**Manuscript type: Original Research Article**

**Comparative transcriptome analysis to** **reveal candidate genes involved in chlorogenic acid biosynthesis in** ***Camellia sinensis***

**Running Title: Reveal chlorogenic acid biosynthesis**

Chao Shen1, Xinzhuan Yao1, Degang Zhao2, Litang Lu1,2

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

Corresponding Author:

Litang Lu 1,2

Xibei Community, Guiyang, Guizhou, 550025, People’s Republic of China

Email address: ltlv@gzu.edu.cn

**Novelty statement:** Transcriptome was firstly performed to reveal the synthesis of chlorogenic acid in *Camellia sinensis.*

**Abstract**

The biosynthetic pathway of chlorogenic acid (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 two tea varieties ‘Qianfu 4’ and ‘Qianmei 419’ using HPLC, and the CGA content in ‘Qianfu 4’ was greater than that in ‘Qianmei 419’. Transcriptome sequencing for *Camellia sinensis* with ‘Qianfu 4’ and ‘Qianmei 419’ 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 ‘Qianfu 4’ and ‘Qianmei 419’, 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 hydroxycinnamoyltransferase (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*.

**Keywords:** *Camellia sinensis*; Chlorogenic acid; Transcriptome; HPLC; DEGs

**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; Leong et al., 2008; Nagao et al., 2007; Iso and Hiroyasu, 2006). 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 utlizing 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-tumour and antiviral activities, and it also has antidiabetic, anti-obesity, anti-inflammatory, and antimicrobial effects (Kozuma et al., 2005; Sawa et al., 1999; Huang et al., 2015; Naveed et al., 2018). Furthermore, CGA showed stronger free radical scavenging activity than ascorbic acid (Pellati et al., 2004; Pavlica and Gebhardt, 2010). 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, 1998). Besides, in transgenic tomatoes, elevated levels of CGA have been shown to reduce UV damage and enhance resistance to microorganisms (Clé et al., 2008; Niggeweg et al., 2004). Furthermore, CGA is also used as a precursor substance for synthesizing G- and S-type lignins 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.

In our research, two tea varieties, ‘Qianfu 4’ and ‘Qianmei 419’, were selected as the study materials in this experiment. ‘Qianfu 4’, a new triploid tea variety, was bred from an asexual propagation after systematic selection of a diplont tea variety ‘Qianmei 419’ seeds treated by 60Co-γRay (Chen et al., 2009). 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. Therefore, HPLC was employed to detect the difference of CGA content between two tea varieties. A comparative transcriptome analysis in two tea varieties with different CGA content was conducted by the Illumina HiSeq™ 2000 platform. Candidate genes may participate in the process of CGA biosynthesis were obtained by annotating and analyzing digital gene expression (DEG). Then, the potential key genes in the CGA synthesis pathway of *Camellia sinensis* were identified by analyzing the differentially expressed genes discovered by RNA-Seq analysis. The results of this study will lay the molecular foundation for further revealing the molecular mechanism of CGA synthesis in *Camellia sinensis* and cultivating new germplasm of tea plants with strong stress resistance and good tea quality.

**Materials and Methods**

**Plant materials**

Leaves of tea (*Camellia sinensis*) used in this study were collected from Guizhou tea institute in Meitan, Guizhou, China. ‘Qianfu 4’ and ‘Qianmei 419’ were grown under the same conditions, and both grow well. The second leaves from the top buds of tea plants were collected. 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 1.000 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 25ml and shook well, filtered through a 0.45um microporous filter membrane, and analyzed by HPLC for CGA content.

A Shimadzu LC-20AT HPLC analyser (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℃. 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 20ul, 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, then the content of CGA was calculated according to the peak area. SPSS software was used to data processing.

**Transcriptome sequencing**

Firstly, the RNeasy Plant Mini Kit (QIAGEN, USA) was employed to extracted total RNA from leaf of ‘Qianfu 4’ and ‘Qianmei 419’. 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 (Aparicio et al., 2006; Conesa et al., 2005). 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**

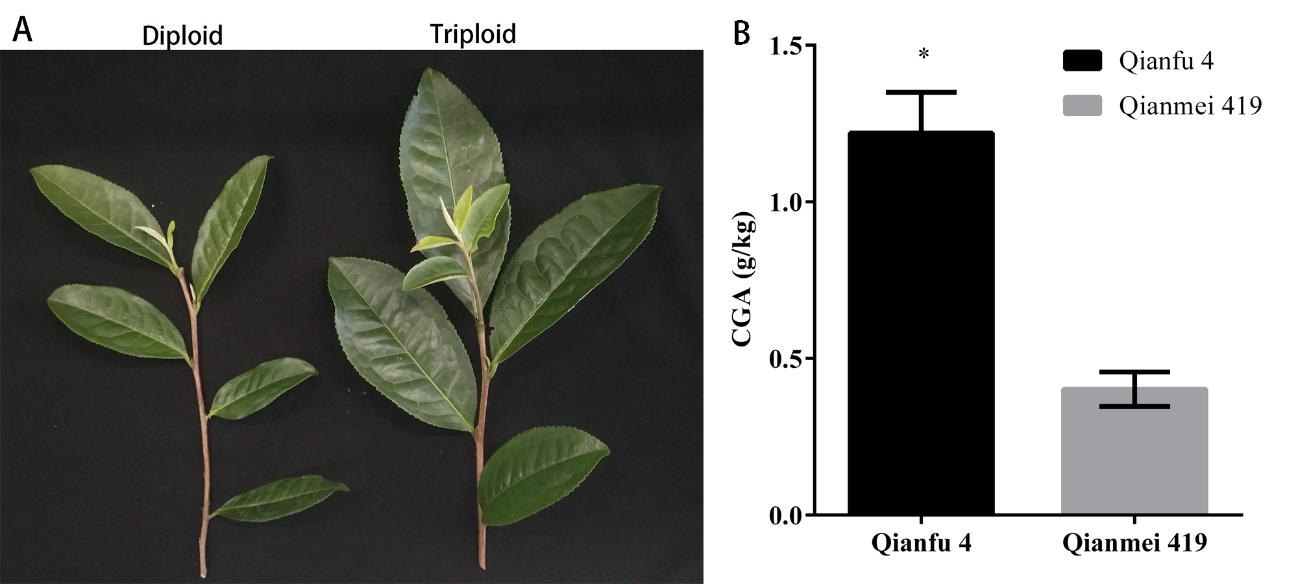
Total RNA of ‘Qianfu 4’ and ‘Qianmei 419’ was extracted to establish two libraries for RNA-Seq analysis. RSEM was employed to calculate gene expression levels (Li and Dewey, 2011). After the aligning the clean reads to the de novo assembled transcriptome, the unigene expression was calculated ultlizing the technique of FPKM (fragment per kilobase per million mapped reads).After that, DESeq2 (Love et al., Ashburner et al., 2000) was implemented to analysis differential expression of ‘Qianfu 4’ and ‘Qianmei 419’. In our research, the FPKM values between ‘Qianfu 4’ and ‘Qianmei 419’ 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 ‘Qianfu 4’ and ‘Qianmei 419’ 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 species) were subjected to CGA content of leaves analyses based on HPLC. The results indicate that the content of CGA in ‘Qianfu 4’ was higher than ‘Qianmei 419’ (P < 0.05) (Fig. 1).

**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.

**Transcriptome Profiling of *Camellia sinensis***

To elucidate the molecular mechanisms and discover the key genes regulating CGA synthesis in *Camellia sinensis*, two C. sinensis samples with different concentrations of CGA were used for transcriptome sequencing (three replicates per sample). 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). Next, the transcripts were assembled by Trinty, and then removed for redundancy and further splicing by Tgicl, 154,097 non-redundant unigenes were obtained, which had average length of 825.67 bp and N50 of 1349 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.

**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% |

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’ containing51,707 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 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, 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 chlorogenic acid synthesis pathway, and 816 transcription factors that may regulate the above enzymes in *Camellia sinensis*. (Table 3)

**Table 2**. The unigenes related to other secondary metabolites.

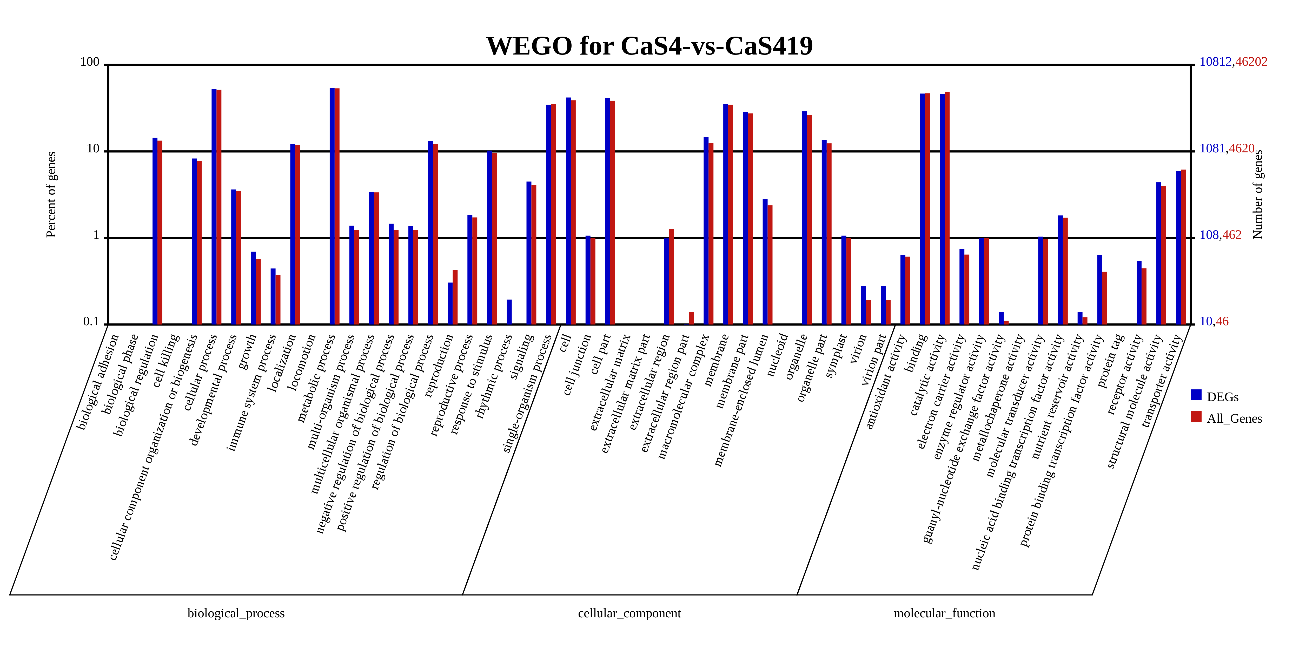
|  |  |  |
| --- | --- | --- |
| Biosynthesis of other secondary metabolites | Unigene numbers | Pathway ID |
| Carbapenem biosynthesis | 4 | ko00332 |
| Betalain biosynthesis | 5 | ko00965 |
| Acridone alkaloid biosynthesis | 11 | ko01058 |
| Anthocyanin biosynthesis | 18 | ko00942 |
| Caffeine metabolism | 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 |

