Full Length Article

Comparative Transcriptomic Analyses of Chlorogenic Acid Biosynthesis Pathways in Diploid and Triploid *Camellia sinensis*

Chao Shen[1], Xinzhuan Yao[1], Degang Zhao[2] and Litang Lu[1,2*]

[1]*College of Tea Science, Guizhou University, Guiyang, Guizhou, People's Republic of China*
[2]*The Key Laboratory of Plant Resources Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), Institute of Agro-Bioengineering, Guiyang, Guizhou, People's Republic of China*
[*]For correspondence: ltlv@gzu.edu.cn

Abstract

Chlorogenic acid (CGA), as a kind of depside in plants, has a variety of beneficial effects on human health, which also plays an important role in helping plants resist a variety of stresses. Therefore, the biosynthetic pathway of CGA has been studied in many plants, however, the synthesis of CGA has not been well elucidated in *Camellia sinensis*. In our research, different CGA levels were detected between triploid tea variety 'Qianfu 4' and diplont tea variety 'Qianmei 419' using HPLC and the CGA content in triploid *Camellia sinensis* was greater than that in diploid *Camellia sinensis*. Transcriptome sequencing for diploid and triploid *Camellia sinensis* was employed to explore genes associated with CGA biosynthesis. Finally, 154,097 unigenes were obtained in total, of which 891 may be related to the biosynthesis of CGA. Furthermore, differentially expressed genes (DEGs) were screened between diploid and triploid *Camellia sinensis*, 32 DEGs were discovered to be related to CGA biosynthesis, including sixteen phenylalanine ammonia lyase (PAL) genes, three 4-coumarate coenzyme A ligase (4CL) genes, nine cinnamate 4-Hydroxylase (C4H) genes, four Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (HQT/HCT), and two hundred and twenty-one TFs including eighty-eight ERFs, forty-one bZIPs, forty-two MYBs and fifty WRKYs, which may also play an important role in the biosynthesis of CGA. Our results will lay the foundation for further exploration of the biosynthesis of CGA and revealing the related regulatory network in *Camellia sinensis*. © 2021 Friends Science Publishers

Introduction

Camellia sinensis belongs to the genus *Camellia*, Theaceae, Theales, Dicotyledoneae, originated in China. All tea trees in the world belong to the basic species *Camellia sinensis (l.) o. Kuntze*, which is divided into two subspecies. *C. sinensis var. sinensis* (China tea) is widely cultivated in China, Japan and Taiwan, while *C. sinensis var. assamica* (Assam tea) is mainly distributed in south and southeast Asia (Chan *et al*. 2007; Wan and Xia 2015). Tea is the second most commonly drank liquid on earth preceded only by water (Sharangi 2009). Tea leaves are rich in a large amount of secondary metabolites such as phenylpropanoids, theanine, and volatile oils, which are beneficial for human health (Wheeler and Wheeler 2004). As a health beverage, drinking tea can help to fight against variable forms of cancer, reduce risk of cardiovascular diseases and prevent diabetes (Beltz *et al.* 2006; Iso *et al.* 2006; Nagao *et al.* 2007; Leong *et al.* 2008). CGA is an important component of phenylpropanoids. Furthermore, the cultivation of tea

trees has high economic value. Therefore, *Camellia sinensis* has extremely high research value and application prospects.

CGA, an important organic acid generated by plant secondary metabolism, exists widely in various plants (Bradfield and Flood 1952). It is a phenylpropanoid compound produced by caffeic acid and quicic acid utilizing the shikimic acid pathway in plants' aerobic respiration (Deng *et al.* 2005). Since the first discovery of CGA in sunflower seeds in 1897, its physiological role has been attracting much attention (Osborne and Campbell 1897). Pharmacological research has demonstrated that CGA exhibits many effects for health including lowering blood pressure, anti-tumor and antiviral activities and it also has antidiabetic, anti-obesity, anti-inflammatory, and antimicrobial effects (Sawa *et al.* 1999; Kozuma *et al.* 2005; Huang *et al.* 2015; Naveed *et al.* 2018). Furthermore, CGA showed stronger free radical scavenging activity than ascorbic acid (Pavlica and Gebhardt 2010; Pellati *et al.* 2004). In addition, CGA also plays an important role in stress resistance for plants themselves. It can act as an

antioxidant in plants. Experimental studies have shown that by inhibiting the accumulation of CGA in tobacco, the death of mature leaf cells can be accelerated, mainly due to the increase in the level of oxidized lipid malondialdehyde caused by the oxidative stress, which indicates that CGA can protect against lipid peroxidation (Tamagnone *et al.* 1998). Besides, in transgenic tomatoes, elevated levels of CGA have been shown to reduce UV damage and enhance resistance to microorganisms (Niggeweg *et al.* 2004; Clé *et al.* 2008). Furthermore, CGA is also used as a precursor substance for synthesizing G- and S-type lignin to help plant cells resist environmental stress (Abdulrazzak *et al.* 2005). Realizing that CGA is an important secondary metabolite playing an important role in both plants and human encourages us to explore the biosynthetic pathway continually.

The biosynthetic mechanism of CGA in many plants has been reported. Based on previous reports, there are three different biosynthetic pathways hypothesized for CGA generation. Phenylalanine ammonia-lyase (PAL), Trans-cinnamate 4-monooxygenase (C4H) and 4-Coumarate-CoA ligase (4CL) are key enzymes involved in the first three steps of each CGA biosynthesis pathway. First, Hydroxycinnamoyl-CoA quinate transferase (HQT) is found to catalyze the production of CGA by caffeic acid coenzyme A and quinic acid in Solanaceae, potatoes, tomatoes, and tobacco (Rhodes *et al.* 1979). Second, Hydroxycinnamoyl D-glucose: quinate hydroxycinnamoyl transferase (HCGQT) is found to catalyze the production of CGA by coumaroyl quinic acid and caffeoyl-D-glucose through isotope tracer method (Villegas and Kojima 1986). Third, hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase (HCT) mediates the synthesis of CGA from coumaroyl quinic acid (Rhodes and Wooltorton 1976). So, it's easy to hypothesize that CGA biosynthesis in different species has some differences. On the CGA biosynthetic pathways, PAL is the first enzyme converting phenylalanine to trans-cinnamic acid. As an entry point enzyme, it plays a vital role in controlling carbon fluxes from primary metabolism to downstream branches of secondary metabolism (Bate *et al.* 1994). Next are C4H and 4CL, which catalyze formation of p-coumaroyl-CoA, as a key node in the phenylpropanoid pathway, it is a precursor substance that forms many important compounds, such as lignin, stilbenes and flavonoids/isoflavonoids (Dixon *et al.* 2002), but the detailed mechanism of the role of these enzymes requires further investigation (Sonnante *et al.* 2010). Silenced or overexpressing HQT in tomato and RNAi suppression of HQT in potato have been utilized to confirm that HQT play a role of rate-limiting enzyme in CGA metabolic pathway (Payyavula *et al.* 2015). Zhang *et al.* (2018) utilized GWAS combined with eQTL analyses to find a significant correlation between PtHCT2 and CGA in poplar trees. Meanwhile, it was reported some transcription factors (MYB, ZIP, WRKY and ERF) are involved in the regulation of CGA biosynthesis (Ye *et al.* 2019). In

Lonicera japonica, LjbZIP8 was found as a transcription inhibitor that can specifically bind to the G-box element of LjPAL2 5 '-UTR, and reduce the content of CGA when it is overexpressed in tobacco (Zha *et al.* 2017). In carrots, DcMYB1, DcMYBs 3 and DcMYBs 5 were all identified as important transcriptional activator, among which DcMYB1 acted the promoter of DcPAL1, and DcMYBs 3 and DcMYBs 5 act on the promoter of DcPAL3 (Maeda *et al.* 2005; Wako *et al.* 2010). Through the study of *arabidopsis* mutants losing function of the WRKY TFs, WRKY family was discovered to regulate the biosynthesis of many phenolic substances and some WRKY TFs were also found to act on PtHCT2 in poplar (Wang *et al.* 2010; Zhang *et al.* 2018). In carrot protoplasts, transient expression analysis was to reveal that DcERF1 could activate the expression of DcPAL3 by binding to its promoter's GCC-box homolog (Kimura *et al.* 2008). So far, the research on CGA in *Camellia sinensis* is still very limited, mainly focused on analysis and detection methods, while there are few studies on biosynthesis. So, the molecular mechanism of CGA synthesis in *Camellia sinensis* is still unclear.

