al., 2009; Yang et al., 2013). The infestation of tea plants by these leafhoppers can cause physiological and biochemical defense responses, which primarily include the activation of salicylic acid, jasmonic acid, and other signaling pathways (Arimura et al., 2011; Zeng et al., 2019). Thus, we used the DuyunMaojian tea plant genome in combination with transcriptomic, proteomic and metabolomic studies to analyze the differentially expressed genes (DEGs), differentially expressed proteins (DEPs), and differential metabolites that play dominant effects in the response to tea green leafhoppers. We explored the important metabolic pathways in the DEGs, DEPs and differential metabolites that are involved in the molecular mechanism of tea plant responses to infestation with green leafhoppers, which provides a theoretical reference to develop insect-resistant cultivars.

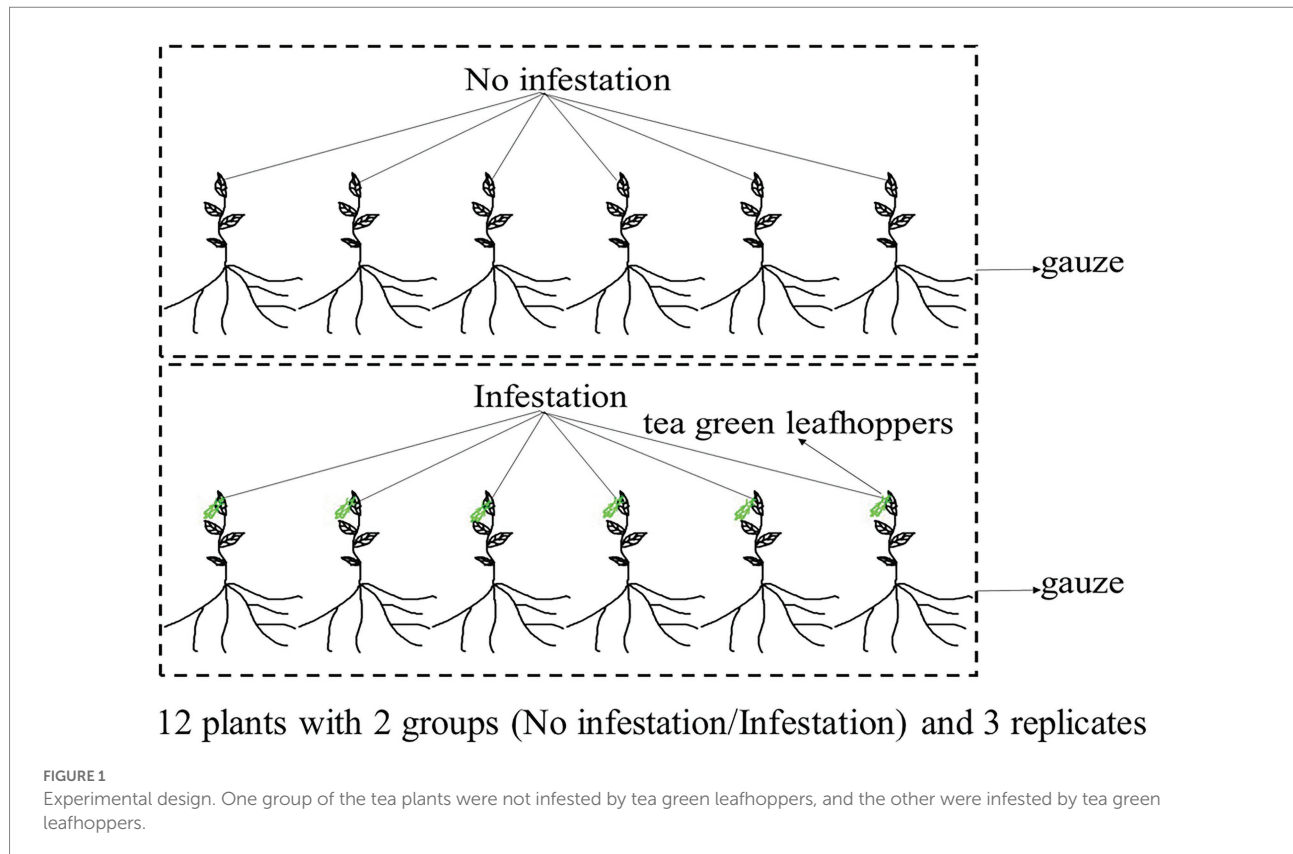
Camellia sinensis genomes have been previously published from the cultivars Yukang 10 (Xia et al., 2017), Shuchazao (Wei et al., 2018; Chen et al., 2020; Xia et al., 2020a), Biyun (Zhang et al., 2020a), DASZ (Zhang et al., 2020b), Huangdan (Wang et al., 2021), Longjing 43 (Wang et al., 2020), and Tieguanyin (Zhang et al., 2021a). Two of the genome versions (Xia et al., 2017; Wei et al., 2018) are short-read-based assemblies that are highly fragmented, and the chromosome-scale genome assemblies of Shuchazao, DASZ, Longjing 43 were obtained using PacBio SMRT (Xia et al., 2020a; Wang et al., 2020) and high-throughput

chromosome conformation capture (Hi-C) technology. Huangdan and Tieguanyin were assembled using PacBio HiFi (Wang et al., 2021; Zhang et al., 2021a) and Hi-C technology. In this study, we used Oxford Nanopore Technology (ONT) PromethION™ (Nicholls et al., 2019) combined with Hi-C technology (Feng et al., 2022; Lin et al., 2022) to assemble the genome of the elite tea cultivar DuyunMaojian. ONT can generate multi-kilobase-long RNA reads and has no bias for GC content and length. Thus, this study will present a theoretical reference for programs to breed tea tree.

Materials and methods

Sample collection, DNA preparation and library construction

An individual plant of DuyunMaojian tea was collected in December 2020 from Duyun City, Guizhou Province, China (N26°20'19" E107°31'11"). The altitude is 786.1 m. Fresh and healthy leaves were frozen in liquid nitrogen after collection, and stored at –80°C before DNA extraction. DNA was extracted from the leaves using the CTAB method, which were examined by NanoDrop spectrophotometry (Thermo Fisher Scientific, Waltham, MA, United States), Qubit fluorimeter (Invitrogen, Carlsbad, CA, United States) and electrophoresis on a 0.35% agarose gel, and large segments were filtered using the BluePippin™ System (Sage Science, Inc., Beverly, MA, United States). We then prepared a library using the large segments of DNA, an ONT Template prep kit (SQK-LSK109; Oxford Nanopore Technologies, Oxford, United Kingdom) and an NEB Next FFPE DNA Repair Mix kit (New England Biolabs, Ipswich, MA, United States). The leaves for transcriptomic, proteomic, and metabolomics analyses were collected from plants same with the genome. Those trees were raised from cuttings planted in Duyun City, Guizhou Province, China (N26°20'19" E107°31'11") in November 2018. The leaves were collected from 12 individual trees with similar heights in June 2021. We made sure that 12 tea plants grown in the same environment and were not infected with pathogens. Six trees were infested by tea green leafhoppers for 24 h and another 6 trees were used as control. The infested and control group were placed in two gauze cages separately (Figure 1). There are three replicates in each group. An Illumina NovaSeq platform was used to sequence the transcriptome of tea tree infested or not by tea green leafhoppers. A DP441 QIAGEN RNAPrep Pure Plant Plus Kit was used to extract all of the RNAs. NEBNext Ultra™ RNA Library Prep Kit for Illumina (New England Biolabs, United States) was used to generate sequencing libraries, and we purified the mRNA. We then fragmented it in NEBNext First Strand Synthesis Reaction Buffer. First strand and second strand cDNA were synthesized using random hexamer primers, M-MuLV Reverse Transcriptase, and DNA Polymerase I and RNase H, respectively. Finally, the library preparations and pair-end reads (PE150) were produced. Tandem



mass tag (TMT) and liquid chromatography–tandem mass spectrometry (LC–MS/MS) were used to quantitatively analyze proteome. The tea tree leaves were frozen in liquid nitrogen and dissolved in lysis buffer. The quality of extracted proteins was detected by SDS–PAGE. We added 10 mM dithiothreitol to 100 μ g of proteins, which was incubated for 1 h at 37°C and followed by alkylation with 40 mM iodoacetamide for 45 min in the dark at room temperature. Trypsin (150 enzyme to protein) was added to the samples, which were then incubated overnight at 37°C. The reaction was terminated by the addition of 0.1% formic acid (FA). Next, samples were desalted using a C_{18} chromatographic column, washed with 70% acetonitrile (ACN), and vacuum freeze-dried according to the manufacturer’s instructions for the TMT kit. The LC–MS/MS spectra were then determined using XBridge Peptide BEH 5 μ m C_{18} . The separated peptides were parsed using an Orbitrap Exploris™ 480 mass spectrometer. Ultra-performance liquid chromatography–quadrupole time-of-flight mass spectrometry (UPLC–QTOF–MS) was used to analyze the metabolome. The LC/MS system for metabolomics analysis was composed of a Waters Acquity I-Class PLUS ultra-high-performance liquid tandem Xevo G2–XS QT of high-resolution mass spectrometry. A Waters Xevo G2–XS QTOF high-resolution mass spectrometer can collect primary and secondary mass spectrometry data in the MSe mode under the control of the acquisition software (MassLynx V4.2, Waters, United States). Dual-channel data acquisition can be performed simultaneously on both low collision energy and high collision energy in each

data acquisition cycle. The low collision energy is 2 V, the high collision energy range is 10–40 V, and the scanning frequency is 0.2 s to obtain a mass spectrum. The settings of the ESI ion source were as follows: capillary voltage: 2,000 V; cone voltage: 30 V; ion source temperature: 150°C; dissolvent gas temperature 500°C; backflush gas flow rate: 50 l/h; and desolventizing gas flow rate: 800 l/h.