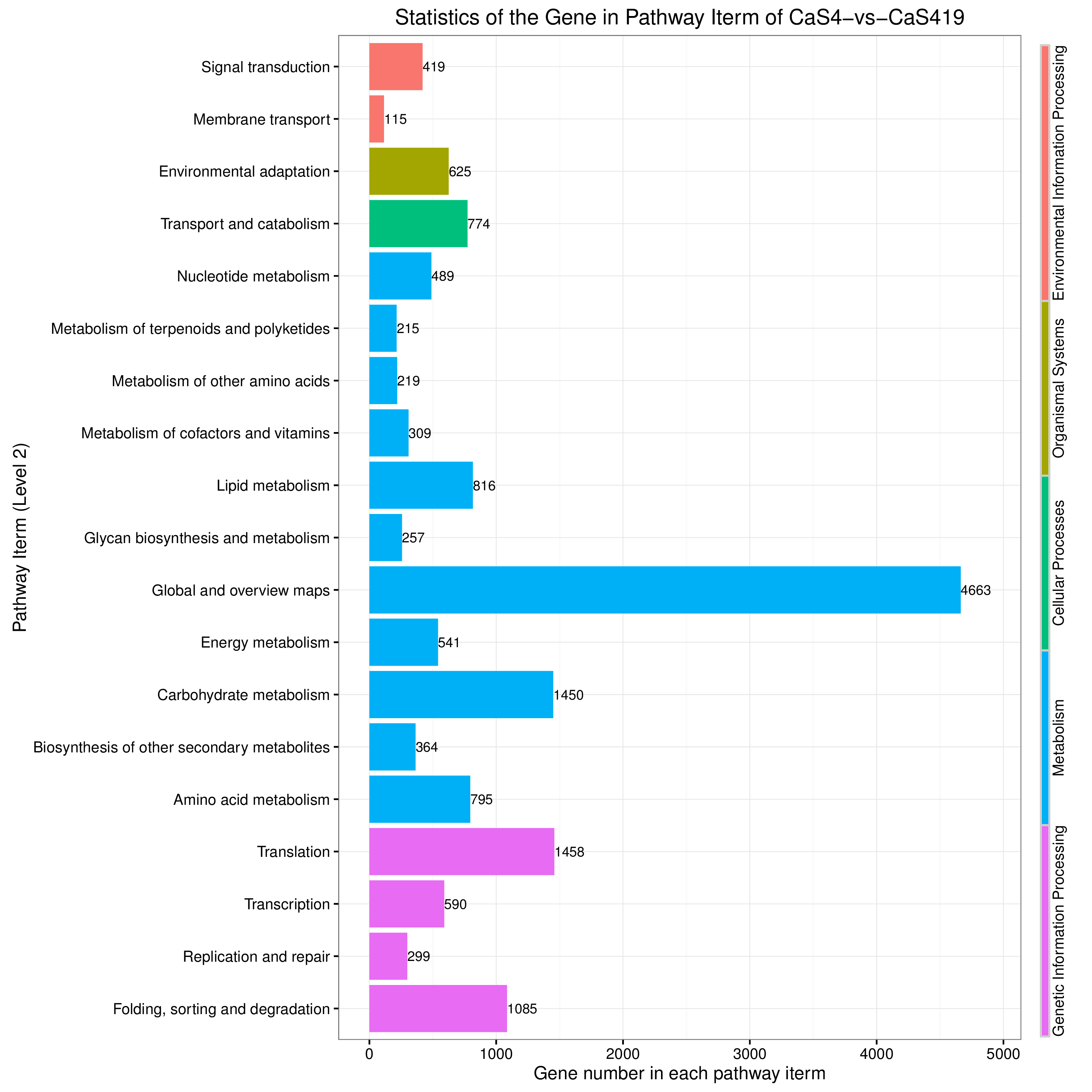
**Functional analysis of DEGs**

In order to screen the candidate unigenes that play an important role in the synthesis of CGA in *Camellia sinensis*, we conducted comparative transcriptome analysis of two tea varieties (‘Qianfu 4’, ‘Qianmei 419’) with different concentrations of CGA. 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 ‘Qianfu 4’ leaves compared with the ‘Qianmei 419’ 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).

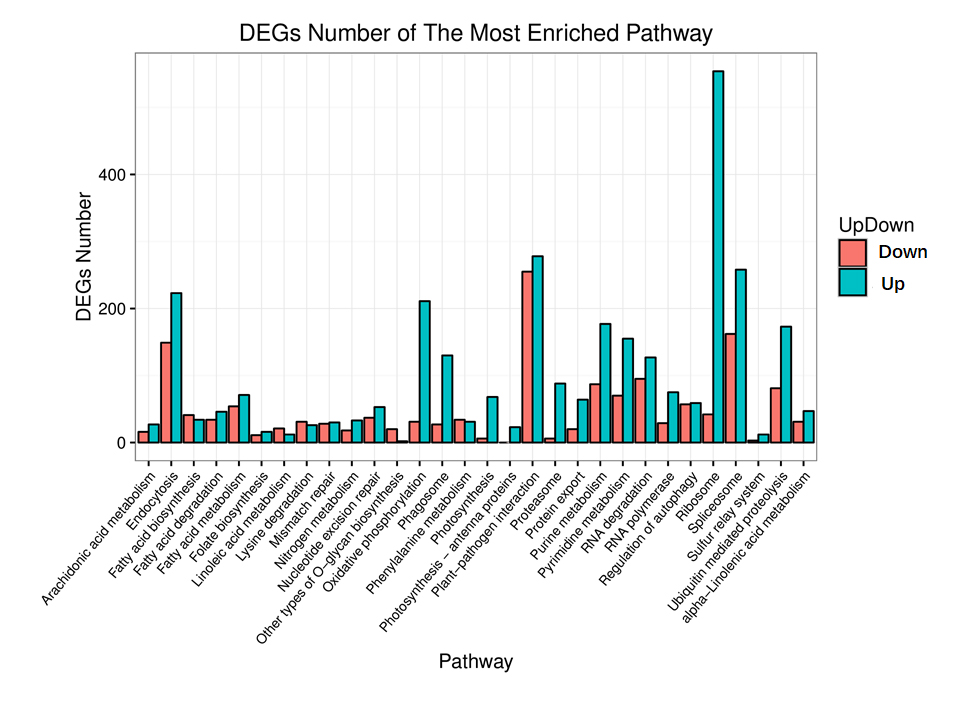


**Fig. 2** WEGO results of differentially expressed genes (DEGs) from Camellia sinensis. The x-coordinate represents the three aspects of GO (cellular component, molecular function, and biological process). The left y-axis indicates the percentage of genes in a category, and the right y-axis indicates the number of genes in a category. The red number represents the background (all genes) and the blue number represents the differential genes.

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 (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. Further research is needed to explore it.



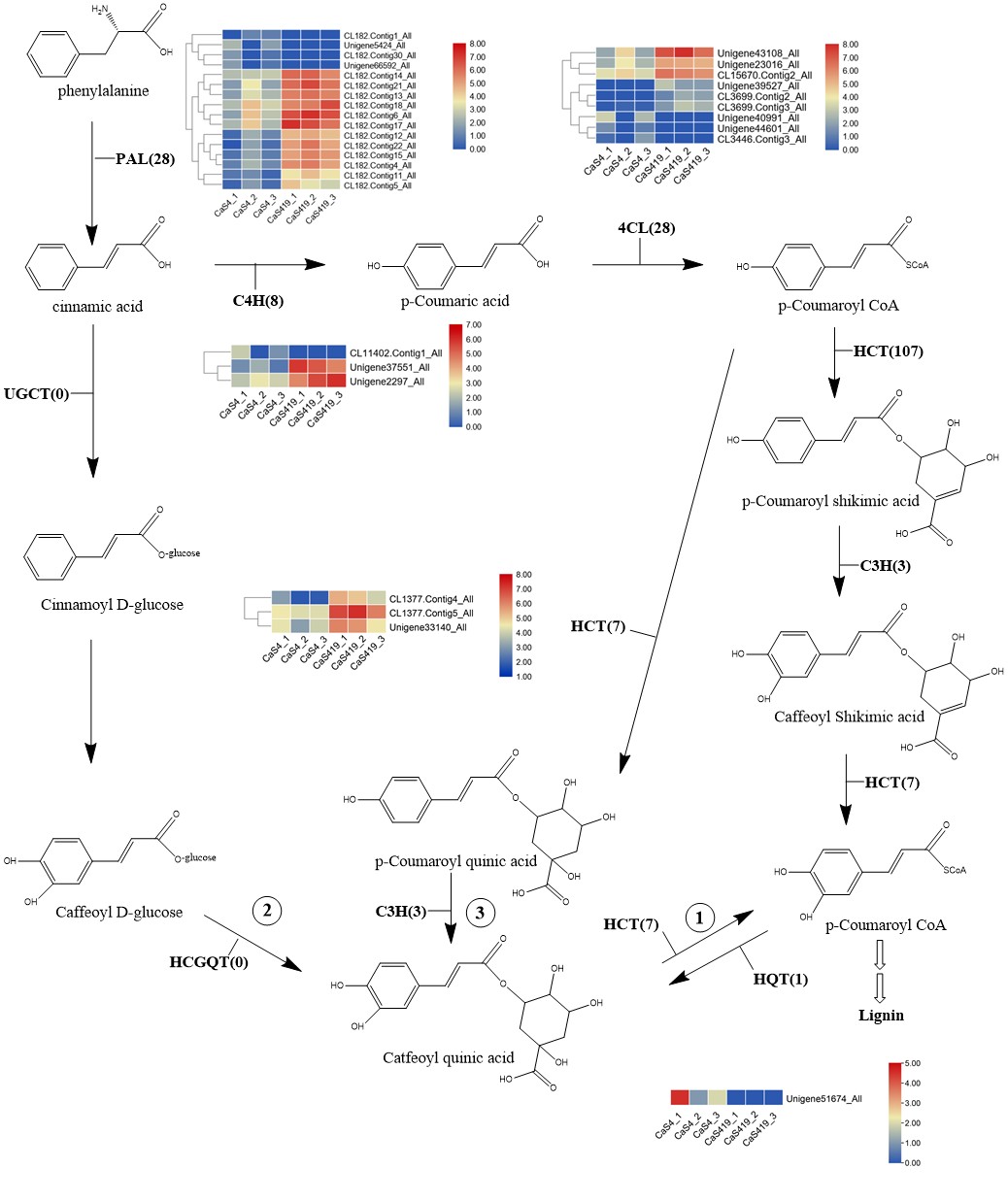
**Fig. 3** KEGG pathway analysis of DEGs in ‘Qianfu 4’-vs- ‘Qianmei 419’. DEGs were distributed in 5 functional categories (cellular processes, environmental information processing, genetic information processing, metabolism, and organismal systems)



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

**Identifying *Camellia sinensis* CGA synthesis-associated unigenes**

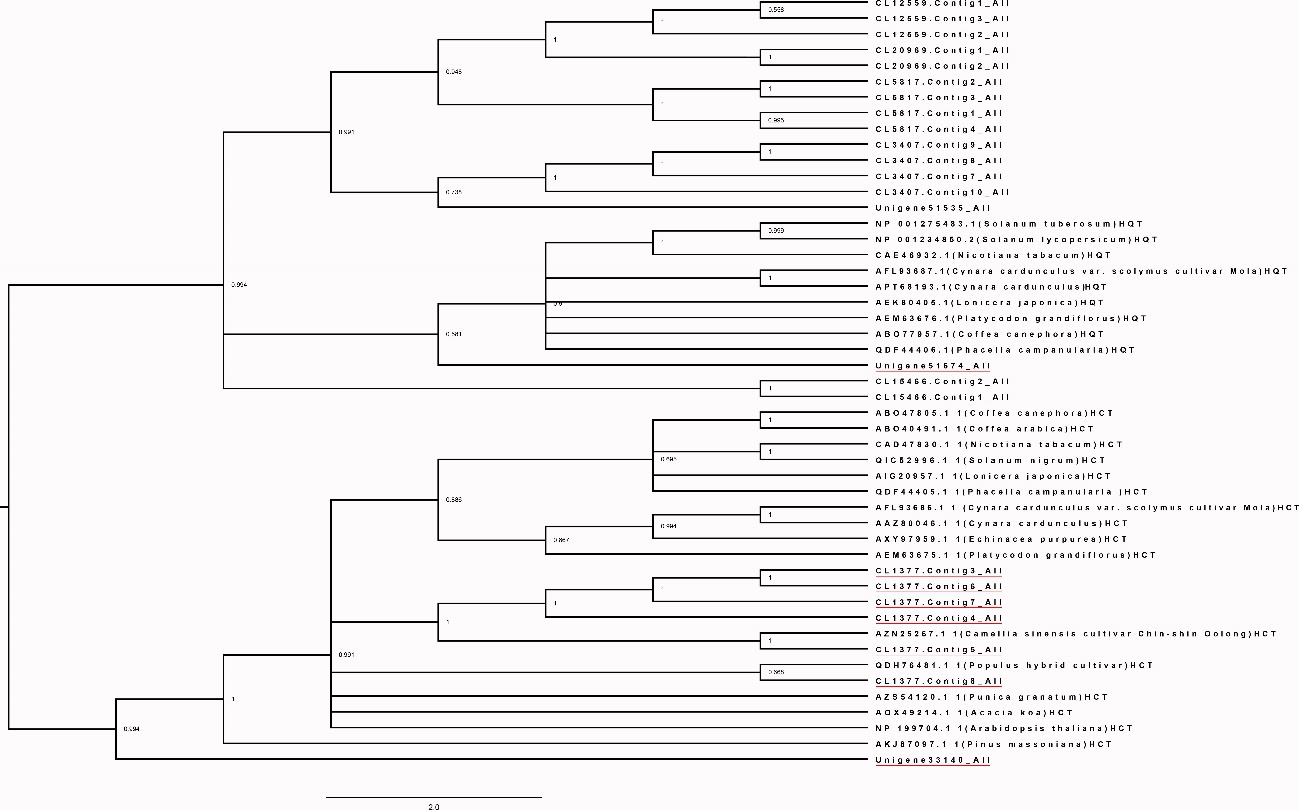
CGA is a phenylpropanoid originated from phenylalanine during plants’ aerobic respiration. According to the information obtained from the transcriptome, based on the annotation information of the major gene databases for unigenes, several gene sequences encoding the enzymes on the CGA synthesis pathway 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 and Table S3). 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 and Table S4, 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 (Figs. 5, S4).