Previous studies have indicated that polyploid plants have more obvious advantages than diploid plants, which are mainly reflected in the morphology and physiology of plants. With the doubling of chromosome composition in the nucleus, a series of changes occur, especially in the improvement of growth rate, genetic gain and metabolic yield (Comai 2005). As described previously (Berkov and Philipov 2002), the content of alkaloids in triploid mandala was 4 times higher than diploid mandala. Xu *et al.* (2014) found that the induction of polyploidy in *E. purpurea* resulted in higher PAL and C_4H expression and promoted the biosynthesis of cichoric acid (Xu *et al.* 2014). Therefore, a triploid tea plant and a diploid tea plant were selected as experimental materials in present experiment to study the key genes of CGA biosynthesis in *Camellia sinensis*. Liu *et al.* (2021) utilized one low-CGA content and one high-CGA content in sweet potato to dissect the mechanisms of CGA biosynthesis by transcriptome data, and found that expression of HCT and other genes were significantly different in the two phenotypes, and therefore inferred that those genes may be the key gene that controlled the synthesis of sweet potato CGA (Liu *et al.* 2021). Wang *et al.* (2020) used the transcriptome to study the CGA biosynthesis of diploid and tetraploid *Lonicera japonica*, they found that changes in the expression level of some key genes in the biosynthesis pathway at different growth stages led to changes in the content of CGA (Wang *et al.* 2020). Therefore, we suspected that there may be some key regulatory genes in tea plants. After polyploidization, the expression levels of those genes may get changed, which produce different CGA content between diploid and triploid tea plants. Through deciphering of the key genes causing the difference of CGA content between triploid and diploid *Camellia sinensis*, it is helpful to reveal the molecular mechanism of CGA biosynthesis in *Camellia sinensis*, and

lay the foundation for the cultivation of new tea germplasm with strong stress resistance and good quality in the future.

Materials and Methods

Plant materials

Leaves of two tea varieties 'Qianfu 4' and 'Qianmei 419' were collected from Guizhou tea institute in Meitan, Guizhou, China. 'Qianfu 4', a new triploid tea variety, was bred from an asexual propagation after systematic selection of a diploid tea variety 'Qianmei 419' seeds treated by ^{60}Co-γRay (Chen *et al.* 2009). Diploid and triploid tea plants were grown in yellow loam soil (pH=5.01). Both of them grew well with sufficient light and water supply in the nursery of the institute. The second leaves from the new young branches of the tea plants were collected on July 25th, 2019. Three biological replicates were taken as experimental samples for both varieties, some of the samples were quick-frozen with liquid nitrogen and then stored at -80°C for subsequent transcriptome sequencing, and the other part was stored at 4°C for CGA content determination.

Determination of CGA content

The fresh leaves of the tea tree to be tested were pulverized, and 1 g of the pulverized uniform leaves was weighed, then poured to a 25 mL volumetric flask, 20 mL methanol (70%, v/v) was added. Ultrasonic extraction was performed for 30 min, (power 40 kHz, 30°C) and then the extract was diluted with methanol (70%, v/v) to 25 mL and well-mixed, filtered through a 0.45 μm microporous filter membrane and analyzed by HPLC for CGA content.

A Shimadzu LC-20AT HPLC analyzer (Shimadzu Corporation., Kyoto, Japan) was used to measure CGA content, equipped with a CTO-10AS VP column oven, SIL-20A autosampler, LC-20AT pump and SPD-20A ultraviolet spectroscopy detector. Separation was carried out using an Agela Technologies Venusil XBP-C18 column (4.6 × 250 mm, 5 μm); the temperature of the column and detector was set to 35°C. The mobile phase consisted of acetonitrile and an aqueous solution of acetic acid 0.5% (v/v), with a flow rate of 1.0 mL/min. The mobile phase ratio was set to 10% acetonitrile and 90% aqueous solution of acetic acid 0.5% (v/v). The injection volume was set at 20 μL and the detection wavelength was 327nm. CGA standard (99.7%) was purchased from ANPEL Laboratory Technology (Shanghai) Inc (Shanghai, China). According to retention time of the CGA standard, CGA of sample was determined and then the content of CGA was calculated according to the peak area. SPSS software was used to data processing.

Transcriptome sequencing

Firstly, the RNeasy Plus Mini kit (Qiagen, Valencia, CA, USA) was employed to extracted total RNA from leaf of diploid and triploid *Camellia sinensis*. After treated with DNase I, mRNA was enriched by magnetic beads with Oligo (dT) through the OligoTex mRNA mini kit (Qiagen, Germany). The mRNA was then fragmented adding a proprietary fragmentation reagent. With the interrupted mRNA as the template, the first strand cDNA was synthesized by using random hexamers, and then the second strand cDNA was synthesized by adding buffer, dNTPs and DNA polymerase I. After purification and recovery of the kit, repair of the sticky end, addition of base' A' at the 3' end and ligation of the sequencing adapters, the obtained fragments were subjected to size selection and PCR amplification enrichment. The constructed library was qualified by Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA) and ABI StepOnePlus Real-Time PCR System (Applied Biosystems, USA) and then sequenced using Illumina HiSeq™ 2000 platform (Illumina Inc, San Diego, CA, USA). There were three biological replicates for RNA-Seq.

De novo transcriptome assembly

In order to ensure the reliability of subsequent analysis results, the original sequencing sequence obtained by sequencing must be filtered through in-house perl script to obtain clean reads. The filtration conditions are as follows: (1) Remove sequences containing adaptors; (2) Remove sequences with N base ratio over 5%; (3) Remove the sequences containing more than 50% of the base mass value less than 10. Transcripts were obtained from the clean reads assembled by the Trinity software (Grabherr *et al.* 2011). Then, Tgicl (Pertea *et al.* 2003) was used to remove redundancy and further splicing for transcripts to obtain the final unigene, which was divided into two parts, one was cluster and in the same cluster there were several unigenes with high similarity (more than 70%) (beginning with CL, followed by the gene family number) and the other were Singletons (beginning with Unigene), representing individual unigene.

Annotation and classification

To functionally annotate genes, the unigene was aligned to sequences available in public databases through BLAST (significance threshold E value ≤ 10^{-5}) (Altschul *et al.* 1990), including Nt (NCBI non-redundant nucleotide sequences); NR (NCBI non-redundant protein sequences) (Deng *et al.* 2006); GO (Gene Ontology) (Ashburner *et al.* 2000); COG(Clusters of Orthologous Groups) (Tatusov *et al.* 2003); KEGG(Kyoto Encyclopedia of Genes and Genomes) (Minoru *et al.* 2004); Swiss-prot (A manually annotated and reviewed protein sequence database) (Apweiler *et al.* 2004). GO annotation of unigenes was obtained by Blast2GO software (v. 2.5.0) combined NR annotation results under the categories of biological function, molecular function and cellular components (Conesa *et al.* 2005; Aparicio *et al.* 2006). Then, WEGO (Web Gene Ontology Annotation Plot) software (Ye 2006)

was implemented to perform GO function classification statistics for all unigene to understand the gene function distribution characteristics of this species from a macroscopic view.

DEG profiling analysis

Total RNA of diploid and triploid *Camellia sinensis* was extracted to establish two libraries for RNA-Seq analysis. RSEM was employed to calculate gene expression levels (Li and Dewey 2011). After the aligning the clean reads to the de novo assembled transcriptome, the unigene expression was calculated utilizing the technique of FPKM (fragment per kilobase per million mapped reads). After that, DESeq2 (Ashburner *et al.* 2000; Love *et al.* 2014) was implemented to analysis differential expression of diploid and triploid *Camellia sinensis*. In our research, the FPKM values between two tea varieties of all unigenes were compared by a threshold of FDR < 0.001 and |log2ratio| > 1 to observe differential gene expression. GO and KEGG enrichment analyses were performed to investigate the functional distribution of differentially expressed genes (DEGs) for exploring DEGs involved in CGA synthesis.

qRT-PCR validation of of RNA-seq

To verify the accuracy of RNA-Seq results, ten DEGs were randomly selected for qRT-PCR analysis. Total RNA was extracted from leaf samples of diploid and triploid *Camellia sinensis* with Quick RNA isolation Kit (Waryong, Beijing, China), then the first-strand cDNA was synthesized by First Strand cDNA Synthesis Kit (with gDNA Eraser) (GENENODE, Beijing, China). qRT-PCR reactions utilized for gene expression analysis were conducted in a Bio Rad CFX ConnectTM Real-Time System (Bio-Rad Laboratories, Hercules, California, USA) by using ChamQ Universal SYBR qPCR Master Mix (Vazyme, Nanjing, China) following the instructions. Primers for each gene were designed by Primer Premier 5.0 software (Table S1).