Genome sequencing and assembly

Nanopore sequencing of 2 μ g of gDNA was prepared using an NEB Next FFPE DNA Repair Mix kit (New England Biolabs) and subsequently processed using the ONT Template prep kit (Oxford Nanopore Technologies). The large segment library was premixed with loading beads and then pipetted onto a previously used and washed R9 flow cell. The library was sequenced on the ONT PromethION platform with the Corresponding R9 cell and ONT sequencing reagents kit (EXP–FLP001.PRO.6; Oxford Nanopore Technologies). The genome was assembled based on three ways: initial wtdbg2 (Ruan and Li, 2020) assembly and then SMARTdenovo assembly (Liu et al., 2021), followed by error correction using racon (Vaser et al., 2017) software and adjustment by Pilon (Walker et al., 2014) software. The assembly results were evaluated by the ratio of sequencing reads, Core Eukaryotic Genes Mapping Approach (CEGMA; Parra et al., 2007), and Benchmarking Universal Single-Copy Orthologs (BUSCO; Simão et al., 2015).

Hi-C sequencing

We constructed Hi-C fragment libraries that ranged from insert sizes of 300–700 bp as illustrated in Rao et al. (2014) and sequenced them using the Illumina Novaseq 6,000 System (San Diego, CA, United States). The low-quality reads were removed and the clean data truncated, and the trimmed reads were then aligned to the assembly genome using a Burrows-Wheeler Aligner (BWA; Li and Durbin, 2009). Only unique paired-end reads that could be aligned and had a mapping quality > 20 were conserved. The valid interaction pairs were employed to correct the scaffolds and then ordered by LACHESIS (Burton et al., 2013). Finally, the vast majority of the sequences were located on the chromosomes.

Repeat annotation

We integrated *de novo* and homology-based methods to recognize repetitive sequences in the genome of DuyunMaojian tea plants. We used LTR_FINDER (Xu and Wang, 2007) to look for homology. *De novo* predictions were conducted using RepeatScout (Price et al., 2005), and then PASTECClassifier (Hoede et al., 2014) and RepeatMasker were used to predict the repetitive sequences.

Gene prediction and functional annotation

We used *ab initio*, homology, and RNA-Seq methods to predict the protein-coding genes of the tea plant. We used Genscan (Burge and Karlin, 1997), Augustus (Stanke and Waack, 2003), GlimmerHMM (Majoros et al., 2004), GeneID, and SNAP (Korf, 2004) to perform *ab initio* predictions. We compared the protein-coding genes from *Arabidopsis thaliana*, black cottonwood (*Populus trichocarpa*), coffee (*Coffea canephora*), rice (*Oryza sativa*), and kiwifruit (*Actinidia chinensis*) to the DuyunMaojian tea plant genome. The homology methods were carried out using GeMoMa (Keilwagen et al., 2016, 2018) software. We used TransDecoder, GeneMarkS-T (Tang et al., 2015), and PASA (Campbell et al., 2006) to predict the transcriptome. We then integrated the three predictions described using EVM software. We annotated the tRNA genes using tRNAscan-SE (Lowe and Eddy, 1997). We recognized the microRNA and rRNA based on the Rfam (Griffiths-Jones et al., 2005) database. We predicted the pseudogenes using BLAST (Kent, 2002) and GeneWise. We functionally annotated the protein-coding genes of the tea plant genome by performing BLAST searches against the NR, gene ontology (GO), EuKaryotic Orthologous Groups (KOG), Kyoto Encyclopedia of Genes and Genomes (KEGG), and TrEMBL databases.

Comparative genomic analysis

We applied Orthofinder v2.5.1 software to recognize gene families of the tea plant and 14 other plant species, including Shuchazao, Yunkang10, DASZ, *Camellia oleifera* Abel, kiwifruit (*Actinidia chinensis*), coffee (*C. canephora*), *Arabidopsis thaliana*, cacao (*Theobroma cacao*), blueberry (*Vaccinium corymbosum*), *Rhododendron delavayi*, rice (*Oryza sativa*), apple (*Malus domestica*), grape (*Vitis vinifera*), *Citrus clementina*. We generated high-quality single-copy genes that were used to build the phylogenetic tree among the 15 species described above. The expansion or contraction events of gene families were computationally identified by CAFE v4.2 (De Bie et al., 2006). The synonymous substitution rates (*Ks*) of genes were calculated using wgdv1.1.132. We searched for LTR-RT sequences using LTR_FINDERv1.07 and LTRharvestv1.5.9 software. Collinearity analyses were conducted using VGSC35 software.

The DEGs, DEPs and differential metabolites of the transcriptomic, proteomic and metabolomics analyses

The level of expression of the transcriptome genes was calculated using fragments per kilobase of transcript per million mapped reads (FPKM). The DEGs were investigated through DESeq with fold change ≥ 2 and FDR < 0.01 defined as DEGs. TMT and liquid chromatography–tandem mass spectrometry (LC–MS/MS) were used to quantitatively analyze the proteins of tea leaves infested with tea green leafhoppers. Proteins with fold change ≥ 2 and FDR < 0.01 were assigned as DEPs. UPLC-QTOF-MS techniques were used to analyze the metabolites qualitatively and quantitatively. We used principal component analysis (PCA), orthogonal projections to latent structures-discrimination analysis (OPLS-DA), and fold change (FC) to filter out the differential metabolites. The protein–protein interaction network (PPIN) of the DEPs were constructed based on two methods: (1) the Interolog (Matthews et al., 2001) method and BLAST were conducted between the DEPs and STRING databases (Szklarczyk et al., 2021); and (2) The existing DEP PPINs were extracted from STRING databases and TeaGPIN (Singh et al., 2021). We obtained the DEP PPINs by integrating the three DEPs PPINs.

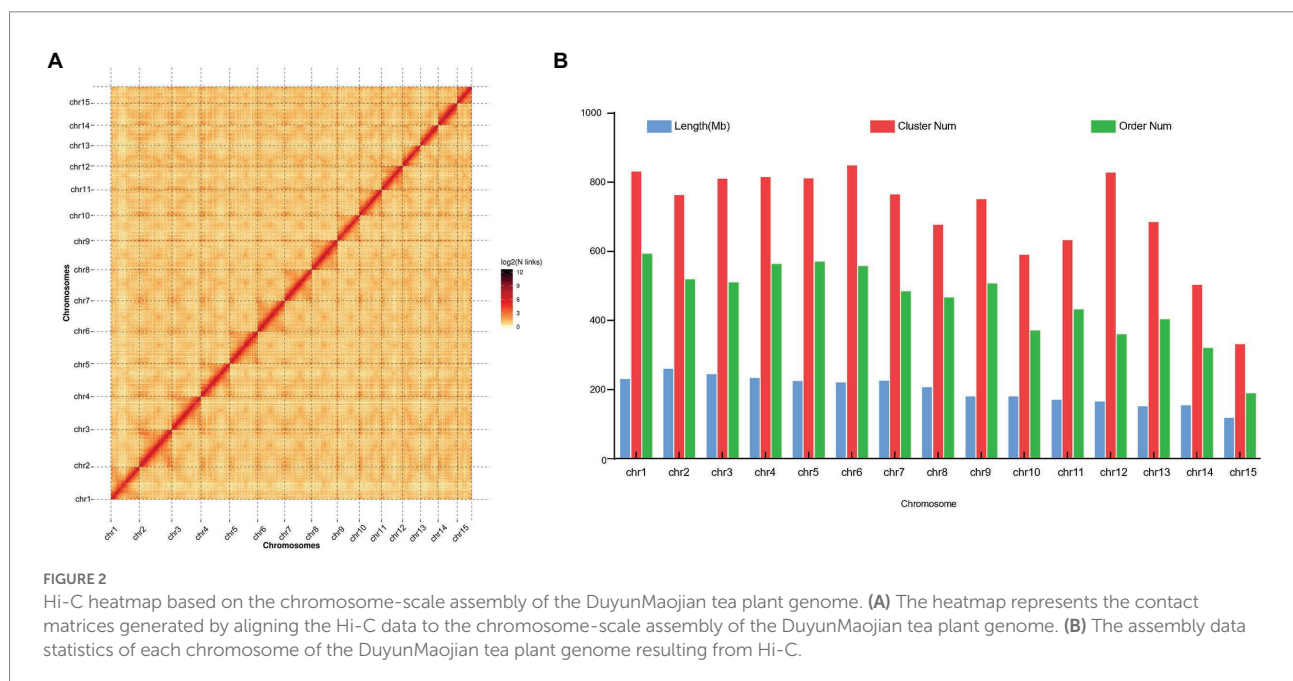
Results

Chromosome-scale assembly and annotation of the DuyunMaojian tea tree genome

The genome of the tea plant designated DuyunMaojian was sequenced and assembled using a combination of ONT and Hi-C. ONT produced 289.67 Gb of clean data, and the sequencing

TABLE 1 Summary statistics of the DuyunMaojian tea plant assembly genome compared with those of other cultivars and species.