**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 colour scale represents log2 transformed FPKM values. Blue indicates low expression, and red indicates high expression

**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. Therefore, the data obtained from transcriptome sequencing in this research can be utilized to further investigatation on genes related to CGA biosynthesis.



**Fig. 6** Phylogenetic tree of HQT/HCT protein sequences available in public databases.

**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 it's 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 diplont tea variety and a triploid tea variety were selected for research. HPLC revealed that the content of CGA in leaves of ‘Qianfu 4’ was significantly higher than that of ‘Qianmei 419’. The RNA from leaves of two tea varieties was extracted for transcriptome sequencing. 154,097 unigenes were obtained in total. Then, the DEGs of *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 widely found in many plant species for antioxidant. 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 and NO.3 synthetic pathways 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 (Chang, et al., 2009; Shadle et al., 2003). 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. Wu Y et al 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. C4H is the second enzyme of phenylpropanoid pathway after PAL, converting cinnamic acid to p-coumaric acid. There were 1 of 8 genes coding C4H found to have the expression level consisting with CGA content (CL11402.Contig1\_All). 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). Because 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. 6), 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 ‘Qianfu 4’, while the expression level of CL1377.Contig4 All, CL 1377.Contigs All, and Unigene33140 All are higher in ‘Qianmei 419’. Overall, CGA synthesis pathway of many enzymes in ‘Qianmei 419’ expression quantity is higher than in ‘Qianfu 4’, while CGA content in the ‘Qianfu 4’ rather than ‘Qianmei 419’. However, only one potential gene encoding HQT was highly expressed in ‘Qianfu 4’. Combined with studies on the model plant *Arabidopsis thaliana*, in which both C3H and HCT were present, but no chlorogenic acid 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). Therefore, we speculated that chlorogenic acid synthesis in tea plants may also be dominated by HQT. At the same time, in the third synthetic route, the last of the chlorogenic acid 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 chlorogenic acid synthesis in *Camellia sinensis*, resulting in low chlorogenic acid content in ‘Qianmei 419’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.

**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.

**References**

Chan, E.W.C, Y.Y. Lim, and Y.L. Chew, 2007. Antioxidant activity of *Camellia sinensis* leaves and tea from a lowland plantation in Malaysia. Food Chemistry., 102:1214-1222.

Wan, X.C. and T. Xia, 2015. Secondary metabolism of tea plant, Science press, Beijing, China.

Sharangi, A.B, 2009. Medicinal and therapeutic potentialities of tea (*Camellia sinensis* L.) – A review. Food Research International., 42: 529-535.

Wheeler, D.S. and W.J. Wheeler, 2004. The Medicinal Chemistry of Tea. Drug Development Research., 61:45-65.

Beltz, L.A., D.K. Bayer, A.L. Moss and I.M. Simet, 2006. Mechanisms of Cancer Prevention by Green and Black Tea Polyphenols. Anti-Cancer Agents in Medicinal Chemistry., 6:389-406.

Leong, H., P.S. Mathur and G.L. Greene, 2008. Inhibition of mammary tumorigenesis in the C3(1)/SV40 mouse model by green tea. Breast Cancer Research and Treatment., 107:359-369.

Nagao, T., T. Hase and I. Tokimitsu, 2007. A Green Tea Extract High in Catechins Reduces Body Fat and Cardiovascular Risks in Humans. Obesity., 15:1473-1483.

Iso, H., C. Date, K, Wakai, M. Fukui and A. Tamakoshi, 2006. The Relationship between Green Tea and Total Caffeine Intake and Risk for Self-Reported Type 2 Diabetes among Japanese Adults. Annals of Internal Medicine., 144:554-562.

Bradfield, A.E. and A.E. Flood, 1952. Chlorogenic Acids in Fruit Trees. Nature., 170:168-169.

Deng, L., H.Yuan and Z.Y. Yu, 2005. Recent advances in research on chlorogenic acid. Chemistry & Bioengineering., 7:4-6.

Osborne, T.B. and G.F. Campbell, 1897. The proteids of the sunflower seed. Journal of the American Chemical Society., 19:487-494.

Kozuma, K., S.Tsuchiya, J. Kohori, T. Hase and I. Tokimitsu, 2005. Antihypertensive Effect of Green Coffee Bean Extract on Mildly Hypertensive Subjects. Hypertension Research., 28:711-718.

Sawa, T., M. Nakao, T. Akaike, K. Ono and H. Maeda, 1999. Alkylperoxyl Radical-Scavenging Activity of Various Flavonoids and Other Phenolic Compounds: Implications for the Anti-Tumor-Promoter Effect of Vegetables. Journal of Agricultural and Food Chemistry., 47:397-402.

Huang, K., X.C. Liang, Y.L. Zhong, W.Y. He and Z. Wang, 2015. 5-Caffeoylquinic acid decreases diet-induced obesity in rats by modulating PPARα and LXRα transcription. Journal of the Science of Food and Agriculture., 95:1903-1910.

Naveed, M., V. Hejazi, M. Abbas, A.A. Kamboh, G.J. Khan, M. Shumzaid, F. Ahmad, Babazadeh, D., F.F. Xia, F. Modarresi-Ghazanij, W.H. Li and X.H. Zhou, 2018. Chlorogenic acid (CGA): A pharmacological review and call for further research. Biomedicine & Pharmacotherapy., 97:67-74.

Pellati, F., S. Benvenuti, L. Magro, M. Melegari and F. Soragni, 2004. Analysis of phenolic compounds and radical scavenging activity of Echinacea spp. Journal of Pharmaceutical and Biomedical Analysis., 35:289-301.

Pavlica, S. and R, Gebhardt, 2010. Protective effects of flavonoids and two metabolites against oxidative stress in neuronal PC12 cells. Life Sciences., 86:79-86.

Tamagnone, L., A. Merida, N. Stacey, K. Plaskitt, A. Parr, C.F. Chang, D. Lynn, J.M. Dow, K. Roberts and C. Martin, 1998. Inhibition of Phenolic Acid Metabolism Results in Precocious Cell Death and Altered Cell Morphology in Leaves of Transgenic Tobacco Plants. The plant cell., 10:1801-1816.

Clé, C., L.M. Hill, R. Niggeweg, C.R. Martin, Y. Guisez, E. Prinsen and M.A.K. Jansen, 2008. Modulation of chlorogenic acid biosynthesis in Solanum lycopersicum; consequences for phenolic accumulation and UV-tolerance. Phytochemistry., 69:2149-2156.

Niggeweg, R., A.J. Michael and C. Martin, 2004. Engineering plants with increased levels of the antioxidant chlorogenic acid. Nature Biotechnology., 22:746-754.

Abdulrazzak, N., B. Pollet, J. Ehlting, K. Larsen, C. Asnaghi, S. Ronseau, C. Proux, M. Erhardt, V. Seltzer, J.P. Renou, P. Ullmann, M. Pauly, C. Lapierre and D.W. 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 Physiology., 140:30-48.

Rhodes, M.J.C., L.S.C. Wooltorton and E.J. Lourenco, 1979. Purification and properties of hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase from potatoes. Phytochemistry., 18:1125-1129.

Villegas, R.J.A. and M. Kojima, 1986. Purification and characterization of hydroxycinnamoyl D-glucose. Quinate hydroxycinnamoyl transferase in the root of sweet potato, Ipomoea batatas Lam. Journal of Biological Chemistry., 261:8729-8733.

Rhodes, M.J.C. and L.S.C. Wooltorton, 1976. The enzymic conversion of hydroxycinnamic acids to p-coumarylquinic and chlorogenic acids in tomato fruits. Phytochemistry., 15:947-951.

Bate, N.J., J. Orr, W. Ni, A. Meromi, T. Nadler-Hassar, P.W. Doerner, R.A. Dixon, R.A. Lamb and 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. Proceedings of the National Academy of Sciences., 91:7608-7612.

Dixon, R.A., L. Achnine, P. Kota, C.J. Liu, M.S.S. Reddy, L. Wang, 2002. The phenylpropanoid pathway and plant defence – a genomics perspective. Molecular Plant Pathology., 3:371-390.

Sonnante, G., R.D. Amore, E. Blanco, C.L. Pierri, M.D. Palma, J. Luo, M. Tucci and C. Martin, 2010. Novel Hydroxycinnamoyl-Coenzyme A Quinate Transferase Genes from Artichoke Are Involved in the Synthesis of Chlorogenic Acid. PLANT PHYSIOLOGY 153:1224-1238.

Payyavula, R.S., R. Shakya, V.G. Sengoda, J.E. Munyaneza, P. Swamy and D.A. Navarre, 2015. Synthesis and regulation of chlorogenic acid in potato: Rerouting phenylpropanoid flux in\r, HQT\r, -silenced lines. Plant Biotechnology Journal., 13:551-564.

Zhang, J., Y. Yang, K. Zheng, M. Xie, K. Feng, S.S. Jawdy, L.E. Gunter, P. Ranjan, V.R. Singan, N. Engle, E. Lindquist, K. Barry, J. Schmutz, N. Zhao, T.J. Tschaplinski, J. LeBoldus, G.A. Tuskan, J.G. Chen and 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 Phytologist., 220: 502-516.

Ye, J., W. Han, P. Deng, Y. Jiang, M. Liu, L. Li and Z.Q. Li, 2019. Comparative transcriptome analysis to identify candidate genes related to chlorogenic acid biosynthesis in Eucommia ulmoides Oliv. Trees.

Zha, L.P., S. Liu, J. Liu, C. Jiang, S.L. Yu, Y. Yuan, J. Yang, Y.L. Wang and L.Q. Huang, 2017. DNA Methylation Influences Chlorogenic Acid Biosynthesis in Lonicera japonica by Mediating LjbZIP8 to Regulate Phenylalanine Ammonia-Lyase 2 Expression. Frontiers in Plant Science., 8:1178-.

Maeda, K., S. Kimura, T. Demura, J. Takeda and 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 Molecular Biology., 59:739-752.

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

Wang, H.Z., U. Avci, J. Nakashima, M.G. Hahn, F. Chen and R.A. Dixon, 2010. Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. Proceedings of the National Academy of Sciences., 107:22338-22343.

Kimura, S., Y. Chikagawa, M. Kato, K. Maeda and 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. Journal of Plant Research., 121:499-508.

Chen, Z.W., Z.M. Liu and J.L. Wang, 2009. Breeding of Qianfu 4, new triploid tea variety. Southwest China Journal of Agricultural Sciences., 22:1194-1197.

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q.D. Zeng, Z.H. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F.D. Palma, B.W. Birren, C. Nusbaum, K.L. Toh, N. Friedman and A. Regev, 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology., 29:644-652.

Pertea, G., X.Q. Huang, F. Liang, V. Antonescu, R. Sultana, S. Karamycheva, Y.D. Lee, J. White, F. Cheung, B. Parvizi, J. Tsai and J. Quackenbush, 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics., 19: 651-652.

Altschul, S.F., W. Gish, W. Miller, E.W. Miller and D.J. Lipman, 1990. Basic local alignment search tool. Journal of Molecular Biology., 215:403-410.

Deng, Y.Y., J.Q. Li, S.F. Wu, Y.P. Zhu and F.C. He, 2006. Integrated nr Database in Protein Annotation System and Its Localization. Computer Engineering., 32:71-72.

Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein and J.M. Cherry, 2000. Gene ontology: Tool for the unification of biology. Gene., 25:25-29.

Tatusov, R.L., N.D. Fedorova, J.D. Jackson, A.R. Jacobs, B. Kiryutin, E.V. Koonin, D.M. Krylov, R. Mazumder, S.L. Mekhedov, A.N. Nikolskaya, B.S. Rao, S. Smirnov, A.V. Sverdlov, S. Vasudevan, I.Y. Wolf, J.J. Yin and D.A. Natale, 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics., 4:1-14.

Minoru, K., G. Susumu, K. Shuichi, O. Yasushi and H. Masahiro, 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res., 32:277-280.