Results

Determination of CGA content

In Total, 6 samples (three biological replicates for two tea varieties) were subjected to CGA content of leaves analyses based on HPLC. The results indicate that the content of CGA in triploid *Camellia sinensis* was higher than diploid *Camellia sinensis* (*P < 0.05*) (Fig. 1).

Transcriptome Profiling of *Camellia sinensis*

Diploid and triploid *Camellia sinensis* samples contained different concentrations of CGA detected by HPLC. In order to elucidate the molecular mechanisms and discover the key genes regulating CGA biosynthesis in *Camellia*
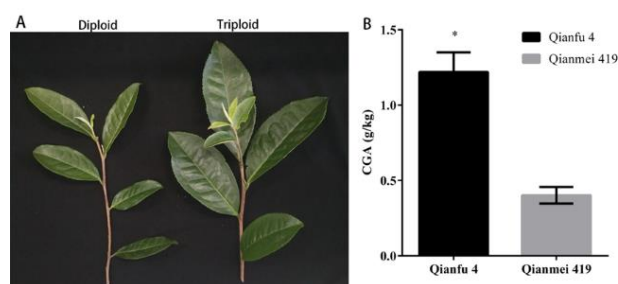
Fig. 1: The leaves used for transcriptome sequencing and chlorogenic acid (CGA) contents of Camellia sinensis. **A**: The leaves of *Camellia sinensis*. **B**: The CGA contents of 'Qianfu 4' and 'Qianmei 419'. Three individual experimental were performed for each varieties, and significant differences were analysed using student's t test, different alphabet indicated *P* value *< 0.05*

sinensis, transcriptome sequencing for two tea varieties (three replicates per sample) was employed to explore which key gene was responsible for the difference. A total number of 318,515,326 raw reads were got then through discarding the low-quality raw reads, 317,609,528 clean reads were obtained, which have an average length of 150 bp (Table S2). Through de novo transcriptome assembly by the Trinity software and remove the reluctances by TGICL software, we obtained 154,097 non-redundant unigenes achieved with a N50 of 1349 bp and an average length of 825.67 bp. The distribution of unigene lengths is presented in (Fig. S1).

Functional annotation and categorization

In order to get more accurate gene function annotations as much as possible, Nr, Nt, Swiss-Prot, KEGG, COG and GO databases were employed to annotate gene function of all the 154,097 unigenes based on sequence similarity. 103,448 (67.13%) unigenes were annotated in at least one database-based BLAST search. The annotated results of unigenes in six databases were listed in Table 1 based on significant similarity to sequences in the database.

Gene Ontology (GO) is used to describe characterization of genes and products in organisms (Ashburner *et al.* 2000). According to result of Nr annotation (90,547 unigenes) Blast2GO software combined with WEGO software were utilized to perform GO function classification statistics for 45,820 unigenes, they were distributed in 'biological process', 'cellular component', and 'molecular function' belonging to the three main categories of GO and were further divided into 55 sub-categories. 'Biological process' is the ontology that contained the most unigenes (98,552), followed by 'cellular component' (91,167 unigenes), 'molecular function' containing 51,707 unigenes ranked the lowest. The most highly represented 'biological process' subcategories were 'metabolic process' (24,750 unigenes, 48.49%). The sub-category of 'cell' (18,015 unigenes, 35.29%) ranked first in 'cellular

Table 1: Summary for the annotation of unigenes

	Total Unigene	Annotated	Percent (%)
Nt	154,097	89,933	58.36%
Nr	154,097	90,547	58.76%
COG	154,097	35,298	22.91%
SWISSPROT	154,097	61,318	39.79%
KEGG	154,097	67,980	44.12%
GO	154,097	45,820	29.73%
All annotated unigenes	154,097	103,448	67.13%

Table 2: The unigenes related to other secondary metabolites

Biosynthesis of other secondary metabolites	Unigene numbers	Pathway ID
Carbapenem biosynthesis	4	ko00332
Betalain biosynthesis	5	ko00965
Acridone alkaloid biosynthesis	11	ko01058
Anthocyanin biosynthesis	18	ko00942
Caffeine metabolism	24	ko00232
Benzoxazinoid biosynthesis	35	ko00402
Indole alkaloid biosynthesis	38	ko00901
Isoflavonoid biosynthesis	59	ko00943
Glucosinolate biosynthesis	61	ko00966
Monobactam biosynthesis	63	ko00261
Flavone and flavonol biosynthesis	128	ko00944
Tropane, piperidine and pyridine alkaloid biosynthesis	152	ko00960
Isoquinoline alkaloid biosynthesis	171	ko00950
Stilbenoid, diarylheptanoid and gingerol biosynthesis	203	ko00945
Flavonoid biosynthesis	288	ko00941
Phenylpropanoid biosynthesis	764	ko00940

Table 3: The numbers of unigenes involved in CGA biosynthesis

Gene	Enzyme No.	Numbers
Functional gene		75
Phenylalanine ammonia-lyase (PAL)	EC:4.3.1.24	28
4-Coumarate-CoA ligase (4CL)	EC:6.2.1.12	28
Trans-cinnamate 4-monooxygenase (C4H)	EC:1.14.13.11	8
Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase (HQT/HCT)	EC:2.3.1.133	8
Coumaroylquinate (coumaroylshikimate) 3′- monooxygenase (C3H)	EC:1.14.13.36	3
Transcription factors		816
MYB transcription factor	–	207
ZIP transcription factor	–	150
WRKY transcription factor	–	193
Ethylene response factor	–	266

component' category while catalytic activity (22,501 unigenes, 44.08%) was the dominant 'molecular_function' sub-categories. (Fig. S2).

COG database is constructed based on the classification of 21 complete genomes of bacteria, algae and eukaryotes uploaded to NCBI database, which is employed to make prediction for possible functions of unigene and carry on the statistical analysis of function. Through the prediction of COG functional classification of unigenes, a total of 35,298 (22.91%) unigenes were annotated with 25 COG classifications. The 'general function prediction only' (9286, 26.31%) was largest one, followed by 'transcription' (5108, 14.47%), 'Posttranslational modification, protein turnover, chaperones' (4438, 12.58%), 'Translation, ribosomal structure and biogenesis' (4199, 11.90%).

'Extracellular structures' and 'Nuclear structure' had the fewest unigenes (Fig. S3).

KEGG is a database that integrates genomic, chemical, and systemic functional information to facilitate the study of metabolic pathways of some metabolites. According to the KEGG annotation information, the pathway annotation of unigene can be further obtained. To further explore the function of genes, the unigenes were compared to a KEGG database. A total of 67,890 unigenes were assigned to136 different pathways in KEGG database. There were 11,956 (17.59%) unigenes involving in metabolic pathways, of which 6236 (9.17%) unigenes relating to secondary metabolism and 2024 (2.98%) unigenes were classified to biosynthesis of other secondary metabolites (including 16 pathways) (Table 2). These pathways provided a foundation for investigating biosynthesis of CGA, among them, the pathway ranked the first in other secondary metabolism was phenylpropanoid biosynthesis (ID: ko00940) (764, 37.75%), followed by Flavonoid biosynthesis (ID: ko00941) (288 unigenes, 14.23%) and Stilbenoid, diarylheptanoid and gingerol biosynthesis (ID: ko00945) (203 unigenes, 10.03%). CGA is a product of the phenylpropanoid biosynthesis pathway of plant secondary metabolism. Based on the analysis of transcriptome data and the existing literature reports, we screened out a total of 75 genes encoding 6 enzymes in the CGA synthesis pathway and 816 transcription factors that may regulate enzymes of CGA biosynthesis pathway in *Camellia sinensis* (Table 3).
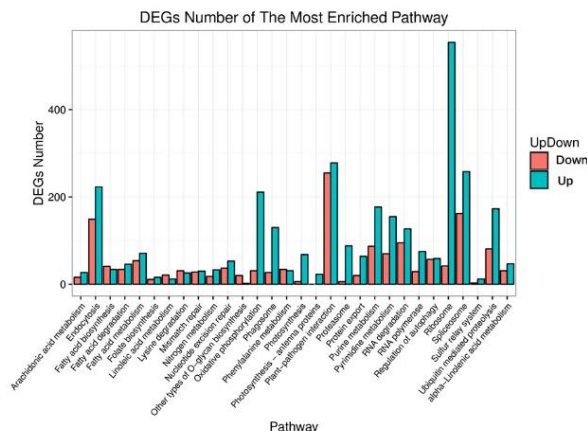
Functional analysis of DEGs

In order to screen the candidate unigenes that play an important role in the production of different concentrations of CGA between diploid and triploid *Camellia sinensis*, DEGs in the transcriptome data of two tea varieties was analyzed. FDR < 0.001 and |log2ratio| > 1 was set as the threshold value for screening DEGs, a total of 23,813 DEGs were obtained in the triploid tea leaves compared with the diploid tea leaves, including 16,459 up-regulated unigenes and 7,354 down-regulated unigenes. GO enrichment analysis was performed for the DEGs in the two groups. In total, 52 GO terms were enriched with 10,744 DEGs, including 21 biological process terms, 16 cellular component terms, and 15 molecular function terms. The main 'biological processes' terms were 'metabolic process' (1765) and 'cellular process' (1754). 'binding' (1758) and 'catalytic activity' (1757) were major parts of the 'molecular function category' and 'cell part' (1755) and 'cell' (1755) were the most popular frequent terms in the 'cellular component' functional group (Fig. 2).