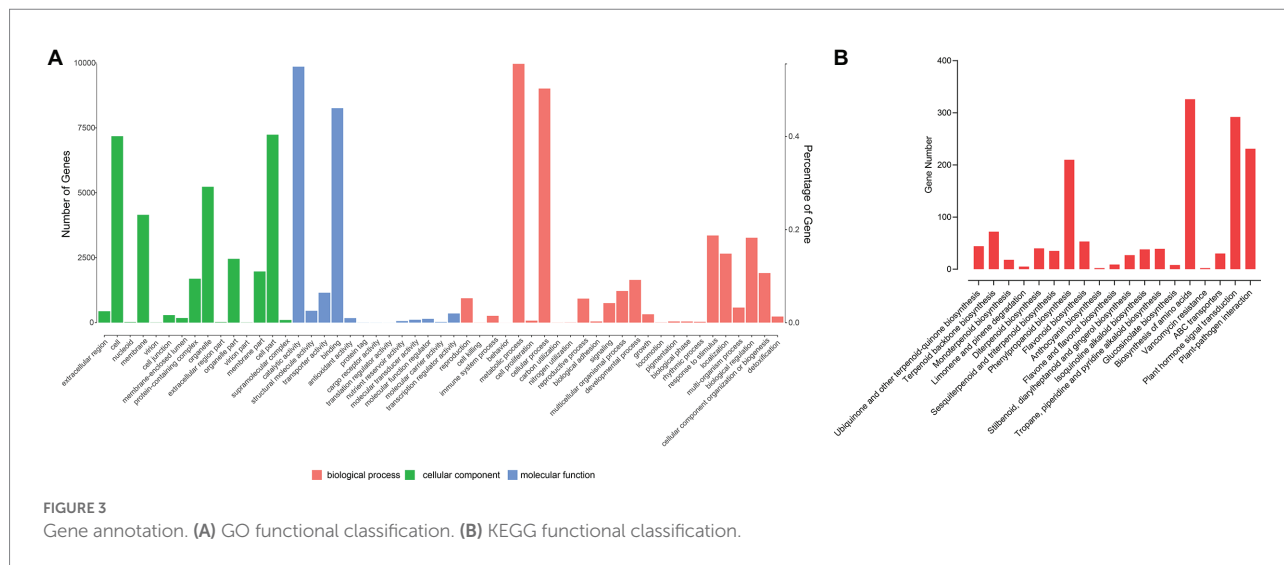
Genomic feature	DuyunMaojian using HIC	Shuchazao	Yunkang 10	DASZ	<i>Camellia oleifera</i>
Genome size (Gb)	2.97	2.94	3.02	3.11	2.89
Contig number	15,771	7,031	258,790	5,453	4,075
Contig N50 (kb)	723.697	600.46	19.96	2589.771	1,002
Contig max (kb)	6,456.183	2,885.68	257.648	16,831.527	/
Scaffold N50 (Mb)	207.721	218.1	0.45	204.21	185.364
Scaffold max (Mb)	250.703	/	3.51	336.927	/
GC content (%)	38.54	38.25	39.62	38.98	37.51
Gene number	34,896	50,525	36,951	33,021	42,426
Average gene length (bp)	6,961.34	4,906	3,549	8,050	3,955
Average exon length (bp)	232.25	245	237	211.7	317
Average intron length (bp)	1,117.79	973	640	1330.34	776
Reference	This study	Xia et al. (2020a)	Xia et al. (2017)	Zhang et al. (2020a,b)	Lin et al. (2022)



depth was 85.56X (Supplementary Tables S1–S3). We got 3.35 Gb genome assembly and contig N50 was 778.61 kb. The GC content was 38.56% (Table 1). To validate the genome assembly quality, we first mapped all the high-throughput clean reads (916,216,962) to the assembled genome, the mapping rate was 91.66% and a proper mapping rate was 85.47% (Supplementary Table S4). Secondly, a CEGMAv2.5 result demonstrated that 90.61% of the core genes were found in the DuyunMaojian tea tree genome (Supplementary Table S5). Third, BUSCO v2.0 data demonstrated that 87.78% of the key genes were located in the assembly (Supplementary Table S6). A Hi-C analysis (Supplementary Tables S7–S10) was introduced to enhance the quality of tea plant genome and build a chromosome-scale assembly. After Hi-C assembly and manual adjustment, 2.97 Gb of the genome was mapped to 15 pseudo-chromosomes that anchored 96.35% of the assembled sequences and 2.79 Gb can confirm the

order and direction (Figure 2). The redundant sequences from the heterozygous genome were removed by the Hi-C heat map, and the final genome was 3.08 Gb. After error correction, the contig N50, scaffold N50 and GC contents were 723.7 kb, 207.72 Mb and 38.54%, respectively (Table 1). The Hi-C heatmap perfectly showed 15 pseudo-chromosomes, which were designated chr1 to chr15, and the longest and shortest were chr2 and chr15 at 260,782,366 bp and 117,955,952 bp, respectively.

We identified 34,896 protein-coding genes and 2.67 Gb (86.77%) of repetitive sequences in the genome. Of the 34,896 genes, 33,481 (95.95%) were annotated with the GO, KEGG, KOG, Pfam, TrEMBL and NR databases and we identified 16,674 SSRs. The prediction of noncoding RNA genes produced 104 miRNA, 261 rRNA, and 669 tRNA (Supplementary Table S11). In addition, 6,502 pseudogenes were annotated in our genome. Finally,



we obtained 2,997 motifs and 37,323 domains. Retrotransposon elements include LTR and non-LTR. The LTR-RT include *Copia* and *Gypsy*, while the non-LTR include LINE and SINE. *Copia* and *Gypsy* were the predominant LTR-RTs and accounted for 12.5 and 43.68% of the DuyunMaojian tea plant genome, respectively (Supplementary Table S12). LINE is an autonomously active transposable element in the tea plant genome and accounts for about 1.68% of the genome. There were 27,830 SINEs, which accounted for 0.17% of the DuyunMaojian tea plant genome. They are small retrotransposable elements, which do not encode any transposable genes that are active. *Copia*, *Gypsy*, LINE, and SINE contribute to the expansion of the DuyunMaojian tea plant genome. Thus, retrotransposon elements may perform important functions in the responses to external stimulus.

Functional annotation of tea tree genome

There were 17,919 genes annotated in the GO database (Supplementary Table S13). In cellular components, there were 15 secondary categories in which 7,177 genes, 7,236 genes and 5,229 genes were annotated to cell, cell part and organelle, respectively. In molecular function, 9,856, 8,257, and 1,149 genes were annotated to catalytic activity, binding, and transporter activity, respectively. In biological process, 9,963, 9,013, 3,354, 1,641, 318, 258, and 233 genes were annotated to metabolic process, cellular process, response to stimulus, developmental process, growth, immune system process, and detoxification, respectively (Figure 3A). A total of 11,407 genes were annotated to 128 pathways. A total of 44 genes were participated in ubiquinone and other terpenoid-quinone biosynthesis; 72 genes were took part in terpenoid backbone biosynthesis; 18 genes participated in monoterpenoid biosynthesis; 210 genes were took part in phenylpropanoid biosynthesis; 53 genes were involved with flavonoid biosynthesis; 70 genes were involved in α -linolenic acid

metabolism; 292 genes participate in plant hormone signal transduction, and 231 genes participate in plant-pathogen interaction (Figure 3B). There were 17,700 genes annotated to the KOG database in all; 1,839 genes were classified to signal transduction mechanisms; 442 genes were classified into cell cycle, cell division, chromosome partitioning, and 213 genes were annotated to defense mechanism.

Comparative genomic analysis

Gene duplication within a species, evolution among species, and classification of species-specific genes (Figure 4A) were analyzed by comparing the genomes of DuyunMaojian tea plant with those of cacao (*T. cacao*), coffee (*C. canephora*), apple (*M. domestica*), *C. clementina*, kiwifruit (*A. chinensis*), blueberry (*V. corymbosum*), rice (*O. sativa*), *R. delavayi*, grape (*V. vinifera*), *A. thaliana*, Yunkang 10, Shuchazao, DASZ, and *C. oleifera*. The protein sequences of the DuyunMaojian tea plant and the other 14 species were classified using Orthofinderv2.5.1 (Emms, 2019). Thirty thousand and eight hundred and fifty three genes were grouped into 16,548 gene families (Table 2; Figure 4B). In addition, 215 gene families were specific to DuyunMaojian, and could be involved in specific biological processes, such as the large amount of synthesis of tea polyphenolics. IQ-TREEv1.6.11 (Nguyen and Schmidt, 2014), MAFFTv7.205 (Katoh et al., 2009), Gblocks0.91b (Talavera and Castresana, 2007), and ModelFinder (Kalyaanamoorthy et al., 2017) software were used to study the evolutionary relationships between Duyun Maojian and the other 14 species using single-copy protein sequences. The results showed that DuyunMaojian had the closest relationship with already available tea tree genome of Shuchazao and Yunkang 10, followed by DASZ, *C. oleifera*, *V. corymbosum*, *R. delavayi*, *A. chinensis*, *C. canephora*, *V. vinifera*, *T. cacao*, *A. thaliana*, *C. clementina*, *M. domestica*, and *O. sativa* (Figure 4C). Shuchazao is evaluated to have diverged from Yunkang 10 about 15 (15–27)