Apweiler, R., R. Bairoch, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H.Z. Huang, R. Lopez, M. Magrane, M.J. Martin, D.A. Natale, C. O'Donovan, N. Redaschi and L.S.L. Yeh, 2004. UniProt: the Universal Protein knowledgebase. Nucleic Acids Research., 32:115-119.

Aparicio, G., S. Götz, A. Conesa, D. Segrelles and M. Talon, 2006. Blast2GO goes Grid: Developing a Grid-Enabled Prototype for Functional Genomics Analysis. Studies in health technology and informatics., 120:194-204.

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

Ye, J., 2006. WEGO: a web tool for plotting GO annotations. Nucleic Acids Research., 34:293-297.

Li, B., and C.N. Dewey, 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323-0.

Love, M.I., W. Huber and S. Anders, 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology., 15:550-0.

Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L.I. Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin and G. Sherlock, 2000. Gene ontology: Tool for the unification of biology. Gene., 25:25-29.

Shi, C.Y., H. Yang, C.L. Wei, O. Yu, Z.Z. Zhang, C.J. Jiang, J. Sun, Y.Y. Li, Q. Chen, T. Xia and X.C. Wan, 2011. Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. BMC Genomics., 12:131-0.

Liu, F., Y. Wang, Z.T. Ding, L. Zhao, J. Xiao, L.J. Wang, J. Xiao and S.B. Ding, 2017. Transcriptomic analysis of flower development in tea (*Camellia sinensis*, (L.)). Gene., 631:39-51.

Jayaswall, K., P. Mahajan, G. Singh, R. Parmar, R. Seth, A. Raina, M.K. Swarnkar, A.K. Singh, R. Shankar and R.K. Sharma, 2016. Transcriptome Analysis Reveals Candidate Genes involved in Blister Blight defense in Tea (*Camellia sinensis* (L) Kuntze). Scientific Reports., 6:30412-.

Zheng, C., L. Zhao, Y. Wang, J.Z. Shen, Y.F. Zhang, S.S. Jia, Y.S. Li and Z.T. 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: e0125031.

Strickler, S.R., A. Bombarely and L.A. Mueller, 2012. Designing a transcriptome next-generation sequencing project for a nonmodel plant species1. American Journal of Botany., 99:257-266.

Dhawan, O.P., and U.C. Lavania, 1996. Enhancing the productivity of secondary metabolites via induced polyploidy: a review. Euphytica., 87:81-89.

Lepelley, M., G. Cheminade, N. Tremillon, A. Simkin, V. Caillet and 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 Science., 172: 978-996.

Zhang, J.R., M.L. Wu, W.D. Li and G.B. Bai, 2017. Regulation of chlorogenic acid biosynthesis by hydroxycinnamoyl CoA quinate hydroxycinnamoyl transferase in Lonicera japonica. Plant Physiol Biochem., 7:74-79.

He, L., X.L. Xu, Y. Li, C.F. Li, Y.J. Zhu, H.X. Yan, Z.Y. Sun, C. Sun, J.Y. Song, Y. Bi, J. Shen, R.Y. Cheng, Z.Z. Wang, W. Xiao and S.L. 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: e62922-.

Chen, Z.X., N. Tang, Y.M. You, J.B. Lan, Y.Q. Liu and Z.G. 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: e0137212-.

Singh, K., S. Kumar, A. Rani, A. Gulati and P.S. Ahuja, 2008. Phenylalanine ammonia-lyase (PAL) and cinnamate 4-hydroxylase (C4H) and catechins (flavan-3-ols) accumulation in tea. Functional & Integrative Genomics., 9:125-134.

Chang, J., J. Luo and A.G. He, 2009. Regulation of polyphenols accumulation by combined overexpression/silencing key enzymes of phenylpropanoid pathway. Acta Biochimica et Biophysica Sinica., 41:123-130.

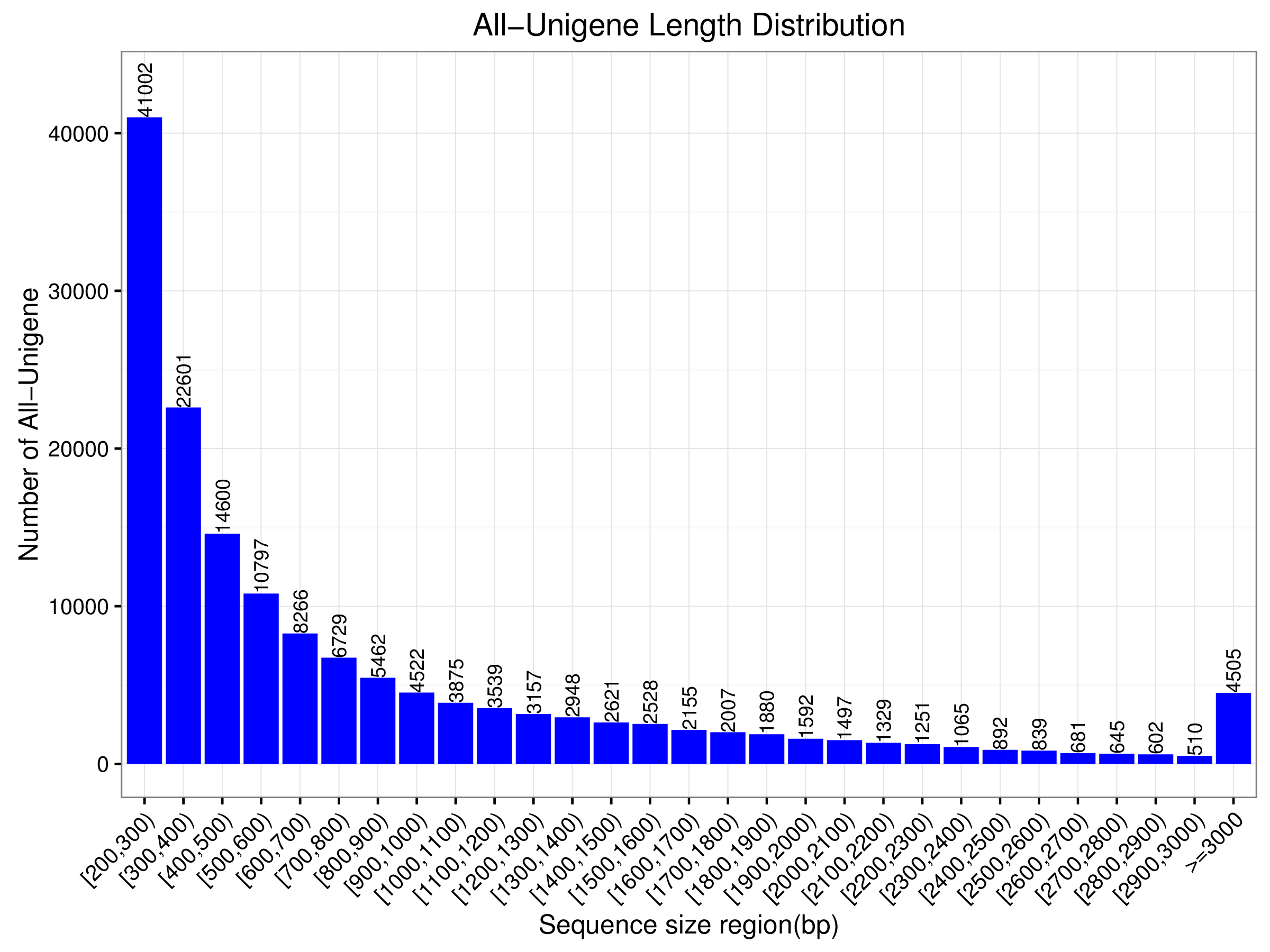
Shadle, G.L., S.V. Wesley, K.L. Korth, F. Chen, C. Lamb and R.A. Dixon, 2003. Phenylpropanoid compounds and disease resistance in transgenic tobacco with altered expression of l-phenylalanine ammonia-lyase. Phytochemistry (Amsterdam)., 64:153-161.

Wu, Y.L., W.Z. Wang, W.Z. Li, X.L. Dai, G.L. Ma, D.W. Xing, M.Q. Zhu, L.P. Gao and T. Xia, 2017. Six phenylalanine ammonia-lyases from, *Camellia sinensis*: Evolution, expression, and kinetics. Plant Physiology and Biochemistry., 118:413-421.

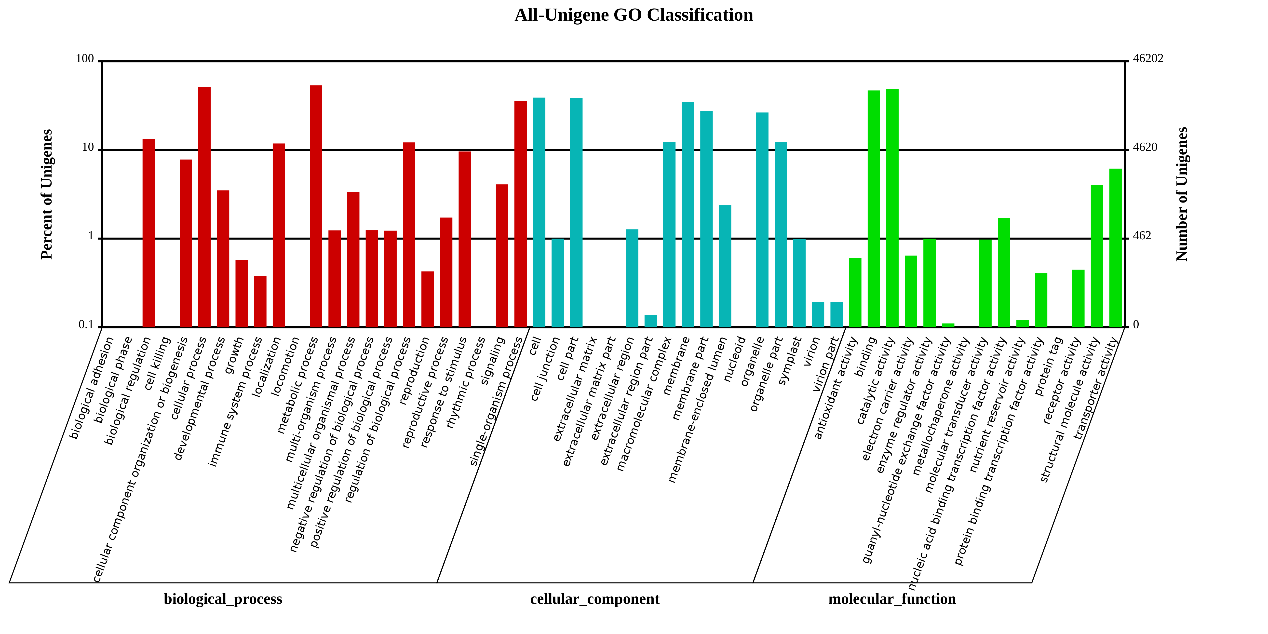
Miguel, S., G. Legrand, L. Duriot, M. Delporte, B. Menin, C. Michel, A. Olry, G. Chataigné, A. Salwinski, J. Bygdell, D. Vercaigne, G. Wingsle, J.L. Hilbert, F. Bourgaud, A. Hehn and D. Gagneu, 2020. A GDSL lipase-like from Ipomoea batatas catalyzes efficient production of 3,5-diCQA when expressed in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

Valiñas, M.A., M.L. Lanteri, A.T. Have and A.B. Andreu, 2017. Chlorogenic acid, anthocyanin and flavan-3-ol biosynthesis in flesh and skin of Andean potato tubers (*Solanum tuberosum subsp. andigena*). Food Chemistry., 229:837-846.