To further characterize the DEGs, KEGG enrichment analysis was performed to analyze functional genes in the biosynthesis pathways associated with CGA synthesis in *Camellia sinensis*. As a result, 15,133 DEGs were mapped to 19 KEGG pathways iterm (Level 2), respectively (Fig. 3). Metabolism was the group containing the most DEGs
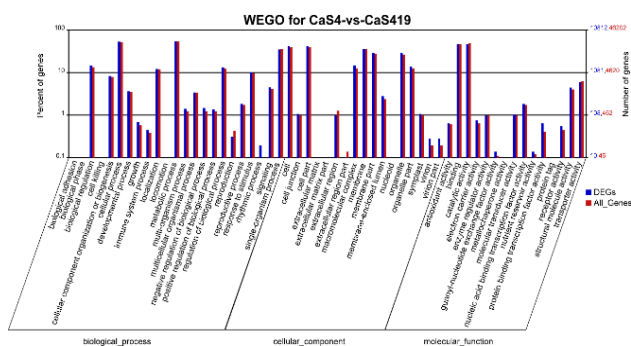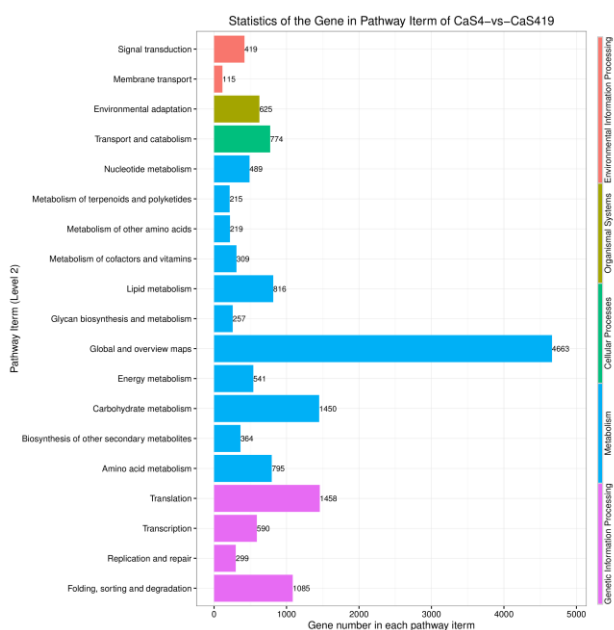
Fig. 2: WEGO results of differentially expressed genes (DEGs) from Camellia sinensis. The x-coordinate represents the three aspects of GO (cellular component, molecular function, and biological process). The left y-axis indicates the percentage of genes in a category, and the right y-axis indicates the number of genes in a category. The red number represents the background (all genes) and the blue number represents the differential genes

Fig. 3: KEGG pathway analysis of DEGs in 'Qianfu 4'-*vs*.- 'Qianmei 419'. DEGs were distributed in 5 functional categories (cellular processes, environmental information processing, genetic information processing, metabolism and organismal systems)

(10,118 DEGs). Then, the top 30 the most enriched KEGG pathways are shown in Fig. 4, including numbers of up and down-regulated unigenes of each pathways. 'Ribosome' (596 DEGs, ko03010) were the most popular pathways, next is 'Plant-pathogen interaction' (533 DEGs, ko04626). CGA biosynthesis belongs to 'Biosynthesis of other secondary metabolites' containing 364 DEGs and 'Phenylpropanoid biosynthesis' (135 DEGs, ko00940) was the top one enriched pathways of this category. Although there were 8,680 DEGs which were not functionally annotated, some of them may get involved in CGA biosynthesis progress.

Fig. 4: Enrichment of Pathway differentially expressed genes up and down-regulated histogram

Further research is needed to explore it.

Identifying *Camellia sinensis* CGA synthesis-associated unigenes

CGA is a phenylpropanoid originated from phenylalanine during plants' aerobic respiration. According to the information obtained from the transcriptome, based on the annotation information of the major gene databases for unigenes, several gene sequences encoding the enzymes on the CGA synthesis pathway were obtained. Here, 75 unigenes encoding these enzymes were obtained, including twenty-eight PAL, twenty-eight 4CL, eight C4H, eight HCT/HQT and three C3H unigenes in *Camellia sinensis* (Fig. 5). However, gene sequences encoding two key enzymes of one of CGA synthesis pathway were not found (HCGQT, UDP glucose: cinnamate glucosyl transferase). In addition, Transcription factors are key regulatory protein of biosynthesis of secondary metabolites, which mediate the transcriptional regulation. On the basis of previous reports, MYB family, ZIP family, WRKY family and ERF family were regard to control the expression of genes relating to CGA biosynthesis. As shown in Table 3, a total of 207, 150, 193 and 266 unigenes were identified to code MYB, ZIP, WRKY, and ERF in this study, respectively. The expression levels of these unigenes and TFs above may affect CGA content. Hence, the expression profiles of these genes and related TFs still need further research to verify their function (Fig. 5, S4).

qRT-PCR validation of DEGs

In order to verify the accuracy of transcriptome sequencing, we selected ten DEGs to detect their relative expression levels by qRT-PCR. The results showed that there were indeed differences between the two varieties and were consistent with the results obtained by transcriptome

sequencing. Although the actual measured data was different from the expression level of the transcriptome, the expression trend was consistent with transcriptome (Fig. 6). Therefore, the data obtained from transcriptome sequencing in this research can be utilized to further investigation on genes related to CGA biosynthesis.

Discussion

Tea is so popular because it contains a large amount of polyphenols, vitamins, theanine, flavonoids and other biologically active substances that are beneficial to the human body, and consumed everyday by millions of people. CGA, as a component of tea, is also beneficial for human health. However, researches on CGA of tea are very limited, especially in biosynthesis parts. Candidate genes for the biosynthesis of flavonoid, theanine and caffeine have been revealed by transcriptomic sequencing of *Camellia sinensis* (Shi *et al.* 2011). Nevertheless, transcriptome sequencing has not been reported to reveal the biosynthesis and metabolic pathways of CGA in *Camellia sinensis*. At present, there are gene studies carried out in *Camellia sinensis* to identify genes related to flower development (Liu *et al.* 2017), blister blight defense (Jayaswall *et al.* 2016) and chilling and freezing responsive (Zheng *et al.* 2015) by high-throughput sequencing technology. Therefore, the Illumina HiSeq™ 2000 platform was employed to digital gene expression (DGE) and on this basis, to further explore key pathways for CGA synthesis in *Camellia sinensis*. DGE profiling technology can be used to explore the function of some genes in many biosynthetic pathways owing to its ability in quickly detecting the DEGs among multiple groups of samples (Strickler *et al.* 2012).

It was reported that polyploidy can enhance the yield of secondary metabolites (Dhawan and Lavania 1996). Hence, in our research, a diploid tea variety and a triploid tea variety were selected for research. HPLC revealed that the content of CGA in leaves of triploid tea variety was significantly higher than that of diploid tea variety. The RNA from leaves of two tea varieties was extracted for transcriptome sequencing. 154,097 unigenes were obtained in total. Then, the DEGs of diploid and triploid *Camellia sinensis* leaves were collected from DGE profiling by RNA-Seq analysis. These obtained DEGs are the basis for further exploration of the mechanism of CGA biosynthesis in *Camellia sinensis*.