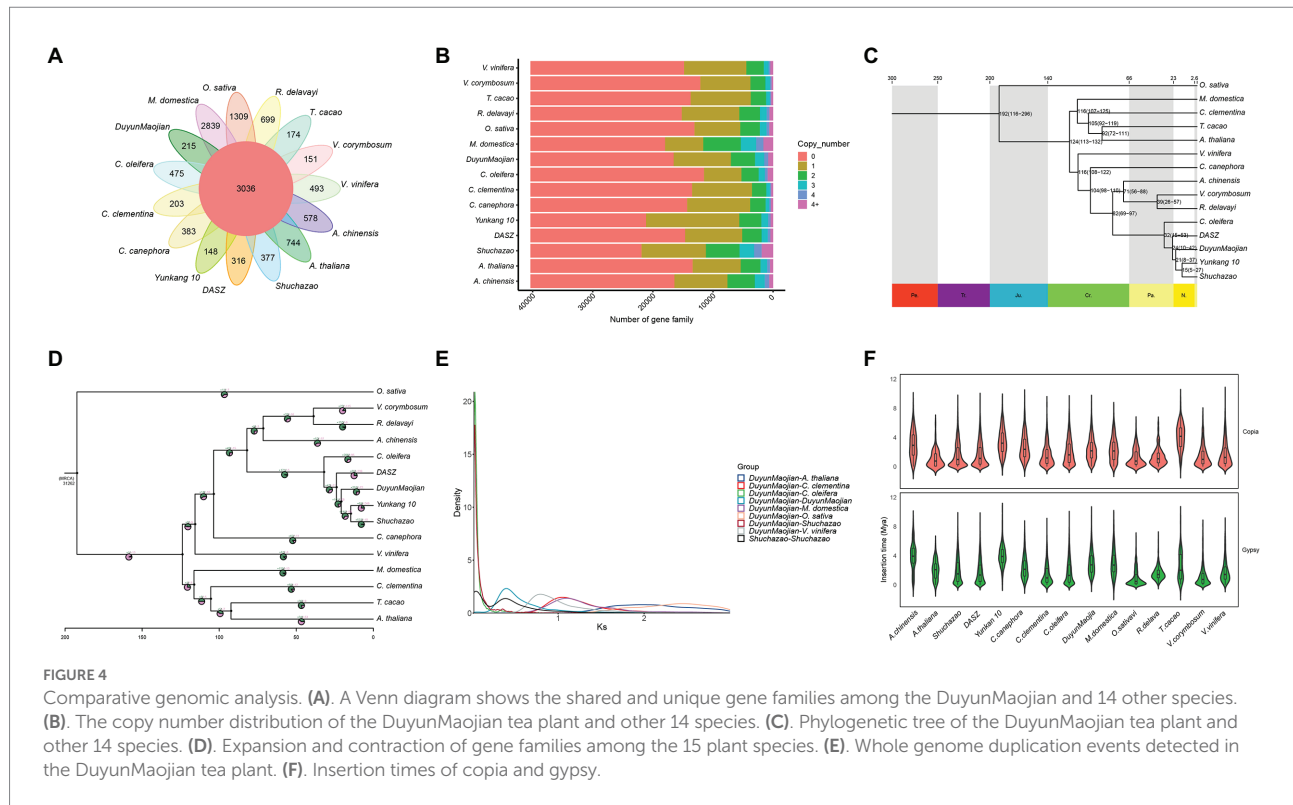
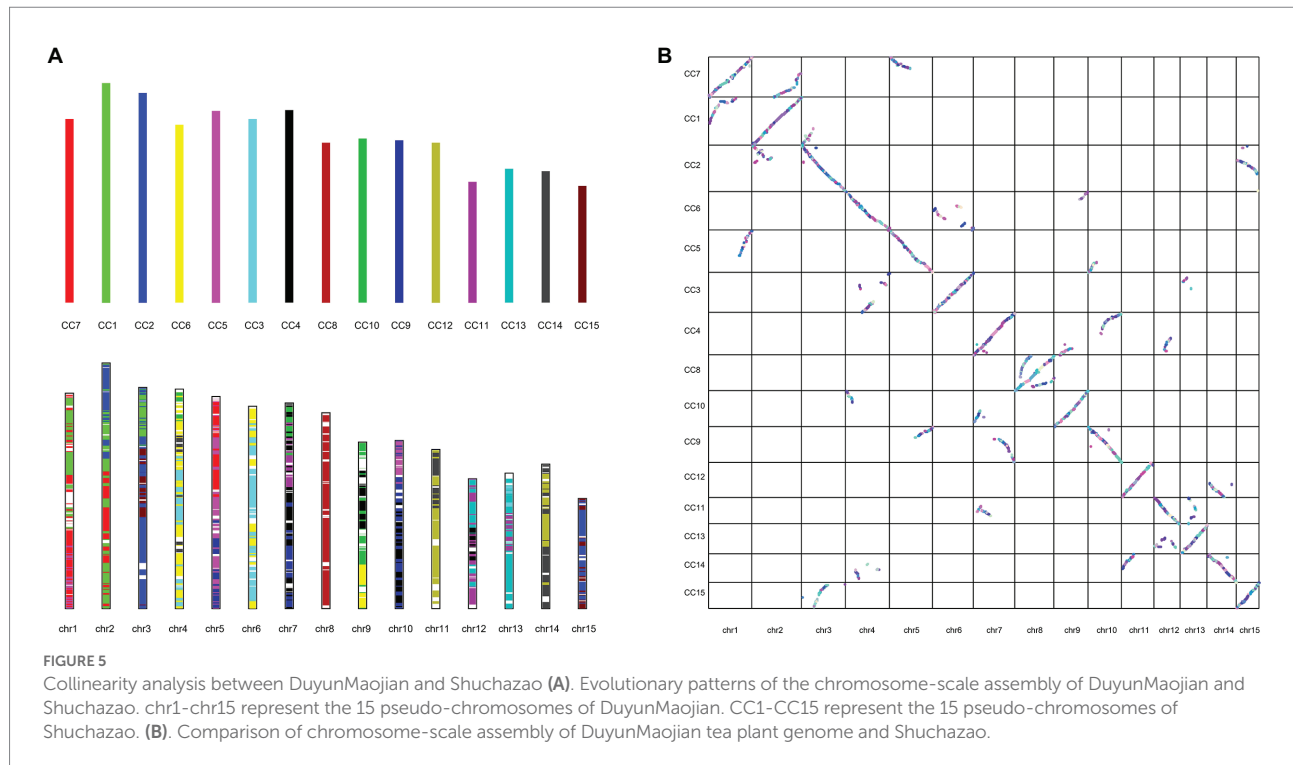


TABLE 2 Classification of gene family.

Name	Total gene	Cluster gene number	Total family	Unifamily
<i>Vaccinium corymbosum</i>	20,771	18,538	12,109	151
<i>Rhododendron delavayi</i>	32,410	26,819	15,208	699
Shuchazao	50,525	48,637	21,900	377
<i>Theobroma cacao</i>	21,254	20,614	13,712	174
DuyunMaojian	34,896	30,853	16,548	215
<i>Oryza sativa</i>	28,317	24,773	13,093	1,309
<i>Arabidopsis thaliana</i>	27,533	24,544	13,398	744
<i>Malus domestica</i>	55,414	43,649	17,974	2,839
<i>Vitis vinifera</i>	29,421	23,785	14,808	493
<i>Coffea canephora</i>	25,100	22,141	14,306	383
<i>Citrus clementina</i>	22,207	20,342	13,495	203
<i>Actinidia chinensis</i>	39,444	30,001	16,420	578
Yunkang 10	36,951	31,246	21,130	148
DASZ	28,981	24,339	14,630	316
<i>Camellia oleifera</i>	33,173	31,856	11,504	475

million years ago (MYA). DuyunMaojian is estimated to have diverged from Shuchazao and Yunkang 10 approximately 21 (8–37) MYA and split from DASZ approximately 24 (10–42) MYA. The cultivar DuyunMaojian is also evaluated to have diverged from *C. oleifera* approximately 32 (15–53) MYA and split from blueberry, *R. delavayi* and kiwifruit about 82 (69–97) MYA and from grape about 116 (108–122) MYA. We found that 201 gene families had expanded and 72 contracted in the DuyunMaojian tea plant genome using CAFÉv4.2 (De Bie et al.,

2006; Figure 4D). The K_s of genes were calculated using wgdv1.1.1 (Zwaenepoel, 2019) and revealed that a recent whole-genome duplication (WGD) event ($K_s = 0.39$) was occurred in DuyunMaojian tea plant genome (Figure 4E). We searched for LTR-RT sequences using LTR_FINDERv1.07 (Xu and Wang, 2007) and LTRharvestv1.5.9 (Ellinghaus et al., 2008) software and the insertion time of *Copia* and *Gypsy* among the 15 species (Figure 4F). Collinearity analyses were conducted using VGSC (Xu et al., 2016) software, and there were 42,543 collinear genes



between DuyunMaojian and Shuchazao (Figure 5). All these findings will greatly enhance our understanding of the diversification history of the DuyunMaojian tea plant genome.