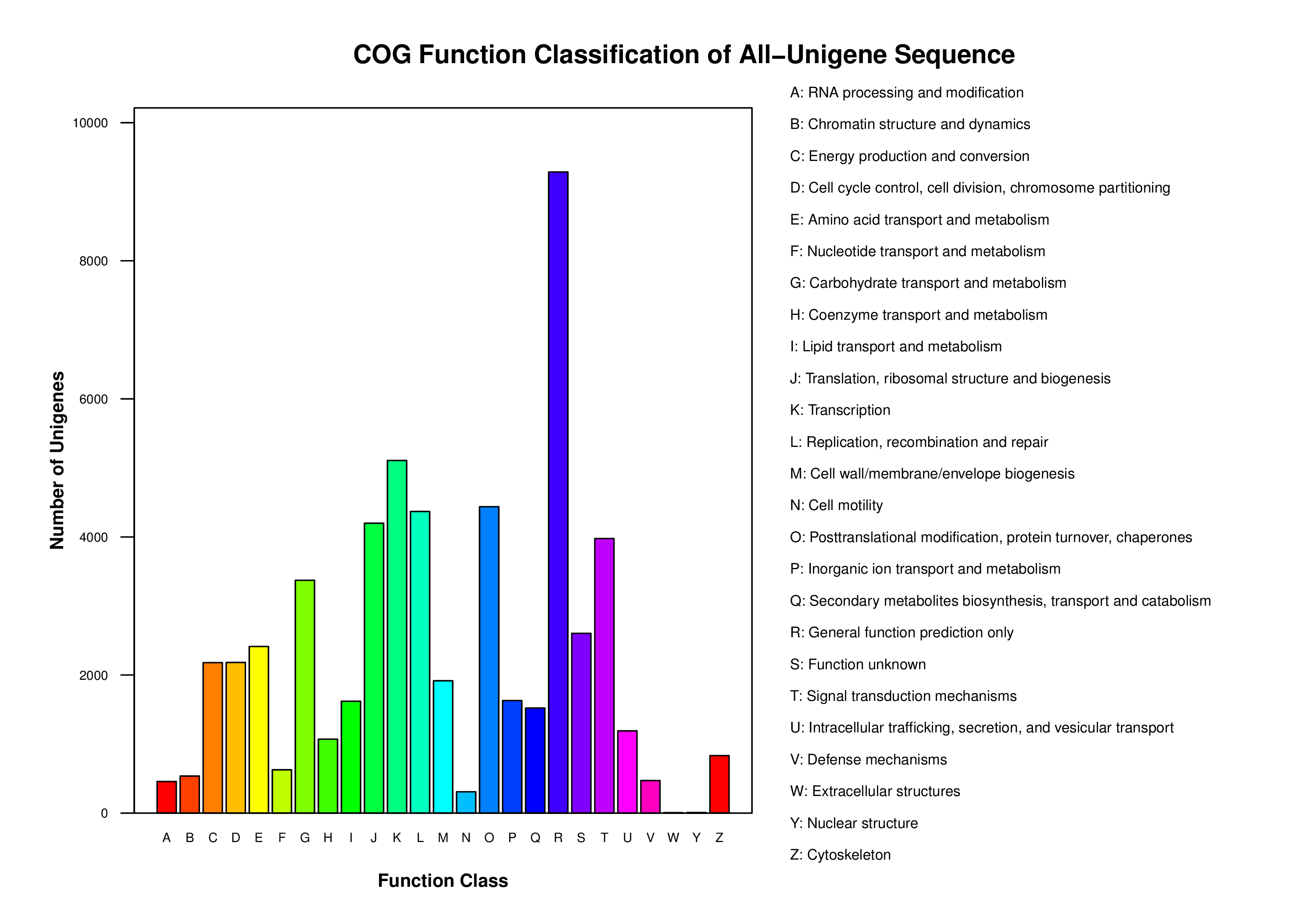
**Supply information**



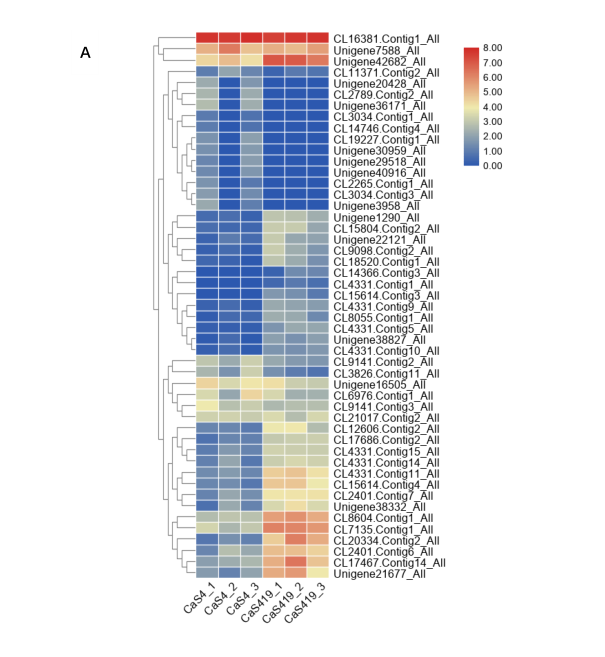
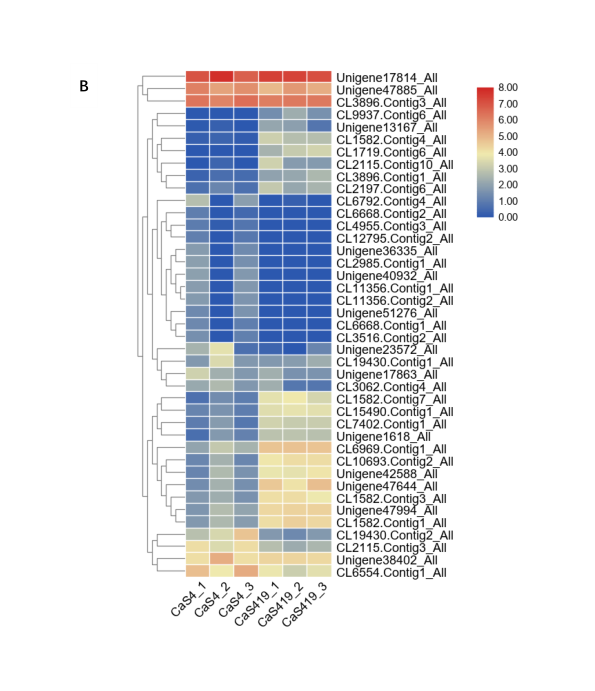
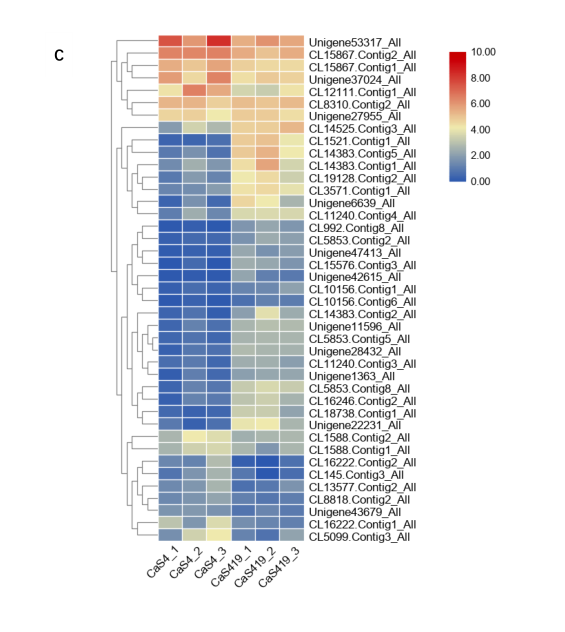
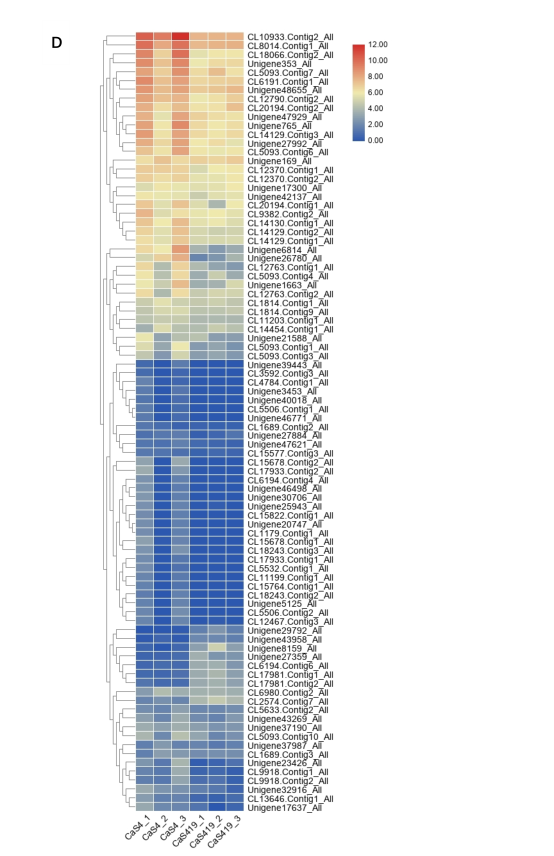
**Fig. S1** Length distribution of the *Camellia sinensis* unigenes.



**Fig. S2** Histogram presentation of GO classification. The results are summarized in three main categories, 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.

****

**Fig. S3** Histogram presentation of the COG classification of all unigenes. In total, 56,330 unigenes showing significant homology to the COGs database are grouped into at least 25 categories. The capital letters in x-axis indicate the COG categories as listed on the right and the y-axis indicates the number of unigenes in each category.

**Fig. S4** All DEGs encoding the WRKY (A), bZIP (B), MYB(C) and ERF (D). Transcription factors were hierarchically clustered and mapped using the FPKM values. Colours indicate log2 transformed FPKM values. Green indicates low expression, and red indicates high expression.

**Table S1.** The primer sequences and amplicon characteristics used for qRT-PCR.

|  |  |  |
| --- | --- | --- |
| Unigene ID | Forward primer (5'-3') | Reverse primer (5'-3') |
| CL182.Contig14\_All | TACTCTGGCATTCGGTTCG | TTGAGCGTTTCTCCGTTGG |
| CL15670.Contig2\_All | CAAGCGAAAGCATCCAAGG | CCCGACGAATAAGGTAGCG |
| CL1377.Contig9\_All | ACACCCAGCGTCTACTTTT | TCACGAGCCAGCCGTCCAG |
| CL10933.Contig2\_All | GCGGTAGACATTGACGGAG | AAAGCAGGGAACTGGGCAT |
| CL15867.Contig2\_All | TCCGAAGACGAGGAAACAC | TCCTCTGCCGTCCTTCCTG |
| CL11642.Contig3\_All | GGCTTCATTTGGGAGTTCA | GGCTTCATTTGGGAGTTCA |
| CL14810.Contig2\_All | TTGCTCCCATAGAGTCCCA | CCAACGAAAACCACCTGAT |
| Unigene32202\_All | CGGTAGTTCCGTAGCAGGTT | GCCCCAAGAAGCATCAAACAT |
| CL938.Contig2\_All | TCGGTGACTTTGTGGTGGA | TGAGGATGAAGGCGGTGGT |
| CL290.Contig15\_All | CCGCTGTCTGTCTCCTTAC | TCGTCTATGGTAACGCCTC |

**Table S2.** Summary for the *Camellia sinensis* transcriptome

|  |  |
| --- | --- |
| Total number of raw reads | 318,515,326 |
| Total number of clean reads | 317,609,528 |
| Total number of clean nucleotides (bp) | 47,641,429,200 |
| Average Read Length (bp) | 150 |
| Total number of contigs | 154,097 |
| Total length of contigs (bp) | 127,232,652 |
| Mean length of contigs (bp) | 825 |
| Total number of unigenes | 154,097 |
| Total length of unigenes (bp) | 127,233,270 |
| Mean length of unigenes | 826 |
| N50 | 1,349 |