CGA is an important antioxidant for both plants and animals and widely found in many plant species. It was synthesized via phenylpropanoid pathway which has been revealed in many plants, such as artichoke (Sonnante *et al.* 2010), coffee (Lepelley *et al.* 2007) and *Lonicera japonica* (Zhang *et al.* 2017). To date, CGA biosynthesis is supposed to occur through three main routes in in the published study (He *et al.* 2013). However, only the No. 1 pathway that HQT catalyzes the synthesis of CGA by transesterification of caffeoyl-CoA and quinic acid while No. 3 pathway that

the p-coumarolylquinic acid produced under the catalysis of HCT is then hydroxylated by C_3H to produce CGA are more common in plants. In the second pathway, CGA is formed from caffeoyl D-glucoside and quinic acid with catalysis of HCGQT, which was only studied in sweet potato (Villegas and Kojima 1986). Various important enzymes of two pathways (the first and third) involved in CGA synthesis have been identified in many species including *L. japonica*, coffee, artichoke, tobacco, tomato and so on. Among them, PAL, C4H and 4CL are three consecutively catalyzed enzymes shared upstream of CGA synthesis in the study based on studies of *L. japonica*, coffee, artichoke and so on (Chen *et al.* 2015). Singh *et al.* (2008) have reported that PAL and C4H mainly regulate the synthesis of phenylpropanoid and flavonoid in tea. PAL catalyzing the conversion of phenylalanine to cinnamate serves as the entry point from primary metabolism into secondary metabolism, which is the dominant control point of carbon fluxes in phenylpropanoid pathway. Overexpression of PAL gene in tobacco significantly increasing the content of CGA suggests that it plays an important regulatory role in the biosynthesis of CGA (Shadle *et al.* 2003; Chang *et al.* 2009). PAL is encoded by a polygenic family, for example, PtrPAL gene family consists of five members in poplar, PAL family of cucumber has 7 members and tomato has number of PAL up to 26 (Tsai *et al.* 2006; Chang *et al.* 2008; Shang *et al.* 2012). Wu *et al.* (2017) have screened out six PAL genes from *Camellia sinensis* based on the sequencing data available from eight transcriptome projects (Wu *et al.* 2017). In present study, we have found 28 genes coding PAL, among them, the expression level of 4 PAL genes (Unigene5424_All, CL182.Contig1_All, CL182.Contig30_All, Unigene66592_All) corresponded well with CGA content. Yu *et al.* (2021) used overexpression of IbPAL1 to increase the content of CGA in sweet potato, which also confirmed the regulation effect of PAL on the synthesis of CGA (Yu *et al.* 2021). C4H is the second enzyme of phenylpropanoid pathway after PAL, converting cinnamic acid to p-coumaric acid. There was 1 of 8 genes coding C4H found to have the expression level consisting with CGA content (CL11402.Contig1_All). Xu *et al.* (2014) also found that induction of polyploidy in *E. purpurea* resulted in higher PAL and C4H expression (Xu *et al.* 2014). Then, the next step is the conversion of p-coumaric acid to p-coumaroyl-CoA catalyzed by 4CL. When Shi *et al.* (2011) studied candidate genes for flavonoids synthesis through deep sequencing of the *Camellia sinensis* transcriptome, 22 unigenes coding 4CL were obtained. Similarity, 28 unigenes were found in this study, and 3 of them were positively correlated with the CGA content (Unigene40991_All, Unigene44601_All, CL3446.Contig3_All). When Ye *et al.* (2019) employed the transcriptome to study the synthesis of CGA in *Eucommia ulmoides*, they also found that the expression level of 4CL in cultivars containing higher concentrations of CGA was
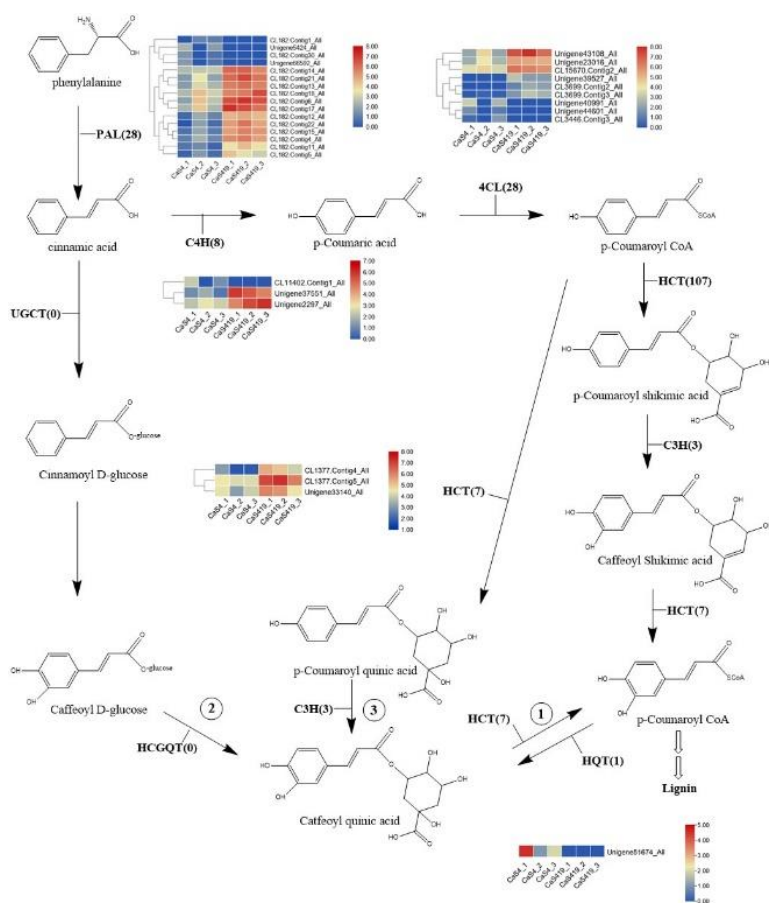
Fig. 5: Proposed pathways for the biosynthesis of chlorogenic acid (CGA) in Camellia sinensis. The three different routes of CGA biosynthesis are labelled 1, 2 & 3. The names and expression patterns of the enzyme are displayed in each step. The expression patterns of each DEGs are shown by heatmap. The color scale represents log2 transformed FPKM values. Blue indicates low expression, and red indicates high expression
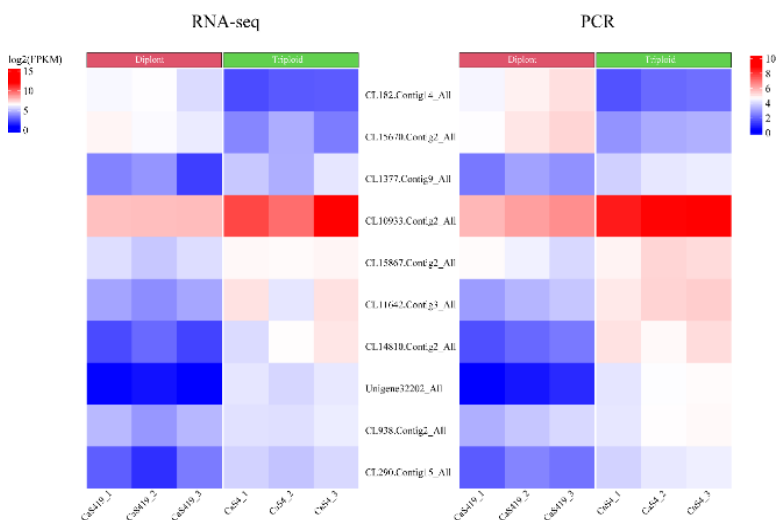
Fig. 6: Validation of RNA-Seq results by qRT-PCR for ten randomly selected *Camellia sinensis* genes. The expression of selected DEGs is shown. A red color indicates high expression for the gene
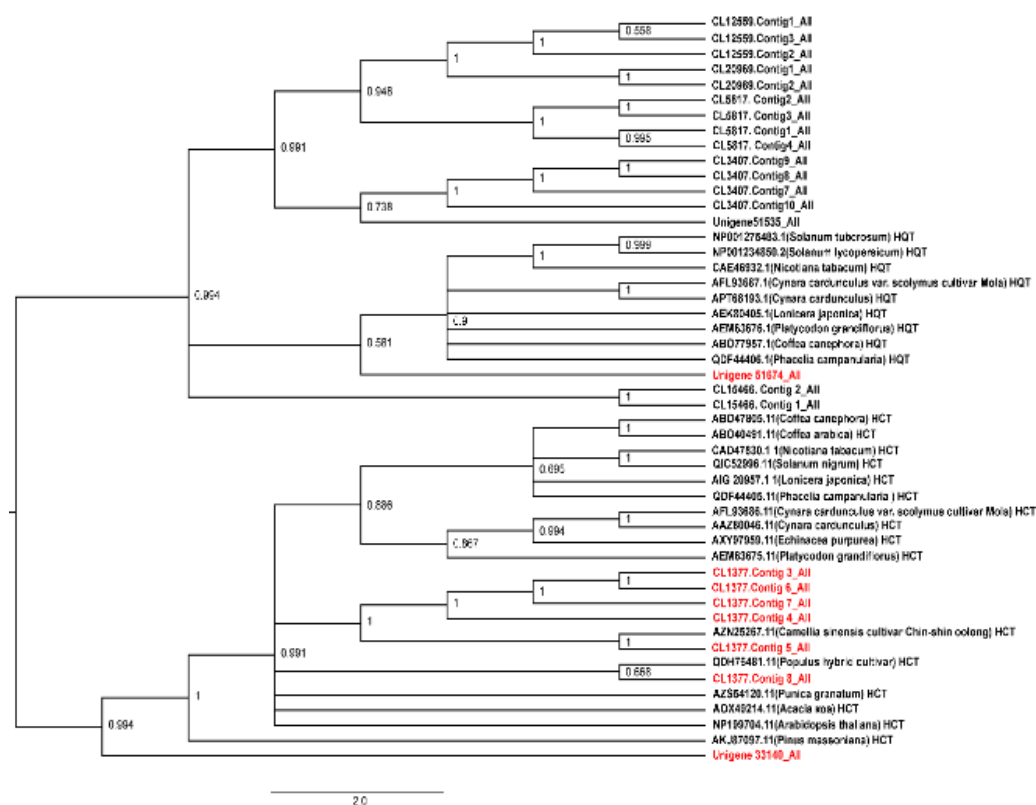
Fig. 7: Phylogenetic tree of HQT/HCT protein sequences available in public databases