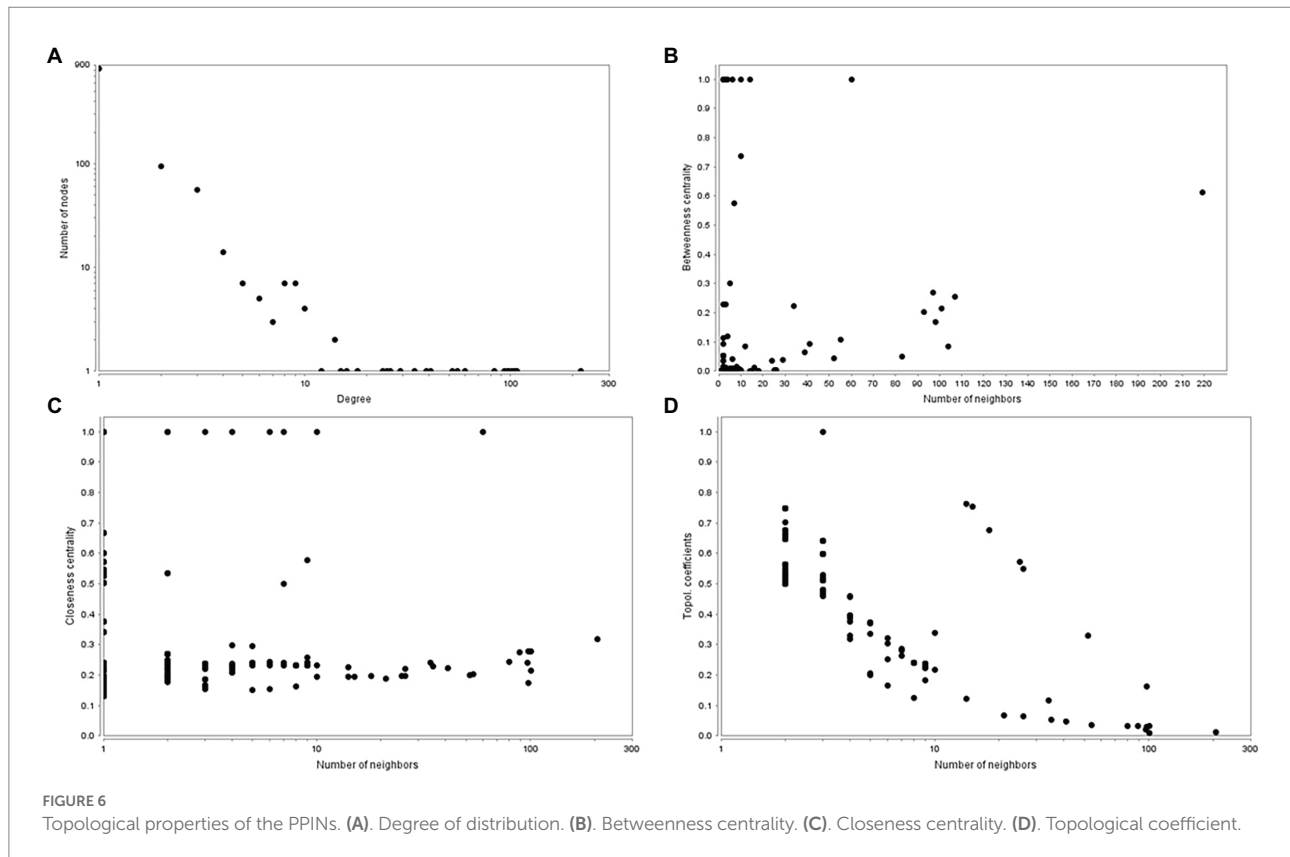
Multi-omics analysis of the response of tea plants to infestation with tea green leafhoppers

Tea green leafhopper (Liu et al., 2017; Yang, 2017) can significantly reduce the yield and quality of tea during the growth of plant. These insects can cause a loss of 11–55% in tea yields in the absence of effective means of prevention and treatment. Thus, we also used the tea plant transcriptome, proteome and metabolome in combination with its genome to analyze the DEGs, DEPs and differential metabolites after leafhopper infestation. We paid more attention on tea plants responds to infestation with tea green leafhoppers in a short time rather than a long-term mechanism. At the same time, we are trying to screen for the key DEGs, DEPs, and differential metabolites between the two groups. Therefore, we only selected 24h for comparison of two groups after infestation by tea green leafhoppers. There were 1,575 DEGs, 871 DEPs, and 41 differential metabolites, respectively. A total of 281 DEGs, 353 DEPs, and eight differential metabolites identified through these processes were annotated to the KEGG database, respectively (Supplementary Tables S14–S16). We found that phenylpropanoid biosynthesis pathway, flavonoid biosynthesis pathway, and α -linolenic acid metabolism were simultaneously identified in the transcriptome, proteome, metabolome and genome. Therefore, we hypothesized that proteins and metabolites

involved in the three pathways play important roles in the resistance to tea green leafhopper. The results indicated that phenylpropanoid biosynthesis, α -linolenic acid metabolism, and flavonoid biosynthesis contained 50 DEPs that were filtered out and found in the tea plant transcriptome, proteome, metabolome and genome (Supplementary Table S17). Thus, we hypothesized that the three pathways and 50 DEPs play important roles in response to tea green leafhoppers, which will provide a theoretical foundation to breed tea plants resistant to insects and ultimately result in increases in tea yield and quality.

Protein–protein interaction network analysis of the response of tea plants to infestation with tea green leafhoppers

The predicted tea tree protein–protein interaction network (PPIN) was available in STRING database, TeaGPIN (Singh et al., 2021) and TeaLIPIN (Singh et al., 2019). There were 820 proteins and 1,067 protein–protein interactions (PPIs) in the predicted network based on the existing PPIN in the STRING database and the interolog method. We also extracted DEPs PPIN from TeaGPIN, in which there were 389 proteins and 583 PPIs (Supplementary Table S18). There were 362 shared PPIs in both network. We obtained the DEPs PPIN by integrating the two DEPs PPINs. In the merged DEPs PPIN (Supplementary Table S19), there were 1,047 nodes and 1,430 PPIs and a degree distribution of $P(k) \approx 53.574x^{-1.008}$, $R^2 = 0.684$. The topological properties (Figure 6) of the PPIN showed that the degree of distribution of the proteins obeyed a power-law distribution (Figure 6A). There

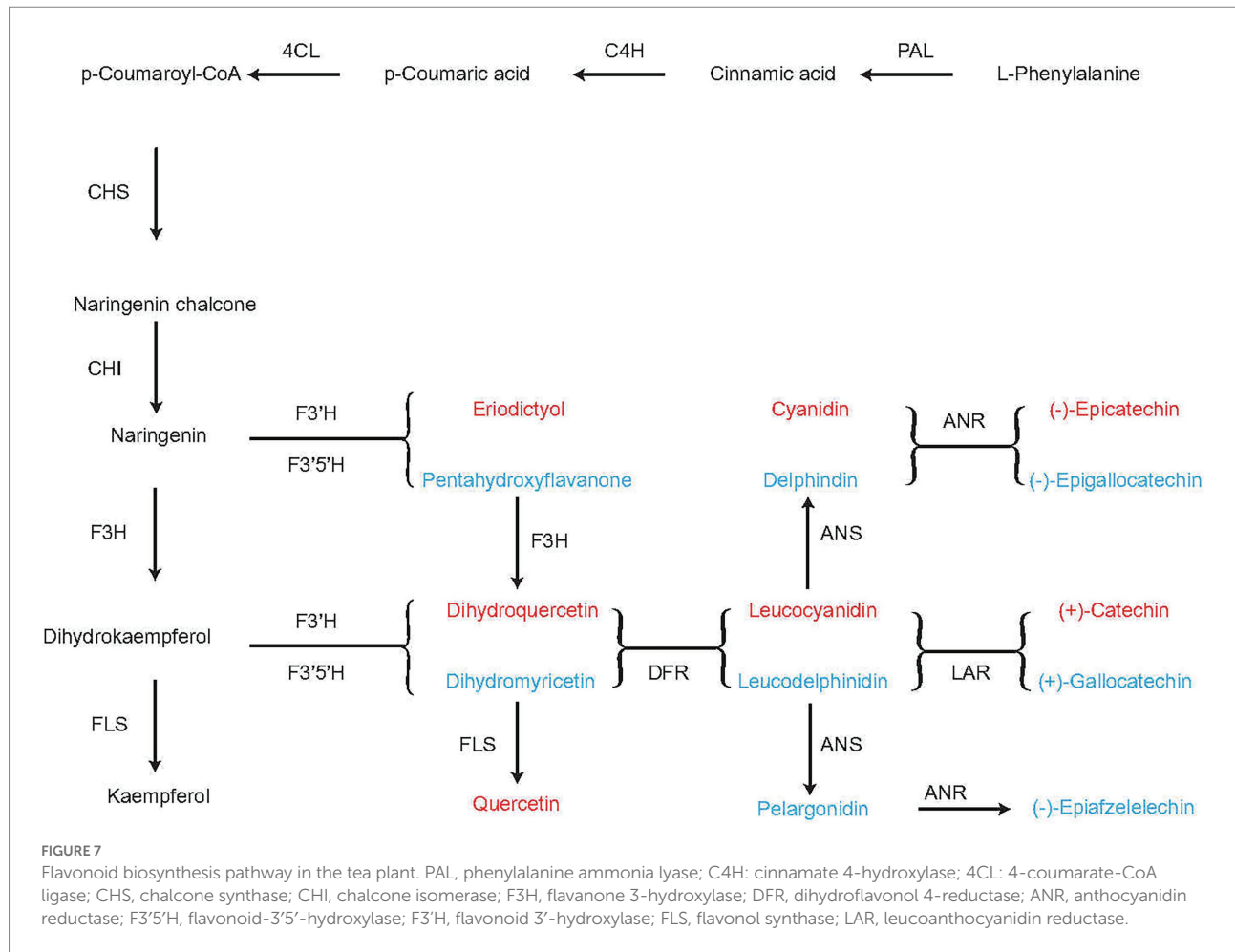


were a few proteins that have a high degree, which is a scale-free phenomenon. Highly connected proteins with central roles in the PPIN are hubs, and the degree of the hub proteins was calculated (Supplementary Table S20). These are the 21 proteins with the largest number of interactions, suggesting that they are also the most important proteins (hubs) in the PPIN. The betweenness centrality (Figure 6B) was a measure of a node's centrality in a network equal to the number of shortest paths from all vertices to all the others that pass through that node. Its value was between 0 and 1. $C_b(i) = \sum_{m \neq i \neq n} (\sigma_{mn}(i) / \sigma_{mn})$, where m and n were the nodes in the network that differed from i , and σ_{mn} denoted the number of the shortest paths from m to n . $\sigma_{mn}(i)$ was the number of shortest paths from m to n that i laid on. The closeness centrality (Figure 6C) of each node was also between 0 and 1 and was used to identify important positions within the network. $C_c(i) = 1 / \text{avg.}(S[u, v])$, where $S[u, v]$ was the length of shortest path between the nodes u and v . $C_c(i)$ of node i is the reciprocal of the average shortest path length. The topological coefficient (Figure 6D) of protein i is $T_i = \text{avg.}(J(i, j)) / n_i$, and $J(i, j)$ that indicate that all nodes j share at least one neighbor with i . The topological coefficient decreases with the increasing number of neighbors, revealing that the nodes with many neighbors are not artificially clustered together. Thus, the merged DEPs PPINs of the 50 DEPs is reliably based on the topological property analysis described above. We hypothesized that the 21 hubs with the largest degrees, particularly peroxidase, played important roles in the response to tea green leafhopper infestation.