**Table S3.** Expression profiles of genes involved in CGA biosynthesis of *Camellia sinensis*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Gene | Unigene ID | FPKM | | | | | |
| CaS4\_1 | CaS4\_2 | CaS4\_3 | CaS419\_1 | CaS419\_2 | CaS419\_3 |
| PAL | CL182.Contig10\_All | 2.48 | 17.13 | 10.02 | 35.77 | 45.17 | 30.93 |
|  | CL182.Contig11\_All | 0.71 | 1.44 | 0.28 | 12.41 | 21.42 | 12.53 |
|  | CL182.Contig12\_All | 0.37 | 3.35 | 1.22 | 23.59 | 29.03 | 21.36 |
|  | CL182.Contig13\_All | 2.32 | 8.36 | 4.15 | 49.91 | 62.45 | 47.8 |
|  | CL182.Contig14\_All | 4.63 | 5.79 | 6.04 | 67.52 | 75.06 | 43.51 |
|  | CL182.Contig15\_All | 0.6 | 2.84 | 0.88 | 31.74 | 48.21 | 36.04 |
|  | CL182.Contig16\_All | 0 | 0.86 | 0 | 4.26 | 6.43 | 3.53 |
|  | CL182.Contig18\_All | 4.46 | 20.85 | 7.08 | 45.91 | 52.68 | 102.15 |
|  | CL182.Contig19\_All | 0.97 | 23.79 | 1.67 | 28.45 | 46.37 | 29.59 |
|  | CL182.Contig1\_All | 0 | 1.14 | 0.94 | 0 | 0 | 0 |
|  | CL182.Contig20\_All | 18.5 | 0 | 10.97 | 0 | 0.04 | 0 |
|  | CL182.Contig21\_All | 2.14 | 10.87 | 2.66 | 65.79 | 106.58 | 57.53 |
|  | CL182.Contig22\_All | 0 | 2.54 | 0.49 | 33.03 | 57.92 | 29.56 |
|  | CL182.Contig25\_All | 6.86 | 0.03 | 3.76 | 0 | 0.05 | 0 |
|  | CL182.Contig26\_All | 0.74 | 0 | 0.92 | 0 | 0 | 0 |
|  | CL182.Contig27\_All | 2.25 | 0 | 0 | 0 | 0 | 0 |
|  | CL182.Contig29\_All | 24.83 | 0 | 11.42 | 0.05 | 0 | 0 |
|  | CL182.Contig30\_All | 1.99 | 0 | 0.61 | 0 | 0 | 0 |
|  | CL182.Contig3\_All | 0 | 0.19 | 0 | 0.33 | 0 | 0 |
|  | CL182.Contig4\_All | 1 | 3.9 | 1.34 | 41.68 | 52.96 | 31.49 |
|  | CL182.Contig5\_All | 0.25 | 2.16 | 0.44 | 20.27 | 9.64 | 6.92 |
|  | CL182.Contig6\_All | 2.48 | 20.27 | 5.36 | 96.26 | 144.46 | 97.13 |
|  | CL182.Contig8\_All | 0.06 | 0 | 0 | 0.12 | 0 | 0 |
|  | CL182.Contig9\_All | 0.83 | 11.04 | 1.86 | 5.11 | 10.68 | 7.73 |
|  | Unigene32983\_All | 5.85 | 30.56 | 29.35 | 94.44 | 41.48 | 100.33 |
|  | Unigene51359\_All | 7.64 | 0 | 2.14 | 0 | 0 | 0 |
|  | Unigene5424\_All | 4.02 | 0 | 2.37 | 0 | 0 | 0 |
|  | Unigene66592\_All | 2.2 | 0 | 0.68 | 0 | 0 | 0 |
| C4H | CL11402.Contig1\_All | 4.16 | 0 | 1.26 | 0 | 0 | 0 |
|  | CL11402.Contig2\_All | 118.27 | 95.83 | 82.1 | 386.72 | 424.88 | 336.27 |
|  | CL11402.Contig3\_All | 81.19 | 0.03 | 26.12 | 0.26 | 0.23 | 0 |
|  | CL11402.Contig4\_All | 14.66 | 28.48 | 26.92 | 78.64 | 108.32 | 137.43 |
|  | CL15243.Contig1\_All | 10.79 | 0 | 3.16 | 0 | 0 | 0 |
|  | Unigene2297\_All | 3.64 | 6.54 | 4.52 | 22.92 | 46.45 | 63.01 |
|  | Unigene33772\_All | 9.71 | 0.03 | 7.26 | 0.04 | 0 | 0 |
|  | Unigene37551\_All | 1.13 | 2.38 | 0.58 | 56.92 | 39.79 | 24.1 |
| 4CL | CL14470.Contig1\_All | 0.17 | 0.47 | 0.21 | 0.73 | 0 | 0.75 |
|  | CL14470.Contig2\_All | 0.06 | 0.06 | 0.16 | 0.28 | 0.17 | 0.85 |
|  | CL15670.Contig1\_All | 1.87 | 1.93 | 1.1 | 13.5 | 10.41 | 7.48 |
|  | CL15670.Contig2\_All | 11.62 | 21.31 | 9.98 | 88.26 | 70.56 | 56.61 |
|  | CL15670.Contig3\_All | 1.83 | 0 | 0.21 | 0 | 0 | 0 |
|  | CL3446.Contig1\_All | 0.83 | 0 | 0.34 | 0 | 0 | 0 |
|  | CL3446.Contig2\_All | 0.63 | 0 | 0 | 0 | 0 | 0 |
|  | CL3446.Contig3\_All | 0.46 | 0 | 2.47 | 0 | 0 | 0 |
|  | CL3699.Contig1\_All | 1.46 | 2.74 | 2.06 | 3.47 | 8.52 | 15.15 |
|  | CL3699.Contig2\_All | 0 | 0 | 0 | 1.35 | 3.79 | 2.95 |
|  | CL3699.Contig3\_All | 0.07 | 0.27 | 0.18 | 2.29 | 6.74 | 4.63 |
|  | CL4461.Contig14\_All | 8.12 | 0.06 | 2.04 | 0 | 0 | 0.07 |
|  | Unigene18504\_All | 0.03 | 0 | 0 | 0.77 | 2.87 | 0 |
|  | Unigene20468\_All | 0.5 | 0.34 | 0 | 0.18 | 0.85 | 0.54 |
|  | Unigene23016\_All | 4.21 | 14.36 | 5.08 | 44.12 | 40.32 | 35.3 |
|  | Unigene39130\_All | 2.66 | 0 | 1.07 | 0 | 0 | 0 |
|  | Unigene39527\_All | 0 | 0.02 | 0.03 | 6.57 | 2.72 | 2.63 |
|  | Unigene40991\_All | 8.51 | 0 | 5.18 | 0 | 0 | 0 |
|  | Unigene43108\_All | 3.47 | 20.98 | 3.47 | 147.72 | 244.59 | 97.25 |
|  | Unigene44601\_All | 1.81 | 0 | 1 | 0 | 0 | 0 |
|  | Unigene45693\_All | 0 | 0.28 | 0 | 0 | 0.43 | 0.35 |
|  | Unigene4923\_All | 0 | 0 | 0 | 0 | 0 | 0.45 |
|  | Unigene49930\_All | 11.22 | 0.02 | 3.19 | 0 | 0.03 | 0 |
|  | Unigene61411\_All | 0 | 0 | 0 | 0.14 | 0.27 | 0.97 |
|  | Unigene62822\_All | 0.43 | 0 | 0 | 0 | 0 | 0.72 |
|  | Unigene65195\_All | 0 | 0 | 0 | 1.36 | 0 | 0 |
|  | Unigene67615\_All | 2.05 | 0 | 0.36 | 0 | 0 | 0 |
|  | Unigene9215\_All | 0 | 0.51 | 0 | 0.77 | 0 | 0.76 |
| HCT/  HQT | Unigene51674\_All | 4.55 | 0.1 | 1.99 | 0 | 0 | 0 |
|  | CL1377.Contig3\_All | 1.64 | 0.81 | 0.49 | 4.85 | 3.78 | 1.21 |
|  | CL1377.Contig6\_All | 0.93 | 0.48 | 0.41 | 9.5 | 6.75 | 1.52 |
|  | CL1377.Contig7\_All | 4.5 | 1.06 | 1.48 | 11.71 | 11.5 | 5.74 |
|  | CL1377.Contig4\_All | 7.42 | 2.45 | 2.65 | 42.11 | 32.58 | 15.67 |
|  | CL1377.Contig5\_All | 22.7 | 18.04 | 18.51 | 118.15 | 153.89 | 67.49 |
|  | CL1377.Contig8\_All | 7.37 | 0 | 2.83 | 0 | 0 | 0 |
|  | Unigene33140\_All | 20.32 | 7.66 | 15.44 | 64.5 | 54.36 | 23.48 |
| C3'H | CL1401.Contig10\_All | 1.15 | 0 | 0 | 0 | 0 | 0 |
|  | CL1401.Contig11\_All | 8.58 | 0 | 3.71 | 0.07 | 0 | 0 |
|  | CL1401.Contig9\_All | 99.57 | 59.02 | 53.37 | 227.39 | 253.2 | 177.43 |

**Table S4.1.** Expression profiles of bZIPs involved in CGA biosynthesis of *Camellia sinensis* (DEGs)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Gene ID | log2  Fold | FPKM | | | | | |
| CaS4\_1 | CaS4\_2 | CaS4\_3 | CaS419\_1 | CaS419\_2 | CaS419\_3 |
| CL1582.Contig4\_All | -3.93 | 0.15 | 0.25 | 0.09 | 8.15 | 5.67 | 5.63 |
| CL1719.Contig6\_All | -8.69 | 0 | 0 | 0 | 4.37 | 7.24 | 8.83 |
| Unigene38402\_All | 1.47 | 16.67 | 34.85 | 16.66 | 19.21 | 19.41 | 18.16 |
| CL6969.Contig1\_All | -1.15 | 3.11 | 7.03 | 4.39 | 24.53 | 25.25 | 24.48 |
| CL1582.Contig7\_All | -2.26 | 0.58 | 1.63 | 0.99 | 11.69 | 14.2 | 9.19 |
| Unigene47994\_All | -1.27 | 2.12 | 5.46 | 3.08 | 18.29 | 19.97 | 20.07 |
| Unigene47644\_All | -1.91 | 1.63 | 4.22 | 1.79 | 23.04 | 15.63 | 26.98 |
| CL19430.Contig2\_All | 3.89 | 5.89 | 9.47 | 24.54 | 2.32 | 1.5 | 2.33 |
| CL15490.Contig1\_All | -1.76 | 1.26 | 1.96 | 1.01 | 10.79 | 12.06 | 11.33 |
| CL9937.Contig6\_All | -7.23 | 0 | 0 | 0 | 1.69 | 3.85 | 1.87 |
| CL2115.Contig3\_All | 3.00 | 16.81 | 10.12 | 17.07 | 5.69 | 3.7 | 4.78 |
| CL1582.Contig1\_All | -1.31 | 2.29 | 4.92 | 2.56 | 16.82 | 20.79 | 18.28 |
| CL10693.Contig2\_All | -1.64 | 1.37 | 4.25 | 1.38 | 14.59 | 16.99 | 16.99 |
| CL6554.Contig1\_All | 2.62 | 26.8 | 13.75 | 35.67 | 12.99 | 7.9 | 11.26 |
| CL3896.Contig1\_All | -1.96 | 0.23 | 0.52 | 0.45 | 3 | 3.36 | 4.9 |
| CL1582.Contig3\_All | -1.29 | 2.32 | 3.83 | 1.89 | 16.58 | 17.07 | 13.34 |
| Unigene17814\_All | 1.18 | 129.24 | 219.96 | 107 | 169.73 | 163.18 | 138.45 |
| CL2115.Contig10\_All | -4.14 | 0 | 0.28 | 0 | 8.42 | 2.5 | 2.4 |
| CL4955.Contig3\_All | 5.55 | 0.96 | 0.1 | 0.73 | 0 | 0 | 0 |
| CL11356.Contig1\_All | 5.67 | 2.55 | 0 | 2.23 | 0 | 0 | 0 |
| Unigene36335\_All | 5.57 | 2.46 | 0 | 1.55 | 0 | 0 | 0 |
| Unigene47885\_All | 1.93 | 66.89 | 40.84 | 54.2 | 29.04 | 47.12 | 34.6 |
| CL2985.Contig1\_All | 5.30 | 2.73 | 0 | 1.79 | 0 | 0 | 0 |
| CL19430.Contig1\_All | 1.76 | 2.29 | 10.05 | 2.31 | 2.29 | 2.52 | 3.86 |
| Unigene13167\_All | -3.51 | 0.04 | 0.17 | 0 | 3.26 | 2.77 | 0.74 |
| CL6668.Contig1\_All | 5.25 | 1.45 | 0 | 1.03 | 0 | 0 | 0 |
| CL11356.Contig2\_All | 5.18 | 2.4 | 0 | 2.08 | 0 | 0 | 0 |
| CL12795.Contig2\_All | 5.13 | 1.09 | 0 | 0.79 | 0 | 0 | 0 |
| Unigene23572\_All | 3.86 | 4.42 | 11.62 | 0.56 | 0.27 | 0 | 1.33 |
| Unigene51276\_All | 5.04 | 1.51 | 0 | 1.98 | 0 | 0 | 0 |
| Unigene40932\_All | 5.03 | 2.8 | 0 | 2.31 | 0 | 0 | 0 |
| CL6792.Contig4\_All | 4.83 | 5.64 | 0 | 2.65 | 0 | 0.15 | 0 |
| CL3896.Contig3\_All | 1.45 | 80.09 | 63.62 | 86.98 | 69.45 | 74.37 | 74.92 |
| CL7402.Contig1\_All | -1.29 | 0.97 | 2.5 | 0.78 | 8.66 | 7.43 | 7.64 |
| Unigene42588\_All | -1.10 | 1.29 | 5.15 | 1.96 | 12.59 | 11.82 | 15.09 |
| CL3516.Contig2\_All | 4.87 | 1.79 | 0 | 1.1 | 0 | 0 | 0 |
| CL2197.Contig6\_All | -1.40 | 0.57 | 1.34 | 0.56 | 7.05 | 3.34 | 4.56 |
| CL6668.Contig2\_All | 4.71 | 1.04 | 0 | 0.39 | 0 | 0 | 0 |
| Unigene17863\_All | 2.27 | 8.43 | 4.04 | 2.29 | 4.01 | 1.89 | 2 |
| CL3062.Contig4\_All | 2.13 | 3.48 | 5.1 | 2.23 | 3.98 | 0.78 | 0.75 |
| Unigene1618\_All | -1.05 | 0.53 | 2.34 | 1.17 | 6.18 | 6.29 | 5.9 |