higher than that in low-concentration cultivars (Ye *et al.* 2019). Therefore, the above studies indicated that these genes may have a regulatory effect on the synthesis of CGA in *Camellia sinensis*. However, p-coumaroyl-CoA is a key node in the phenylpropanoid pathway, providing immediate precursor shared by many other important compounds like flavonoids, methoxy guaiacyl- and syringyl-monolignols, so those unigenes coding three enzymes (PAL, C4H and 4CL) may not the key factor affecting the content of CGA in *Camellia sinensis*. Further verification was needed to explore the specific relationship between these unigenes and CGA content.

The conversion of p-coumaroyl-CoA to CGA through two pathways including enzymes C3H, HCT and HQT, and C3H plays an important role in both pathways which catalyzes 3-hydroxylation of p-coumaroyl quinic acid and p-coumaroyl shikimic acid. A total of four unigenes encoding C3H were found, but the correlations between gene expressions and CGA content were very weak. HQT and HCT are homologous genes, belong to the BAHD superfamily of acyl - coenzyme A- dependent acyltransferases, with acyl receptor specificity. Using P-Coumaroyl CoA as acyl donor, shikimic acid and quinic acid could be converted to caffeoyl shikimic acid and p-coumaroyl quinic acid respectively by HCT, and HCT also can catalyze the decomposition of CGA to produce caffeoyl-CoA, which has been demonstrated to be key link

of phenylpropanoid and lignin synthesis in *Arabidopsis*. Nevertheless, HQT only works in the synthesis of CGA, and it convert caffeoyl-CoA and quinic acid to CGA. It has also been reported that HQT can catalyze p-coumaroyl CoA and quinic acid to quinate ester. HQT, a key enzyme in the biosynthesis of CGA, has been proved in many plants (Niggeweg *et al.* 2004). Both HCT and HQT proteins contain the two conserved domains (HXXXD and DFGWG) of the acyltransferase family, which are indistinguishable directly (Miguel *et al.* 2020). In order to further study the characteristics of the acyltransferase contained in tea plants, the protein sequences of HCT and HQT contained in previously reported plants have been collected. Using phylogenetic analysis (Fig. 7), it was found that only Unigene51674_All was clustered into one branch with HQT in other species, while the remaining CL1377.Contig3_All, CL1377.Contig6 All, CL 1377.Contig7 All, CL1377.Contig4 All, CL1377.Contig5 All, CL 1377.Contig8 All, Unigene33 140 All was clustered with HCTs of other species. The expression level of Unigene51674_All was higher in triploid *Camellia sinensis*, while the expression level of CL1377.Contig4 All, CL 1377.Contigs All, and Unigene33140 All are higher in diploid *Camellia sinensis*. Overall, CGA synthesis pathway of many enzymes in diploid *Camellia sinensis* expression quantity is higher than in triploid *Camellia* sinensis, while CGA content in the triploid *Camellia sinensis* rather than

diploid *Camellia sinensis*. However, only one potential gene encoding HQT was highly expressed in triploid *Camellia sinensis*. Combined with studies on the model plant *Arabidopsis thaliana*, in which both C3H and HCT were present, but no CGA was detected, indicating that pathway 3 may not be the main synthesis pathway of chlorogenic acid which may be a compensation mechanism for CGA biosynthesis through HCT when plants are under stress (Niggeweg *et al.* 2004; Valiñas *et al.* 2017). By studying the CGA biosynthesis of *L. japonica*, Wang *et al.* (2020) also observed that the HQT expression level in tetraploid *L. japonica* with high CGA content was higher than that in diploid *L. japonica* with low CGA content (Wang *et al.* 2020). Therefore, we speculated that CGA synthesis in tea plants may also be dominated by HQT. At the same time, in the third synthetic route, the last of the CGA synthesis enzyme in the amount of expression in the two varieties did not change, so the speculated that C3H may be another limiting factor in CGA synthesis in *Camellia sinensis*, resulting in low CGA content in diploid *Camellia sinensis* may be the cause of the high expression of HCT would catalyze CGA to produce caffeoyl-CoA, used in the synthesis of lignin. However, how these genes cooperate to control the synthesis of CGA in *Camellia sinensis* needs further verification.

Conclusion

In summary, transcriptional sequencing was performed on tea trees and the genes related to CGA synthesis of *Camellia sinensis* were discovered based on this basis. Transcriptome analysis showed that there were 2024 (2.98%) genes were classified to biosynthesis of other secondary metabolites (including 16 pathways), which provided foundation for the further research of the secondary metabolism mechanism of *Camellia sinensis*. In addition, 764 genes associated with phenylpropane biosynthesis pathways were identified, and a potential key HQT gene that may dominate the synthesis of chlorogenic acid was found in tea plants. Furthermore, two hundred and twenty-one TFs express differences in expression between the two varieties including eighty-eight ERFs, forty-one bZIPs, forty-two MYBs and fifty WRKYs, all of above genes might play important roles in the synthesis of CGA. However, further experiments are needed to decipher the physiological roles of these enzymes.

Acknowledgements

Author Contributions

LL designed the experiment and provided lab facilities. SC and XY performed experimental work. SC wrote this article and DZ helped in revising the article.

Conflicts of Interest

All other authors declare no conflicts of interest.

Data Availability

Data presented in this study are available on fair request to the corresponding author.

Ethics Approval

Not applicable.