Discussion

We reported the assembly of DuyunMaojian tea plant genome using a combination of the ONT PromethION™ combined with Hi-C technology, and the final genome was 3.08 Gb. A total of 2.97 Gb of the genome was mapped to 15 pseudo-chromosomes, among which 2.79 Gb can confirm the order and direction. The contig N50 was 723.7 kb, which was longer than that of Yunkang 10 (Xia et al., 2017) and shorter than that of Shuchazao (Xia et al., 2020a). The GC contents were 38.54%, which was consistent with those of the Yunkang 10 and Shuchazao genomes (Xia et al., 2020a). The ratio of *Copia* was higher than those in the CSSV1.1 genome, and the ratio of *Gypsy* was lower than those in the CSSV1.1 genome. In addition, we identified 16,674 SSRs, which was far less than those of the CSSV1.2 and CSA genomes.

A comparative genomics analysis showed that DuyunMaojian was the most closely related to Shuchazao and Yunkang 10, followed by DASZ, and *C. oleifera*. DuyunMaojian was evaluated to have diverged from *C. oleifera* approximately 32 (15–53) MYA, split from blueberry, *R. delavayi*, and kiwifruit about 82 (69–97) MYA and from grape about 116 (108–122) MYA. These estimates are similar to those of earlier studies (Shi et al., 2010; Fiz-Palacios et al., 2011; Wu, 2019). The K_s of genes were calculated using wgdv1.1.1 (Zwaenepoel, 2019) and revealed that a recent whole-genome duplication (WGD) event ($K_s = 0.39$) occurred in the DuyunMaojian tea plant genome, which was consistent with the findings of previous studies (Xia et al., 2017, 2020a; Wang et al.,

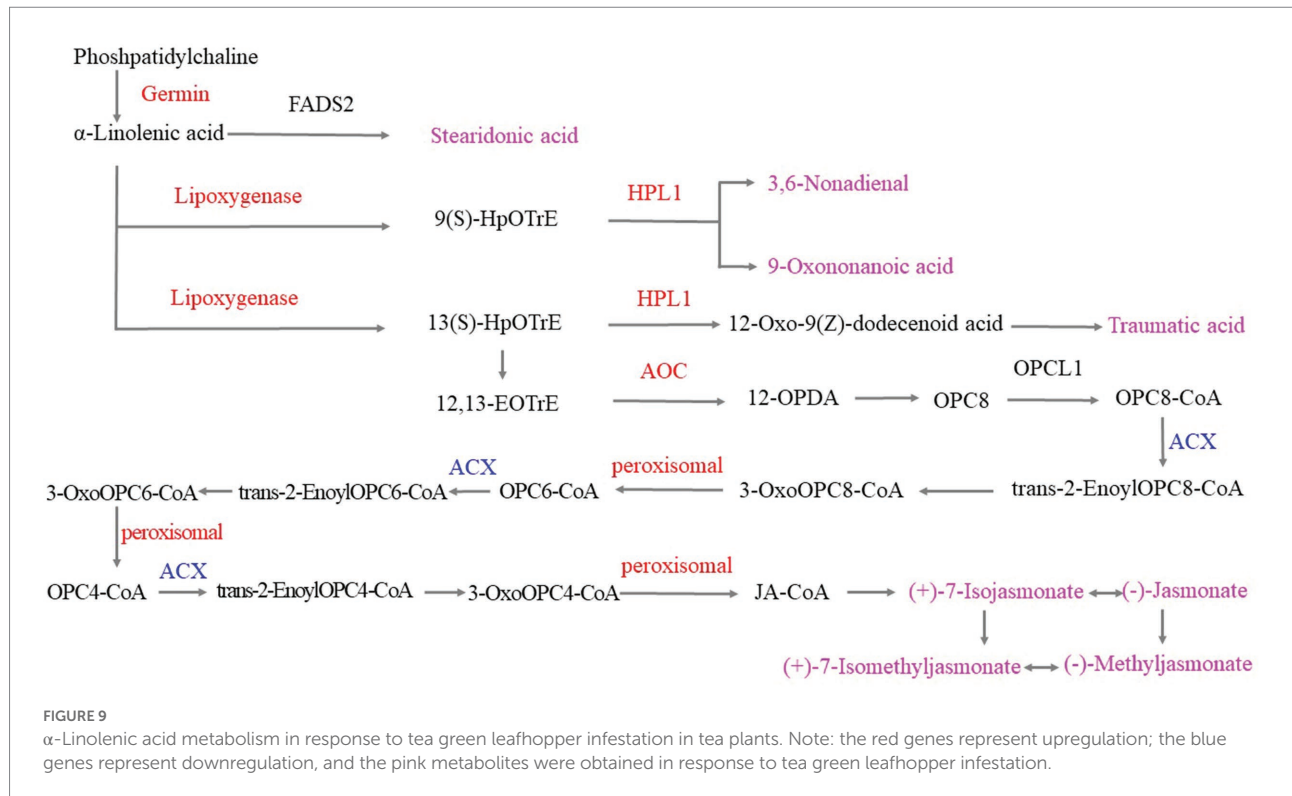


2020). These data provide theoretical references for the cultivation of excellent tea varieties.

There were 33,481 annotated genes in the assembly genome, which were lower than those in the CSSV1.2 genome (Xia et al., 2020a) and higher than those in the CSSV1.1 genome (Wei et al., 2018). Tea plant leaves contain an extraordinarily high level of flavonoids that contribute to its health benefits and significantly affect its flavor, taste, and mouthfeel. Flavonoids, particularly flavonol glycosides and catechins, are the major source of the bitterness and astringency of tea. There were 53 genes annotated to flavonoid biosynthesis, 210 genes annotated to phenylpropanoid biosynthesis, and 70 genes annotated to α -linolenic acid metabolism in the tea plant genome, which contained 50 DEGs filtered out that were identified in the transcriptome, proteome and metabolome. Thus, we hypothesized that the three pathways (Zhao et al., 2020) and 50 DEGs played important roles in the response to tea green leafhopper infestation.

When the tea plants were infested with tea green leafhoppers, they were influenced by DEPs, such as peroxidase. The putative flavonoid biosynthesis *via* the phenylpropanoid pathway are converted from cinnamoyl-CoA to pinobanksin 3-acetate by the enzymes chalcone synthase (CHS), chalcone isomerase (CHI), and flavanone-3-hydroxylase (F3H). L-Phenylalanine can be converted to

p-coumaroyl-CoA through the subsequent action of phenylalanine ammonia lyase (PAL), cinnamate-4-hydroxylase (C4H), and 4-coumarate-CoA-ligase (4CL). *p*-Coumaroyl-CoA can be converted to naringenin through the action of CHS and CHI. Naringenin can be converted to either (–)-epicatechin through the subsequent actions of flavonoid 3'-hydroxylase (F3'H), flavonoid 3', 5'-hydroxylase (F3'5'H), and anthocyanidin reductase (ANR) or to (–)-epigallocatechin through the subsequent action of 3'-hydroxylase (F3'H), flavonoid 3', 5'-hydroxylase (F3'5'H), and anthocyanidin reductase (ANR). F3H catalyzes naringenin to dihydrokaempferol, which can be converted to either kaempferol through the action of flavonol synthase (FLS), or to dihydroquercetin, leucocyanidin, and finally (+)-catechin through the subsequent actions of F3'H, F3'5'H, dihydroflavonol 4-reductase (DFR), and leucoanthocyanidin reductase (LAR) or to (+)-gallocatechin through the subsequent actions of F3'H, F3'5'H, DFR and LAR. Naringenin can be converted either eriodictyol or pentahydroxyflavanone by the enzymes F3'H and F3'5'H, respectively. Eriodictyol and pentahydroxyflavanone can subsequently be converted to dihydroquercetin and dihydromyricetin by F3H, respectively, and finally converted to quercetin by FLS. DFR catalyzes the transformation of dihydromyricetin to leucodelphinidin, which can be converted to (–)-epiafzelechin by ANS and ANR (Figure 7). Thus, in the flavonoid biosynthesis pathway, we obtained



oxyoctadecatrienoic acid (HpOTrE) by the upregulation of lipoxygenase or to 3,6-nonadienal and 9-oxononanoic acid by the upregulation of hydroperoxide lyase 1 (HPL1). Next, α -linolenic acid was catalyzed by lipoxygenase, HPL1 and other enzymes to produce traumatic acid. Moreover, responses to the infestation of tea green leafhoppers included the catalysis of α -linolenic acid by upregulated lipoxygenase to obtain 13(S)HpOTrE, which was catalyzed by the related enzyme and then catalyzed the production of 12,13-EOTrE. Upregulated allene-oxide cyclase (AOC) catalyzed the transformation of 12,13-EOTrE to 12-OPDA, and 12-OPDA was then converted to (+)-7-isojasmonate, (-)-jasmonate, (+)-7-isomethyljasmonate, and (-)-methyl jasmonate by the downregulated ACX and upregulated peroxisomal, which can interconvert these compounds. Thus, we hypothesized that stearidonic acid, 9(S)-HOTrE, 3,6-nonadienal, 9-oxononanoic acid, 10-OPDA, traumatic acid, (+)-7-isojasmonate, (-)-jasmonate, (+)-7-isomethyl jasmonate, and (-)-methyl jasmonate played important roles in the response to tea green leafhopper infestation, which was consistent with the findings of previous studies (Nafie and Hathout, 2011; Xin et al., 2016; Zhao et al., 2020; Figure 9).