**Table S4.2.** Expression profiles of ERFs involved in CGA biosynthesis of *Camellia sinensis* (DEGs)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Gene ID | log2  Fold | FPKM | | | | | |
| CaS4\_1 | CaS4\_2 | CaS4\_3 | CaS419\_1 | CaS419\_2 | CaS419\_3 |
| Unigene6814\_All | 5.35 | 92.16 | 60.32 | 364.97 | 12.35 | 5.6 | 9.3 |
| Unigene169\_All | 1.17 | 80.25 | 150.48 | 92.04 | 113.18 | 101.07 | 125.18 |
| Unigene26780\_All | 5.22 | 34.54 | 112.52 | 234.68 | 2.61 | 4.75 | 12.8 |
| CL1814.Contig1\_All | 1.69 | 26.92 | 46.86 | 29.86 | 24.49 | 28.94 | 22.03 |
| CL11203.Contig1\_All | 2.03 | 20.1 | 25.21 | 26.01 | 14.69 | 14.44 | 13.75 |
| CL1814.Contig9\_All | 1.86 | 22.31 | 34.96 | 37.84 | 20.07 | 20.28 | 23.1 |
| CL10933.Contig2\_All | 4.46 | 1334.11 | 727.68 | 4020.82 | 201.39 | 208.25 | 212.48 |
| CL12370.Contig1\_All | 2.37 | 123.21 | 119.11 | 107.9 | 37.6 | 62.83 | 69.42 |
| CL18066.Contig2\_All | 4.35 | 589.06 | 111.89 | 970.76 | 47.43 | 71.4 | 68.95 |
| Unigene17300\_All | 1.30 | 39.89 | 72.08 | 52.99 | 44.86 | 47.8 | 66.6 |
| Unigene353\_All | 3.65 | 504.52 | 155.16 | 558.18 | 57.53 | 90.3 | 95.32 |
| CL13646.Contig1\_All | 3.91 | 5.93 | 2.17 | 2.4 | 0.59 | 0.34 | 0.78 |
| CL12763.Contig1\_All | 4.18 | 96.79 | 16.38 | 107.23 | 15.56 | 7.75 | 5.4 |
| CL9918.Contig1\_All | 4.46 | 2.44 | 1.57 | 10.13 | 0.46 | 0.6 | 0.24 |
| CL5093.Contig1\_All | 3.85 | 39.69 | 8.54 | 58.03 | 5.67 | 7.74 | 4.21 |
| CL12370.Contig2\_All | 2.25 | 115.03 | 85.36 | 111.45 | 48.99 | 49.69 | 71.15 |
| CL14129.Contig2\_All | 3.05 | 109.34 | 44.78 | 157.04 | 35.88 | 29.17 | 31.26 |
| Unigene23426\_All | 4.49 | 4.73 | 1.6 | 12.96 | 0.18 | 0.44 | 1.11 |
| CL12790.Contig2\_All | 2.77 | 305.92 | 163.08 | 201.36 | 56.68 | 79.2 | 119.68 |
| Unigene37987\_All | 1.77 | 2.01 | 5.42 | 2.35 | 2.51 | 1.6 | 2.26 |
| CL6191.Contig1\_All | 3.03 | 532.1 | 163.65 | 320.62 | 98.03 | 110.1 | 119.65 |
| CL8014.Contig1\_All | 3.18 | 970.86 | 267.34 | 904.79 | 183.29 | 206 | 222.59 |
| Unigene27992\_All | 3.03 | 248.53 | 96.65 | 312.05 | 66.62 | 86.89 | 48.39 |
| CL14129.Contig1\_All | 2.58 | 80.19 | 45.13 | 131.33 | 35.35 | 37 | 37.69 |
| Unigene37190\_All | 2.19 | 10.05 | 6.76 | 10.88 | 6.4 | 4.36 | 4.94 |
| CL14130.Contig1\_All | 2.73 | 141.71 | 61.71 | 173.84 | 47.53 | 57.03 | 41.2 |
| CL6980.Contig2\_All | 1.11 | 5.79 | 16.17 | 10.53 | 11.28 | 12.22 | 10.55 |
| CL15678.Contig1\_All | 6.13 | 6.64 | 0.11 | 2.59 | 0 | 0 | 0 |
| CL5093.Contig6\_All | 2.76 | 152.64 | 76 | 303.47 | 72.1 | 54.3 | 72.99 |
| Unigene32916\_All | 3.36 | 9.07 | 4.57 | 4.24 | 2.18 | 1.51 | 0.48 |
| CL4784.Contig1\_All | 5.98 | 2.39 | 0.09 | 0.5 | 0 | 0 | 0 |
| CL5093.Contig4\_All | 3.26 | 72.69 | 18.08 | 92.09 | 10.16 | 26.29 | 9.79 |
| Unigene1663\_All | 3.36 | 55.76 | 26.8 | 175.47 | 10.37 | 11.79 | 36.92 |
| CL5506.Contig2\_All | 6.00 | 2.74 | 0 | 3.54 | 0 | 0 | 0 |
| Unigene765\_All | 3.19 | 309.09 | 71.1 | 522.75 | 91.9 | 68.23 | 83.84 |
| CL11199.Contig1\_All | 5.93 | 2.65 | 0 | 1.66 | 0 | 0 | 0 |
| CL6194.Contig6\_All | -2.25 | 0.57 | 0.8 | 0.72 | 9.7 | 10.56 | 4.85 |
| CL15678.Contig2\_All | 5.90 | 7.24 | 0 | 9.16 | 0 | 0 | 0 |
| Unigene20747\_All | 5.87 | 3.77 | 0 | 2.19 | 0 | 0 | 0 |
| CL1179.Contig1\_All | 5.86 | 3.64 | 0 | 2.18 | 0 | 0 | 0 |
| CL14129.Contig3\_All | 2.93 | 363.41 | 76.21 | 302.4 | 91.67 | 72.47 | 92.13 |
| CL5093.Contig3\_All | 2.98 | 22.58 | 5.34 | 32.51 | 6.02 | 8.04 | 5.23 |
| Unigene48655\_All | 2.36 | 406.91 | 177.74 | 308.49 | 127.52 | 146.31 | 189.86 |
| CL17933.Contig2\_All | 5.77 | 9.83 | 0 | 5.1 | 0 | 0 | 0 |
| CL1689.Contig3\_All | 1.48 | 2.16 | 6.27 | 5.42 | 3.6 | 3.91 | 3.74 |
| Unigene17637\_All | 4.08 | 8.81 | 3.03 | 2.77 | 1.28 | 0 | 0.49 |
| CL6194.Contig4\_All | 5.72 | 4.33 | 0 | 2.18 | 0 | 0 | 0 |
| CL15822.Contig1\_All | 5.71 | 3.51 | 0 | 2.1 | 0 | 0 | 0 |
| CL20194.Contig1\_All | 2.89 | 135.11 | 44.9 | 147.63 | 40.82 | 13.67 | 59.89 |
| CL18243.Contig3\_All | 5.69 | 4.35 | 0 | 4.22 | 0 | 0 | 0 |
| Unigene42137\_All | 1.90 | 72.13 | 54.37 | 48.77 | 27.2 | 44.51 | 48.86 |
| Unigene47621\_All | 2.76 | 1.36 | 1.04 | 2.18 | 0.63 | 0.54 | 0.43 |
| CL5093.Contig7\_All | 2.73 | 309.32 | 119.68 | 459.33 | 88.96 | 167.66 | 76.97 |
| CL9918.Contig2\_All | 3.33 | 1.99 | 1.51 | 7.73 | 0.95 | 1.15 | 0.3 |
| CL17981.Contig1\_All | -2.29 | 0.67 | 0.56 | 1.06 | 9.82 | 13.43 | 6.7 |
| Unigene29792\_All | -4.36 | 0 | 0 | 0.1 | 2.12 | 4.81 | 2.09 |
| CL14454.Contig1\_All | 1.29 | 11.01 | 39.84 | 18.59 | 16.39 | 25.85 | 18.68 |
| CL18243.Contig2\_All | 5.44 | 1.9 | 0 | 2.21 | 0 | 0 | 0 |
| CL15764.Contig1\_All | 5.45 | 2.33 | 0 | 1.23 | 0 | 0 | 0 |
| CL17933.Contig1\_All | 5.39 | 3.32 | 0 | 1.43 | 0 | 0 | 0 |
| Unigene46498\_All | 5.39 | 4.62 | 0 | 1.78 | 0 | 0 | 0 |
| Unigene30706\_All | 5.22 | 5.1 | 0 | 2.1 | 0 | 0 | 0 |
| Unigene5125\_All | 5.13 | 2.31 | 0 | 2.45 | 0 | 0 | 0 |
| CL2574.Contig7\_All | -1.28 | 2.5 | 3.97 | 2.61 | 14.32 | 23.94 | 14.7 |
| Unigene3453\_All | 5.12 | 1.73 | 0 | 1.34 | 0 | 0 | 0 |
| CL5506.Contig1\_All | 5.10 | 1.82 | 0 | 0.8 | 0 | 0 | 0 |
| Unigene40018\_All | 5.01 | 1.32 | 0 | 0.85 | 0 | 0 | 0 |
| Unigene47929\_All | 2.44 | 221.42 | 49.06 | 292.52 | 111.52 | 82.76 | 75.61 |
| CL15577.Contig3\_All | 2.65 | 1.52 | 1.19 | 1.07 | 0.37 | 0.62 | 0.47 |
| CL1689.Contig2\_All | 2.68 | 1.46 | 0.84 | 0.4 | 0.4 | 0.31 | 0.33 |
| CL5093.Contig10\_All | 2.56 | 13.31 | 3.11 | 17.71 | 8 | 2.67 | 4.16 |
| CL9382.Contig2\_All | 2.43 | 200.54 | 41.45 | 79.34 | 53.47 | 61.3 | 47.16 |
| CL12763.Contig2\_All | 2.45 | 88.33 | 15 | 86.89 | 25.08 | 34.61 | 32.46 |
| CL17981.Contig2\_All | -1.80 | 1.3 | 0.7 | 0.84 | 9.13 | 11.57 | 7.17 |
| Unigene21588\_All | 2.76 | 54.56 | 7.58 | 17.66 | 18.41 | 6.3 | 6.31 |
| CL12467.Contig3\_All | 4.57 | 2.61 | 0 | 2.63 | 0 | 0.14 | 0.01 |
| CL20194.Contig2\_All | 2.07 | 228.94 | 85.96 | 170.19 | 73.76 | 77.78 | 162.51 |
| Unigene25943\_All | 4.56 | 4.16 | 0 | 2.35 | 0 | 0.16 | 0 |
| Unigene8159\_All | -3.39 | 0.46 | 0 | 0.56 | 6.6 | 29.68 | 6.67 |
| Unigene27359\_All | -2.10 | 0.45 | 0.48 | 0.84 | 10.98 | 4.71 | 4.72 |
| CL5633.Contig2\_All | 1.81 | 3.55 | 2.55 | 6.55 | 2.83 | 2.81 | 3.59 |
| Unigene27884\_All | 2.37 | 1.09 | 1.51 | 1.92 | 0.26 | 0.5 | 1.19 |
| CL5532.Contig1\_All | 4.63 | 3.62 | 0 | 1.12 | 0 | 0 | 0 |
| CL3592.Contig3\_All | 4.58 | 0.8 | 0 | 0.5 | 0 | 0 | 0 |
| Unigene43269\_All | 2.17 | 6.37 | 2.64 | 9.92 | 3.3 | 2.37 | 5.28 |
| Unigene43958\_All | -2.35 | 0.07 | 0.47 | 0.09 | 2.99 | 3.01 | 2.07 |
| Unigene39443\_All | 4.48 | 0.69 | 0 | 1.86 | 0 | 0 | 0 |
| Unigene46771\_All | 4.45 | 1.67 | 0 | 0.74 | 0 | 0 | 0 |