References

Abdulrazzak N, B Pollet, J Ehlting, K Larsen, C Asnaghi, S Ronseau, C Proux, M Erhardt, V Seltzer, JP Renou, P Ullmann, M Pauly, C Lapierre, DW Reichhart (2005). A coumaroyl-ester-3-hydroxylase insertion mutant reveals the existence of nonredundant meta-hydroxylation pathways and essential roles for phenolic precursors in cell expansion and plant growth. *Plant Physiol* 140:30–48

Altschul SF, W Gish, W Miller, EW Miller, DJ Lipman (1990). Basic local alignment search tool. *J Mol Biol* 215:403–410

Aparicio, G, S Götz, A Conesa, D Segrelles, M Talon (2006). Blast2GO goes Grid: Developing a Grid-Enabled Prototype for Functional Genomics Analysis. *Stud Health Technol Inform* 120:194–204

Apweiler R, R Bairoch, CH Wu, WC Barker, B Boeckmann, S Ferro, E Gasteiger, HZ Huang, R Lopez, M Magrane, MJ Martin, DA Natale, C O'Donovan, N Redaschi, LSL Yeh (2004). UniProt: The universal protein knowledgebase. *Nucl Acids Res* 32:115–119

Ashburner M, CA Ball, JA Blake, D Botstein, H Butler, JM Cherry, AP Davis, K Dolinski, SS Dwight, JT Eppig, MA Harris, DP Hill, LI Tarver, A Kasarskis, S Lewis, JC Matese, JE Richardson, M Ringwald, GM Rubin, G Sherlock (2000). Gene ontology: Tool for the unification of biology. *Gene* 25:25–29

Bate NJ, J Orr, W Ni, A Meromi, T Nadler-Hassar, PW Doerner, RA Dixon, RA Lamb, Y Elkind (1994). Quantitative relationship between phenylalanine ammonia-lyase levels and phenylpropanoid accumulation in transgenic tobacco identifies a rate-determining step in natural product synthesis. *Proc Natl Acad Sci* 91:7608–7612

Beltz LA, DK Bayer, AL Moss, IM Simet (2006). Mechanisms of cancer prevention by green and black tea polyphenols. *Anti Cancer Agents Med Chem* 6:389–406

Berkov S, S Philipov (2002). Alkaloid production in diploid and autotetraploid plants of *Datura stramonium*. *Pharm Biol* 40:617–621

Bradfield AE, AE Flood (1952). Chlorogenic acids in fruit trees. *Nature* 170:168–169

Chan EWC, YY Lim, YL Chew (2007). Antioxidant activity of *Camellia sinensis* leaves and tea from a lowland plantation in Malaysia. *Food Chem* 102:1214–1222

Chang A, MH Lim, SW Lee, EJ Robb, RN Nazar (2008). Tomato phenylalanine ammonia-lyase gene family, highly redundant but strongly underutilized. *J Biol Chem* 283:33591–33601

Chang J, J Luo, AG He (2009). Regulation of polyphenols accumulation by combined overexpression/silencing key enzymes of phenylpropanoid pathway. *Acta Biochim Biophys Sin* 41:123–130

Chen Z, N Tang, Y You, J Lan, Y Liu, Z Li (2015). Transcriptome analysis reveals the mechanism underlying the production of a high quantity of chlorogenic acid in young leaves of *Lonicera macranthoides* Hand.-Mazz. *PLoS One* 10; Article e0137212

Chen ZW, ZM Liu, JL Wang (2009). Breeding of Qianfu 4, new triploid tea variety. *Southwest Chin J Agric Sci* 22:1194–1197

Clé C, LM Hill, R Niggeweg, CR Martin, Y Guisez, E Prinsen, MAK Jansen (2008). Modulation of chlorogenic acid biosynthesis in *Solanum lycopersicum*; consequences for phenolic accumulation and UV-tolerance. *Phytochemistry* 69:2149–2156

Comai L (2005). The advantages and disadvantages of being polyploid. *Nat Rev Genet* 6:836–846

Conesa A, S Götz, JM Garcia-Gomez, J Terol, M Talon, M Robles (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676

Dhawan OP, UC Lavania (1996). Enhancing the productivity of secondary metabolites via induced polyploidy: A review. *Euphytica* 87:81–89

Deng L, H Yuan, ZY Yu (2005). Recent advances in research on chlorogenic acid. *Chem Bioeng* 7; Article 358

Deng YY, JQ Li, SF Wu, YP Zhu, FC He (2006). Integrated nr database in protein annotation system and its localization. *Comput Eng* 32:71–72

Dixon RA, L Achnine, P Kota, CJ Liu, MSS Reddy, L Wang (2002). The phenylpropanoid pathway and plant defence – a genomics perspective. *Mol Plant Pathol* 3:371–390

Grabherr MG, BJ Haas, M Yassour, JZ Levin, DA Thompson, I Amit, X Adiconis, L Fan, R Raychowdhury, QD Zeng, ZH Chen, E Mauceli, N Hacohen, A Gnirke, N Rhind, FD Palma, BW Birren, C Nusbaum, KL Toh, N Friedman, A Regev (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652

He L, XL Xu, Y Li, CF Li, YJ Zhu, HX Yan, ZY Sun, C Sun, JY Song, Y Bi, J Shen, RY Cheng, ZZ Wang, W Xiao, SL Chen (2013). Transcriptome analysis of buds and leaves using 454 pyrosequencing to discover genes associated with the biosynthesis of active ingredients in *Lonicera japonica* Thunb. *PLoS One* 8; Article e62922

Huang K, XC Liang, YL Zhong, WY He, Z Wang (2015). 5-Caffeoylquinic acid decreases diet-induced obesity in rats by modulating PPARα and LXRα transcription. *J Sci Food Agric* 95:1903–1910

Iso H, C Date, K Wakai, M Fukui, A Tamakoshi (2006). The relationship between green tea and total caffeine intake and risk for self-reported type 2 diabetes among Japanese adults. *Ann Intern Med* 144:554–562

Jayaswall K, P Mahajan, G Singh, R Parmar, R Seth, A Raina, MK Swarnkar, AK Singh, R Shankar, RK Sharma (2016). Transcriptome analysis reveals candidate genes involved in blister blight defense in tea (*Camellia sinensis* (L) Kuntze). *Sci Rep* 6; Article 30412

Kimura S, Y Chikagawa, M Kato, K Maeda, Y Ozeki (2008). Upregulation of the promoter activity of the carrot (*Daucus carota*) phenylalanine ammonia-lyase gene (DcPAL3) is caused by new members of the transcriptional regulatory proteins, DcERF1 and DcERF2, which bind to the GCC-box homolog and act as an activator to theDcPAL3promoter. *J Plant Res* 121:499–508

Kozuma K, S Tsuchiya, J Kohori, T Hase, I Tokimitsu (2005). Antihypertensive effect of green coffee bean extract on mildly hypertensive subjects. *Hypertens Res* 28:711–718

Leong H, PS Mathur, GL Greene (2008). Inhibition of mammary tumorigenesis in the C3(1)/SV40 mouse model by green tea. *Breast Cancer Res Treat* 107:359–369

Lepelley M, G Cheminade, N Tremillon, A Simkin, V Caillet, J McCarthy (2007). Chlorogenic acid synthesis in coffee: An analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1, and CCoAOMT1 genes during grain development in *C. canephora*. *Plant Sci* 172:978–996

Li B, CN Dewey (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12; Article 323

Liu F, Y Wang, ZT Ding, L Zhao, J Xiao, LJ Wang, J Xiao, SB Ding (2017). Transcriptomic analysis of flower development in tea (*Camellia sinensis*, (L.)). *Gene* 631:39–51

Liu Y, WJ Su, LJ Wang, J Lei, SS Chai, WY Zhang, XS Yang (2021). Integrated transcriptome, small RNA and degradome sequencing approaches proffer insights into chlorogenic acid biosynthesis in leafy sweet potato. *PLoS One* 16; Article e0245266

Love MI, W Huber, S Anders (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15; Article 550

Maeda K, S Kimura, T Demura, J Takeda, Y Ozeki (2005). DcMYB1 acts as a transcriptional activator of the carrot phenylalanine ammonia-lyase gene (DcPAL1) in response to elicitor treatment, UV-B irradiation and the dilution effect. *Plant Mol Biol* 59:739–752

Miguel S, G Legrand, L Duriot, M Delporte, B Menin, C Michel, A Olry, G Chataigné, A Salwinski, J Bygdell, D Vercaigne, G Wingsle, JL Hilbert, F Bourgaud, A Hehn, D Gagneu (2020). A GDSL lipase-like from *Ipomoea batatas* catalyzes efficient production of 3, 5-diCQA when expressed in *Pichia pastoris*. *Commun Biol* 3; Article 673

Minoru K, G Susumu, K Shuichi, O Yasushi, H Masahiro (2004). The KEGG resource for deciphering the genome. *Nucl Acids Res* 32:277–280

Nagao T, T Hase, I Tokimitsu (2007). A green tea extract high in catechins reduces body fat and cardiovascular risks in humans. *Obesity* 15:1473–1483

Naveed M, V Hejazi, M Abbas, AA Kamboh, GJ Khan, M Shumzaid, F Ahmad, D Babazadeh, FF Xia, F Modarresi-Ghazanij, WH Li, XH Zhou (2018). Chlorogenic acid (CGA): A pharmacological review and call for further research. *Biomed Pharmacother* 97:67–74