In the PPIN, we hypothesized that the hubs DEPs (Supplementary Table S20) and peroxidase in particular, may play important roles in the response to infestation by tea green leafhoppers.

In brief, the chromosome-scale genome of tea plant is highly promising to help understand the evolution of tea genome and then discover how the tea plant responds to tea green leafhopper. Our results suggest that (1) Duyun Maojian had the closest relationship with Shuchazao and Yunkang 10, followed by DASZ, and *C. oleifera*. (2) A recent whole-genome duplication (WGD) event ($K_s = 0.39$) occurred in the DuyunMaojian tea plant genome (3) The tea plant

could become resistant to tea leafhoppers by the biosynthesis of lignin, which could have potential application values to prevent insect infestations on tea plants. (4) (-)-Epicatechin, (-)-epigallocatechin, (+)-catechin, (+)-gallocatechin, (-)-epiafzelechin, kaempferol, and quercetin played important roles in the response to tea green leafhopper infestation. (5) Stearidonic acid, 9(S)-HOTrE, 3,6-nonadienal, 9-oxononanoic acid, 10-OPDA, traumatic acid, (+)-7-isojasmonate, (-)-jasmonate, (+)-7-isomethyljasmonate, and (-)-methyl jasmonate played important roles in response to tea green leafhopper infestation. (6) The hubs DEPs and peroxidase, in particular, may play important roles in the response to the tea green leafhopper infestation.

Data availability statement

The datasets presented in this study can be found in the NCBI Sequence Read Archive under the BioProject accession number PRJNA841059. The DuyunMaojian tea plant gene and functional annotations are available in the figshare database (DOI: doi.org/10.6084/m9.figshare.19972808.v1).

Author contributions

LL designed the study. DW and RL contributed to the sample preparation and genome sequencing. FW, BZ, and ML conducted the genome assembly, performed genome annotation and interpreted comparative genomic analysis. FW, BZ, XY, HP, RM, and ZC performed multi-omics analysis. FW wrote the

manuscript. LL, BS, and FW modified the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (grant number 32260080 and 31900486); Guizhou Provincial Science and Technology Foundation [(2019)1298]; Guizhou Provincial Education Department [ZDXK(2016)23, (2016)020, QNYSKYPT2018007, (2020)071, and (2015)68]; the Scientific Research Project of Qiannan Normal University for Nationalities (QNYSKYTD2018011, Qnsyk201605, QNSY2018BS018, 2018xjg0520, QNYSXXK 2018005, QNSY2018ZJ006, and 2019xjg0303).

Conflict of interest

ML was employed by Biomarker Technologies Corporation.

References

- Ahmed, G. J., and Li, X. (2022). Hormonal regulation of health-promoting compounds in tea (*Camellia sinensis* L.). *Plant Physiol. Biochem.* 185, 390–400. doi: 10.1016/j.plaphy.2022.06.021
- Arimura, G.-I., Ozawa, R., and Maffei, M. E. (2011). Recent advances in plant early signaling in response to Herbivory. *Int. J. Mol. Sci.* 12, 3723–3739. doi: 10.3390/ijms12063723
- Bhattacharyya, C., Imchen, M., Mukherjee, T., Haldar, S., Mondal, S., Mukherji, S., et al. (2022). Rhizosphere impacts bacterial community structure in the tea (*Camellia sinensis* (L.) O. Kuntze.) estates of Darjeeling, India. *Environ. Microbiol.* 24, 2716–2731. doi: 10.1111/1462-2920.15874
- Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78–94. doi: 10.1006/jmbi.1997.0951
- Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31, 1119–1125. doi: 10.1038/nbt.2727
- Campbell, M. A., Haas, B. J., Hamilton, J. P., Mount, S. M., and Buell, C. R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics* 7, 1–17. doi: 10.1186/1471-2164-7-327
- Chen, J.-D., Zheng, C., Ma, J.-Q., Jiang, C.-K., Ercisli, S., Yao, M.-Z., et al. (2020). The chromosome-scale genome reveals the evolution and diversification after the recent tetraploidization event in tea plant. *Hortic. Res.* 7:63. doi: 10.1038/s41438-020-0288-2
- De Bie, T., Cristianini, N., Demuth, J. P., and Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22, 1269–1271. doi: 10.1093/bioinformatics/btl097
- Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9, 1–14. doi: 10.1186/1471-2105-9-18
- Emms, D. M. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238. doi: 10.1186/s13059-019-1832-y
- Fang, Z., Jin, L., Ye, Y., He, W., Shu, Z., Shao, J., et al. (2022). Effects of different shading treatments on the biomass and Transcriptome profiles of tea leaves (*Camellia sinensis* L.) and the regulatory effect on phytohormone biosynthesis. Pdf. *Front. Plant Sci.* 13, 1–11. doi: 10.3389/fpls.2022.909765
- Feng, F., Yao, Y., Wang, X. Q. D., Zhang, X., and Liu, J. (2022). Connecting high-resolution 3D chromatin organization with epigenomics. *Nat. Commun.* 13:2054. doi: 10.1038/s41467-022-29695-6
- Fitz-Palacios, O., Schneider, H., Heinrichs, J., and Savolainen, V. (2011). Diversification of land plants: insights from a family-level phylogenetic analysis. *BMC Evol. Biol.* 11:341. doi: 10.1186/1471-2148-11-341
- Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S. R., and Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* 33, D121–D124. doi: 10.1093/nar/gki081
- Han, Z.-X., Rana, M. M., Liu, G.-F., Gao, M.-J., Li, D.-X., Wu, F.-G., et al. (2016). Green tea flavour determinants and their changes over manufacturing processes. *Food Chem.* 212, 739–748. doi: 10.1016/j.foodchem.2016.06.049
- Hazarika, L. K., Bhuyan, M., and Hazarika, B. N. (2009). Insect pests of tea and their management. *Annu. Rev. Entomol.* 54, 267–284. doi: 10.1146/annurev.ento.53.103106.093359
- Hoede, C., Arnoux, S., Moisset, M., Chamier, T., Inizan, O., Jamilloux, V., et al. (2014). PASTE: an automatic transposable element classification tool. *PLoS One* 9:e91929. doi: 10.1371/journal.pone.0091929
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Katoh, K., Asimenos, G., and Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* 537, 39–64. doi: 10.1007/978-1-59745-251-9_3
- Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O., and Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinf.* 19, 1–12.
- Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., and Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* 44:e89. doi: 10.1093/nar/gkw092
- Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664.
- Khan, N., Adhmi, V. M., and Mukhtar, H. (2008). Apoptosis by dietary agents for prevention and treatment of cancer. *Biochem. Pharmacol.* 76, 1333–1339. doi: 10.1016/j.bcp.2008.07.015
- Kingdom-Ward, F. (1950). Does wild tea exist? *Nature* 165, 297–299. doi: 10.1038/165297a0
- Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* 5:59. doi: 10.1186/1471-2105-5-59
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Lin, P., Wang, K., Wang, Y., Hu, Z., Yan, C., Huang, H., et al. (2022). The genome of oil-camellia and population genomics analysis provide insights into seed oil domestication. *Genome Biol.* 23:14. doi: 10.1186/s13059-021-02599-2
- Liu, J.-H., Sun, C.-Y., Long, J., and Guo, J.-J. (2017). Complete mitogenome of tea green leafhopper, *Empoasca onukii* (Hemiptera: Cicadellidae) from Anshun,

The remaining 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.2022.1004387/full#supplementary-material>