Niggeweg R, AJ Michael, C Martin (2004). Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nat Biotechnol* 22:746–754

Osborne TB, GF Campbell (1897). The proteids of the sunflower seed. *J Amer Chem Soc* 19:487–494

Pavlica S, R Gebhardt (2010). Protective effects of flavonoids and two metabolites against oxidative stress in neuronal PC12 cells. *Life Sci* 86:79–86

Payyavula RS, R Shakya, VG Sengoda, JE Munyaneza, P Swamy, DA Navarre (2015). Synthesis and regulation of chlorogenic acid in potato: Rerouting phenylpropanoid flux in\r, HQT\r, -silenced lines. *Plant Biotechnol J* 13:551–564

Pellati F, S Benvenuti, L Magro, M Melegari, F Soragni (2004). Analysis of phenolic compounds and radical scavenging activity of *Echinacea* spp. *J Pharm Biomed Anal* 35:289–301

Pertea G, XQ Huang, F Liang, V Antonescu, R Sultana, S Karamycheva, YD Lee, J White, F Cheung, B Parvizi, J Tsai, J Quackenbush (2003). TIGR Gene Indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. *Bioinformatics* 19:651–652

Rhodes MJC, LSC Wooltorton, EJ Lourenco (1979). Purification and properties of hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase from potatoes. *Phytochemistry* 18:1125–1129

Rhodes MJC, LSC Wooltorton (1976). The enzymic conversion of hydroxycinnamic acids to p-coumarylquinic and chlorogenic acids in tomato fruits. *Phytochemistry* 15:947–951

Sawa T, M Nakao, T Akaike, K Ono, H Maeda (1999). Alkylperoxyl radical-scavenging activity of various flavonoids and other phenolic compounds: Implications for the anti-tumor-promoter effect of vegetables. *J Agric Food Chem* 47:397–402

Shadle GL, SV Wesley, KL Korth, F Chen, C Lamb, RA Dixon (2003). Phenylpropanoid compounds and disease resistance in transgenic tobacco with altered expression of l-phenylalanine ammonia-lyase. *Phytochemistry* 64:153–161

Shang QM, L Liang, CJ Dong (2012). Multiple tandem duplication of the phenylalanine ammonia-lyase genes in *Cucumis sativus* L. *Planta* 236:1093–1105

Sharangi AB (2009). Medicinal and therapeutic potentialities of tea (*Camellia sinensis* L.) – A review. *Food Res Intl* 42:529–535

Shi CY, H Yang, CL Wei, O Yu, ZZ Zhang, CJ Jiang, J Sun, YY Li, Q Chen, T Xia, XC Wan (2011). Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genomics* 12; Article 131

Singh K, S Kumar, A Rani, A Gulati, PS Ahuja (2008). Phenylalanine ammonia-lyase (PAL) and cinnamate 4-hydroxylase (C4H) and catechins (flavan-3-ols) accumulation in tea. *Funct Integr Genom* 9:125–134

Sonnante G, RD Amore, E Blanco, CL Pierri, MD Palma, J Luo, M Tucci, C Martin (2010). Novel hydroxycinnamoyl-coenzyme a quinate transferase genes from artichoke are involved in the synthesis of chlorogenic acid. *Plant Physiol* 153:1224–1238

Strickler SR, A Bombarely, LA Mueller (2012). Designing a transcriptome next-generation sequencing project for a nonmodel plant species 1. *Amer J Bot* 99:257–266

Tamagnone L, A Merida, N Stacey, K Plaskitt, A Parr, CF Chang, D Lynn, JM Dow, K Roberts, C Martin (1998). Inhibition of phenolic acid metabolism results in precocious cell death and altered cell morphology in leaves of transgenic tobacco plants. *Plant Cell* 10:1801–1816

Tatusov RL, ND Fedorova, JD Jackson, AR Jacobs, B Kiryutin, EV Koonin, DM Krylov, R Mazumder, SL Mekhedov, AN Nikolskaya, BS Rao, S Smirnov, AV Sverdlov, S Vasudevan, IY Wolf, JJ Yin, DA Natale (2003). The COG database: An updated version includes eukaryotes. *BMC Bioinform* 4; Article 41

Tsai CJ, SA Harding, TJ Tschaplinski, RL Lindroth, YN Yuan (2006). Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in Populus. *New Phytol* 172:47–62

Valiñas MA, ML Lanteri, AT Have, AB Andreu (2017). Chlorogenic acid, anthocyanin and flavan-3-ol biosynthesis in flesh and skin of Andean potato tubers (*Solanum tuberosum* subspp. *andigena*). *Food Chem* 229:837–846

Villegas RJA, M Kojima (1986). Purification and characterization of hydroxycinnamoyl D-glucose. Quinate hydroxycinnamoyl transferase in the root of sweet potato, Ipomoea batatas Lam. *J Biol Chem* 261:8729–8733

Wako T, S Kimura, Y Chikagawa, Y Ozeki (2010). Characterization of MYB proteins acting as transcriptional regulatory factors for carrot phenylalanine ammonia-lyase gene (DcPAL3). *Plant Biotechnol* 27:131–139

Wan XC, T Xia (2015). *Secondary Metabolism of Tea Plant*. Science Press, Beijing, China

Wang HL, YQ Li, SB Wang, DX Kong, SK Sahu, M Bai, HY Li, LZ Li, Y Xu, HP Liang, H Liu, H Wu (2020). Comparative transcriptomic analyses of chlorogenic acid and luteolosides biosynthesis pathways at different flowering stages of diploid and tetraploid *Lonicera japonica*. *Peer J* 8; Article e8690

Wang HZ, U Avci, J Nakashima, MG Hahn, F Chen, RA Dixon (2010). Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. *Proc Natl Acad Sci* 107:22338–22343

Wheeler DS, WJ Wheeler (2004). The medicinal chemistry of tea. *Drug Dev Res* 61:45–65

Wu YL, WZ Wang, WZ Li, XL Dai, GL Ma, DW Xing, MQ Zhu, LP Gao, T Xia (2017). Six phenylalanine ammonia-lyases from, *Camellia sinensis*: Evolution, expression, and kinetics. *Plant Physiol Biochem* 118:413–421

Xu CG, TX Tang, R Chen, CH Liang, XY Liu, CL Wu, YS Yang, DP Yang, H Wu (2014). A comparative study of bioactive secondary metabolite production in diploid and tetraploid *Echinacea purpurea* (L.) Moench. *Plant Cell Tiss Org Cult* 116:323–332

Ye J (2006). WEGO: A web tool for plotting GO annotations. *Nucl Acids Res* 34:293–297

Ye J, W Han, P Deng, Y Jiang, M Liu, L Li, ZQ Li (2019). Comparative transcriptome analysis to identify candidate genes related to chlorogenic acid biosynthesis in *Eucommia ulmoides* Oliv. *Trees* 33:1373–1384

Yu Y, YJ Wang, Y Yue, PY Ma, ZD Jia, XD Guo, YZ Xie, XF Bian (2021). Overexpression of IbPAL1 promotes chlorogenic acid biosynthesis in sweet potato. *Crop J* 9:204–215

Zha L, S Liu, J Liu, C Jiang, S Yu, S Yuan, J Yang, Y Wang, L Huang (2017). DNA methylation influences chlorogenic acid biosynthesis in *Lonicera japonica* by mediating ljbzip8 to regulate phenylalanine ammonia-lyase 2 expression. *Front Plant Sci* 8; Article 1178

Zhang J, Y Yang, K Zheng, M Xie, K Feng, SS Jawdy, LE Gunter, P Ranjan, VR Singan, N Engle, E Lindquist, K Barry, J Schmutz, N Zhao, TJ Tschaplinski, J LeBoldus, GA Tuskan, JG Chen, W Muchero (2018). Genome-wide association studies and expression-based quantitative trait loci analyses reveal roles of HCT2 in caffeoylquinic acid biosynthesis and its regulation by defense-responsive transcription factors in Populus. *New Phytol* 220:502–516

Zhang JR, ML Wu, WD Li, GB Bai (2017). Regulation of chlorogenic acid biosynthesis by hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase in *Lonicera japonica*. *Plant Physiol Biochem* 7:74–79

Zheng C, L Zhao, Y Wang, JZ Shen, YF Zhang, SS Jia, YS Li, ZT Ding (2015). Integrated RNA-Seq and sRNA-Seq analysis identifies chilling and freezing responsive key molecular players and pathways in tea plant (*Camellia sinensis*). *PLoS One* 10; Article 0125031