- Guizhou Province in China. *Taylor Francis* 2, 808–809. doi: 10.1080/23802359.2017.1398616
- Liu, H., Wu, S., Li, A., and Ruan, J. (2021). SMARTdenovo: a *de novo* assembler using long noisy reads. *Gigabyte* 2021, 1–9. doi: 10.46471/gigabyte.15
- 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, 955–964. doi: 10.1093/nar/25.5.955
- Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20, 2878–2879. doi: 10.1093/bioinformatics/bth315
- Matthews, L. R., Vaglio, P., Reboul, J., Ge, H., Davis, B. P., Garrels, J., et al. (2001). Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or “Interologs.” *Genome Res.* 11, 2120–2126. doi: 10.1101/gr.205301
- Mei, X., Liu, X., Zhou, Y., Wang, X., Zeng, L., Fu, X., et al. (2017). Formation and emission of linalool in tea (*Camellia sinensis*) leaves infested by tea green leafhopper (*Empoasca (Matsumurasca) onukii* Matsuda). *Food Chem.* 237, 356–363. doi: 10.1016/j.foodchem.2017.05.124
- Munasinghe, M., Deraniyagala, Y., Dassanayake, N., and Karunarathna, H. (2017). Economic, social and environmental impacts and overall sustainability of the tea sector in Sri Lanka. *Sustain. Prod. Consum.* 12, 155–169. doi: 10.1016/j.spc.2017.07.003
- Na Nagara, V., Sarkar, D., Luo, Q., Biswas, J. K., and Datta, R. (2022). Health risk assessment of exposure to trace elements from drinking black and green tea marketed in three countries. *Biol. Trace Elem. Res.* 200, 2970–2982. doi: 10.1007/s12011-021-02863-3
- Nafe, E., and Hathout, T. (2011). Jasmonic acid elicits oxidative defense and detoxification systems in Cucumis melo L. cells. *Braz. J. Plant Physiol.* 23, 161–174.
- Natukunda, M. I. (2021). Interaction between rag genes results in a unique synergistic transcriptional response that enhances soybean resistance to soybean aphids. *BMC Genomics* 22. doi: 10.1186/s12864-021-08147-3
- Nguyen, L.-T., and Schmidt, H. A. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Nicholls, S. M., Quick, J. C., Tang, S., and Loman, N. J. (2019). Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *GigaScience* 8, 1–9. doi: 10.1093/gigascience/giz043
- Nyhus Dhillon, C., Vossenaar, M., Weiligmann, B., Sanwal, N., Djimeu, E. W., Kneepkens, M., et al. (2022). A nutrition behavior change program moderately improves minimum diet diversity and Handwashing behaviors among tea workers in Assam and Tamil Nadu, India. *Food Nutr. Bull.* 43, 159–170. doi: 10.1177/037957212111070706
- Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23, 1061–1067. doi: 10.1093/bioinformatics/btm071
- Price, A. L., Jones, N. C., and Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics* 21, i351–i358. doi: 10.1093/bioinformatics/bti1018
- Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at Kilobase resolution reveals principles of chromatin looping. *Cells* 159, 1665–1680. doi: 10.1016/j.cell.2014.11.021
- Ross, J. A., and Kasum, C. M. (2002). Dietary flavonoids bioavailability, metabolic effects, and safety. *Annu. Rev. Nutr.* 22, 19–34. doi: 10.1146/annurev.nutr.22.111401.144957
- Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* 17, 155–158. doi: 10.1038/s41592-019-0669-3
- Shi, T., Huang, H., and Barker, M. S. (2010). Ancient genome duplications during the evolution of kiwifruit (*Actinidia*) and related Ericales. *Ann. Bot.* 106, 497–504. doi: 10.1093/aob/mcq129
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Singh, G., Singh, V., and Singh, V. (2019). Construction and analysis of an interologous protein-protein interaction network of *Camellia sinensis* leaf (TeaLIPIN) from RNA-Seq data sets. *Plant Cell Rep.* 38, 1249–1262. doi: 10.1007/s00299-019-02440-y
- Singh, G., Singh, V., and Singh, V. (2021). Genome-wide interologous interactome map (TeaGPIP) of *Camellia sinensis*. *Genomics* 113, 553–564. doi: 10.1016/j.ygeno.2020.09.048
- Souza, C. S. F., Souza, B. H. S., Parrella, R. A. C., Simeone, M. L. F., Nascimento, P. T., França, J. C. O., et al. (2021). Resistance of bmr energy sorghum hybrids to sugarcane borer and fall armyworm. *Braz. J. Biol.* 84, 1–9. doi: 10.1590/1519-6984.251883
- Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19, ii215–ii225. doi: 10.1093/bioinformatics/btg1080
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49, D605–D612. doi: 10.1093/nar/gkaa1074
- Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577. doi: 10.1080/10635150701472164
- Tan, H. R., Lau, H., Liu, S. Q., Tan, L. P., Sakumoto, S., Lassabliere, B., et al. (2019). Characterisation of key odourants in Japanese green tea using gas chromatography-olfactometry and gas chromatography-mass spectrometry. *LWT* 108, 221–232. doi: 10.1016/j.lwt.2019.03.054
- Tang, S., Lomsadze, A., and Borodovsky, M. (2015). Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res.* 43:e78. doi: 10.1093/nar/gkv227
- Tounekti, T., Joubert, E., Hernández, I., and Munné-Bosch, S. (2013). Improving the polyphenol content of tea. *Crit. Rev. Plant Sci.* 32, 192–215. doi: 10.1080/07352689.2012.747384
- Treutter, D. (2006). Significance of flavonoids in plant resistance: a review. *Environ. Chem. Lett.* 4, 147–157. doi: 10.1007/s10311-006-0068-8
- Vaser, R., Sovic, I., and Nagarajan, N. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi: 10.1101/gr.214270.116
- Walker, B. J., Abee, T., Shea, T., Priest, M., Abouelieil, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:15. doi: 10.1371/journal.pone.0112963
- Wang, X., Feng, H., Chang, Y., Ma, C., Wang, L., Hao, X., et al. (2020). Population sequencing enhances understanding of tea plant evolution. *Nat. Commun.* 11:10. doi: 10.1038/s41467-020-18228-8
- Wang, Y.-N., Tang, L., Hou, Y., Wang, P., Yang, H., and Wei, C.-L. (2016). Differential transcriptome analysis of leaves of tea plant (*Camellia sinensis*) provides comprehensive insights into the defense responses to *Ectropis obliqua* attack using RNA-Seq. *Funct. Integr. Genomics* 16, 383–398. doi: 10.1007/s10142-016-0491-2
- Wang, P., Yu, J., Jin, S., Chen, S., Yue, C., Wang, W., et al. (2021). Genetic basis of high aroma and stress tolerance in the oolong tea cultivar genome. *Hortic. Res.* 8:107. doi: 10.1038/s41438-021-00542-x
- Wei, C., Yang, H., Wang, S., Zhao, J., Liu, C., Gao, L., et al. (2018). Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality. *Proc. Natl. Acad. Sci.* 115, E4151–E4158. doi: 10.1073/pnas.1719622115
- Wu, H. (2019). A high-quality *Actinidia chinensis* (kiwifruit) genome. *Hortic. Res.* 6, 1–9. doi: 10.1038/s41438-019-0202-y
- Xia, E., Tong, W., Hou, Y., An, Y., Chen, L., Wu, Q., et al. (2020a). The reference genome of tea plant and Resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. *Mol. Plant* 13, 1013–1026. doi: 10.1016/j.molp.2020.04.010
- Xia, E.-H., Tong, W., Wu, Q., Wei, S., Zhao, J., Zhang, Z.-Z., et al. (2020b). Tea plant genomics: achievements, challenges and perspectives. *Hortic. Res.* 7:7. doi: 10.1038/s41438-019-0225-4
- Xia, E.-H., Zhang, H.-B., Sheng, J., Li, K., Zhang, Q.-J., Kim, C., et al. (2017). The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Mol. Plant* 10, 866–877. doi: 10.1016/j.molp.2017.04.002
- Xin, Z., Li, X., Li, J., Chen, Z., and Sun, X. (2016). Application of chemical elicitor (Z)-3-hexenol enhances direct and indirect plant defenses against tea geometrid *Ectropis obliqua*. *BioControl* 61, 1–12. doi: 10.1007/s10526-015-9692-1
- Xu, Y., Bi, C., Wu, G., Wei, S., Dai, X., Yin, T., et al. (2016). VGSC: a web-based vector graph toolkit of genome Synteny and collinearity. *Biomed. Res. Int.* 2016:7823429. doi: 10.1155/2016/7823429
- Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, W265–W268. doi: 10.1093/nar/gkm286
- Yang, T. (2017). Molecular identification of spiders preying on *Empoasca vitis* in a tea plantation. *Sci. Rep.* 7, 1–10. doi: 10.1038/s41598-017-07668-w
- Yang, Z., Baldermann, S., and Watanabe, N. (2013). Recent studies of the volatile compounds in tea. *Food Res. Int.* 53, 585–599. doi: 10.1016/j.foodres.2013.02.011
- Yue, C., Chen, Q., Hu, J., Li, C., Luo, L., and Zeng, L. (2022). Genome-wide identification and characterization of GARP transcription factor gene family members reveal their diverse functions in tea plant (*Camellia sinensis*). Pdf. *Front. Plant Sci.* 13, 1–21. doi: 10.3389/fpls.2022.947072

- Zeng, L., Watanabe, N., and Yang, Z. (2019). Understanding the biosyntheses and stress response mechanisms of aroma compounds in tea (*Camellia sinensis*) to safely and effectively improve tea aroma. *Crit. Rev. Food Sci. Nutr.* 59, 2321–2334. doi: 10.1080/10408398.2018.1506907
- Zhang, X., Chen, S., Shi, L., Gong, D., Zhang, S., Zhao, Q., et al. (2021a). Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant *Camellia sinensis*. *Nat. Genet.* 53, 1250–1259. doi: 10.1038/s41588-021-00895-y
- Zhang, Q., Li, T., Gao, M., Ye, M., Lin, M., Wu, D., et al. (2022). Transcriptome and Metabolome profiling reveal the resistance mechanisms of Rice against Brown Planthopper. *Int. J. Mol. Sci.* 23:4083. doi: 10.3390/ijms23084083
- Zhang, Q.-J., Li, W., Li, K., Nan, H., Shi, C., Zhang, Y., et al. (2020a). The chromosome-level reference genome of tea tree unveils recent bursts of non-autonomous LTR Retrotransposons in driving genome size evolution. *Mol. Plant* 13, 935–938. doi: 10.1016/j.molp.2020.04.009
- Zhang, Z., Zhang, X., Bi, K., He, Y., Yan, W., Yang, C. S., et al. (2021b). Potential protective mechanisms of green tea polyphenol EGCG against COVID-19. *Trends Food Sci. Technol.* 114, 11–24. doi: 10.1016/j.tifs.2021.05.023
- Zhang, W., Zhang, Y., Qiu, H., Guo, Y., Wan, H., Zhang, X., et al. (2020b). Genome assembly of wild tea tree DASZ reveals pedigree and selection history of tea varieties. *Nat. Commun.* 11:3719. doi: 10.1038/s41467-020-17498-6
- Zhao, X., Chen, S., Wang, S., Shan, W., Wang, X., Lin, Y., et al. (2020). Defensive responses of tea plants (*Camellia sinensis*) against tea green leafhopper attack: a multi-Omics study. *Front. Plant Sci.* 10:1705. doi: 10.3389/fpls.2019.01705
- Zhou, Y., Liu, X., and Yang, Z. (2019). Characterization of Terpene synthase from tea green leafhopper being involved in formation of Geraniol in tea (*Camellia sinensis*) leaves and potential effect of Geraniol on insect-derived Endobacteria. *Biomol. Ther.* 9:808. doi: 10.3390/biom9120808
- Zwaenepoel, A. (2019). Wgd—simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* 35, 2153–2155. doi: 10.1093/bioinformatics/bty915