Table S4.3 Expression profiles of MYBs involved in CGA biosynthesis of *Camellia sinensis* (DEGs)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Gene ID | log2  Fold | FPKM | | | | | |
| CaS4\_1 | CaS4\_2 | CaS4\_3 | CaS419\_1 | CaS419\_2 | CaS419\_3 |
| CL5853.Contig5\_All | -2.25 | 0.36 | 0.71 | 0.36 | 5.14 | 5.59 | 5.3 |
| CL1521.Contig1\_All | -4.55 | 0.29 | 0.32 | 0.58 | 27.44 | 32.12 | 15.5 |
| CL12111.Contig1\_All | 3.09 | 19.01 | 89.83 | 47.14 | 10.62 | 8.88 | 18.94 |
| CL992.Contig8\_All | -4.04 | 0.03 | 0.17 | 0.02 | 2.39 | 3.71 | 2.42 |
| CL1588.Contig2\_All | 2.31 | 5.39 | 16.78 | 12.56 | 4.58 | 5.44 | 5.63 |
| CL14383.Contig5\_All | -3.28 | 0.95 | 2.11 | 0.7 | 28.36 | 41.27 | 16.67 |
| CL15867.Contig2\_All | 2.42 | 86.12 | 81.79 | 87.86 | 45.51 | 31.89 | 45.16 |
| CL5853.Contig8\_All | -2.39 | 0.3 | 1.33 | 0.84 | 8.52 | 11.53 | 8.77 |
| Unigene11596\_All | -1.87 | 0.34 | 1.31 | 0.61 | 5.37 | 6.7 | 5.99 |
| CL19128.Contig2\_All | -2.07 | 0.93 | 2.74 | 1.41 | 16.5 | 21.92 | 9.65 |
| CL1588.Contig1\_All | 2.22 | 5.19 | 9.87 | 11.08 | 5.44 | 2.44 | 5.42 |
| CL15867.Contig1\_All | 2.25 | 44.34 | 29.86 | 51.31 | 25.76 | 22.3 | 22.04 |
| CL5853.Contig2\_All | -1.99 | 0.18 | 0.44 | 0.39 | 2.16 | 4.23 | 3.12 |
| CL8818.Contig2\_All | 2.47 | 1.59 | 2.29 | 3.53 | 1.23 | 0.84 | 1.13 |
| Unigene47413\_All | -2.92 | 0.09 | 0.2 | 0.15 | 3.83 | 2.1 | 2.87 |
| CL18738.Contig1\_All | -2.89 | 0.41 | 0.24 | 0.36 | 8.98 | 8.96 | 3.47 |
| Unigene53317\_All | 3.10 | 180.17 | 62.74 | 317.72 | 43.68 | 68.25 | 49.01 |
| CL16222.Contig1\_All | 3.47 | 7.37 | 2.41 | 12.03 | 1.96 | 1.45 | 1.24 |
| CL5099.Contig3\_All | 3.40 | 2.01 | 9.99 | 17.12 | 1.36 | 0.7 | 3.47 |
| CL16222.Contig2\_All | 4.25 | 1.54 | 1.34 | 6.19 | 0.17 | 0 | 0.64 |
| CL3571.Contig1\_All | -1.83 | 1.46 | 1.87 | 2.84 | 19.29 | 22.42 | 14.24 |
| Unigene37024\_All | 2.53 | 58.26 | 22.24 | 85.09 | 20.36 | 28.32 | 25.03 |
| Unigene6639\_All | -2.87 | 0.27 | 2.11 | 0.41 | 23.64 | 16.63 | 5.24 |
| CL14383.Contig2\_All | -2.89 | 0.29 | 0.7 | 0.14 | 3.05 | 13.64 | 4.26 |
| CL8310.Contig2\_All | 1.38 | 38.83 | 39.75 | 25.72 | 34.71 | 30.55 | 35.93 |
| CL11240.Contig4\_All | -1.20 | 1.11 | 4.4 | 1.54 | 11.52 | 11.92 | 11.26 |
| Unigene42615\_All | -3.39 | 0.04 | 0.1 | 0.05 | 3.61 | 1.2 | 0.95 |
| CL14525.Contig3\_All | -1.10 | 2.76 | 10.25 | 5.77 | 25.48 | 26.87 | 37.71 |
| CL16246.Contig2\_All | -1.96 | 0.19 | 1.34 | 0.95 | 7.34 | 9.41 | 5.1 |
| Unigene22231\_All | -2.71 | 0.94 | 0.21 | 0.62 | 15.01 | 16.94 | 5.33 |
| Unigene28432\_All | -2.01 | 0.2 | 0.85 | 0.63 | 6.05 | 5.31 | 5 |
| CL11240.Contig3\_All | -1.40 | 0.56 | 0.98 | 0.38 | 4.44 | 4.62 | 3.01 |
| Unigene43679\_All | 2.36 | 2.28 | 2.58 | 2.17 | 0.75 | 1.46 | 0.97 |
| CL15576.Contig3\_All | -3.34 | 0 | 0.43 | 0 | 4.46 | 3.79 | 2.31 |
| CL145.Contig3\_All | 3.39 | 0.75 | 2.32 | 5.1 | 0.87 | 0 | 0.54 |
| CL14383.Contig1\_All | -2.00 | 1.79 | 4.61 | 2.44 | 20.62 | 51.35 | 10.63 |
| CL13577.Contig2\_All | 2.46 | 1.65 | 2.41 | 5.15 | 0.62 | 0.9 | 2.26 |
| Unigene1363\_All | -1.60 | 0.13 | 1.15 | 0.41 | 3.07 | 3.82 | 3.75 |
| CL10156.Contig1\_All | -1.49 | 0.17 | 0.52 | 0.28 | 1.39 | 1.62 | 3.23 |
| CL10156.Contig6\_All | -1.92 | 0.04 | 0.23 | 0.06 | 0.62 | 1.25 | 1.04 |
| Unigene27955\_All | 1.13 | 23.86 | 25.8 | 17.13 | 27.77 | 27.73 | 20.52 |

**Table S4.4.** Expression profiles of WRKYs involved in CGA biosynthesis of *Camellia sinensis* (DEGs)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gene ID | log2  Fold | FPKM | | | | | | | | | | | |
| CaS4\_1 | | CaS4\_2 | | CaS4\_3 | | CaS419\_1 | | CaS419\_2 | | CaS419\_3 | |
| CL2401.Contig7\_All | -1.71 | 1.27 | | 3.39 | | 1.76 | | 15.5 | | 15.87 | | 16.57 | |
| CL20334.Contig2\_All | -3.86 | 0.78 | | 1.87 | | 1.15 | | 22.89 | | 73.03 | | 36.94 | |
| CL4331.Contig11\_All | -2.30 | 1.6 | | 2.45 | | 1.4 | | 22.33 | | 25.95 | | 16.61 | |
| CL17686.Contig2\_All | -1.69 | 0.64 | | 1.33 | | 1.15 | | 7.26 | | 8.03 | | 8.29 | |
| CL15614.Contig4\_All | -2.70 | 1.37 | | 1.69 | | 0.82 | | 24.4 | | 25.2 | | 13.95 | |
| CL4331.Contig5\_All | -3.06 | 0.15 | | 0.15 | | 0.12 | | 2.13 | | 3.56 | | 3.32 | |
| CL4331.Contig15\_All | -1.23 | 0.98 | | 2.36 | | 1.25 | | 8.77 | | 8.11 | | 7.88 | |
| CL4331.Contig9\_All | -2.25 | 0.22 | | 0.44 | | 0.15 | | 3.63 | | 3.12 | | 2.58 | |
| Unigene7588\_All | 1.51 | 34.83 | | 77.61 | | 27.03 | | 35.57 | | 29.8 | | 45.24 | |
| CL2401.Contig6\_All | -1.71 | 1.36 | | 5.84 | | 3.55 | | 28.47 | | 28.09 | | 21.04 | |
| CL8604.Contig1\_All | -1.56 | 5.57 | | 6.68 | | 6.25 | | 44.95 | | 55.44 | | 37.27 | |
| CL9141.Contig2\_All | 2.56 | 6.72 | | 3.69 | | 8.21 | | 3.49 | | 2.43 | | 2.21 | |
| CL3826.Contig11\_All | 3.34 | 4.77 | | 1.99 | | 8.23 | | 1.88 | | 0.99 | | 0.68 | |
| CL15614.Contig3\_All | -5.69 | 0 | 0 | | 0 | | 2.31 | | 1.84 | | 0.88 | |
| Unigene1290\_All | -2.45 | 0.45 | 0.52 | | 0.21 | | 6.64 | | 6.06 | | 4.07 | |
| CL2789.Contig2\_All | 6.06 | 4.75 | 0 | | 3.8 | | 0 | | 0 | | 0 | |
| Unigene36171\_All | 6.02 | 5.57 | 0 | | 3.98 | | 0 | | 0 | | 0 | |
| CL17467.Contig14\_All | -2.29 | 2.72 | 3.86 | | 4.96 | | 36 | | 83.19 | | 25.93 | |
| CL15804.Contig2\_All | -2.54 | 0.51 | 0.35 | | 0.31 | | 7.71 | | 7.63 | | 3.39 | |
| CL3034.Contig3\_All | 5.66 | 2.56 | 0 | | 1.6 | | 0 | | 0 | | 0 | |
| CL11371.Contig2\_All | 2.62 | 1 | 3.18 | | 1.35 | | 0.26 | | 0.88 | | 0.69 | |
| Unigene38827\_All | -2.56 | 0.07 | 0.42 | | 0.04 | | 2.79 | | 1.95 | | 2.74 | |
| CL9098.Contig2\_All | -2.82 | 0.18 | 0.43 | | 0.15 | | 7.84 | | 3.92 | | 2.22 | |
| CL9141.Contig3\_All | 2.17 | 14.16 | 6.03 | | 7.83 | | 5.08 | | 5.77 | | 5.79 | |
| CL14366.Contig3\_All | -5.09 | 0 | 0 | | 0 | | 0.23 | | 1.61 | | 1.41 | |
| CL16381.Contig1\_All | 1.45 | 216.71 | 200.05 | | 236.35 | | 182.1 | | 208.06 | | 219.67 | |
| Unigene3958\_All | 5.32 | 3.58 | 0 | | 1.04 | | 0 | | 0 | | 0 | |
| CL19227.Contig1\_All | 5.29 | 2 | 0 | | 2.81 | | 0 | | 0 | | 0 | |
| CL4331.Contig10\_All | -1.89 | 0.16 | 0.46 | | 0.08 | | 2.09 | | 1.85 | | 2.29 | |
| Unigene22121\_All | -1.93 | 0.24 | 0.93 | | 0.44 | | 7.48 | | 3.43 | | 3.04 | |
| CL4331.Contig1\_All | -4.92 | 0 | 0 | | 0 | | 0.36 | | 0.88 | | 0.54 | |
| Unigene29518\_All | 5.16 | 1.36 | 0 | | 2.62 | | 0 | | 0 | | 0 | |
| Unigene21677\_All | -2.33 | 2.55 | 1.23 | | 3.12 | | 37.69 | | 48.87 | | 14.86 | |
| Unigene30959\_All | 5.13 | 1.88 | 0 | | 2.45 | | 0 | | 0 | | 0 | |
| Unigene40916\_All | 5.08 | 1.52 | 0 | | 2.2 | | 0 | | 0 | | 0 | |
| Unigene20428\_All | 5.05 | 3.79 | 0 | | 2.25 | | 0 | | 0 | | 0 | |
| CL3034.Contig1\_All | 4.97 | 0.91 | 0 | | 0.65 | | 0 | | 0 | | 0 | |
| CL12606.Contig2\_All | -1.85 | 1.34 | 1.38 | | 0.91 | | 13.64 | | 14.97 | | 5.54 | |
| CL18520.Contig1\_All | -2.74 | 0.33 | 0.25 | | 0.03 | | 6.57 | | 3.76 | | 1.85 | |
| CL14746.Contig4\_All | 4.85 | 0.98 | 0 | | 0.63 | | 0 | | 0 | | 0 | |
| Unigene38332\_All | -1.69 | 0.58 | 3.61 | | 1.25 | | 9.98 | | 17.39 | | 10.1 | |
| CL4331.Contig14\_All | -1.20 | 1.13 | 3.27 | | 0.81 | | 8.04 | | 9.72 | | 8.59 | |
| Unigene16505\_All | 1.89 | 20.4 | 9.97 | | 15.56 | | 17.32 | | 8.08 | | 7.57 | |
| CL7135.Contig1\_All | -1.62 | 9.25 | 4.92 | | 7.14 | | 69.11 | | 65.83 | | 50.65 | |
| CL21017.Contig2\_All | 1.32 | 8.7 | 8.81 | | 7.73 | | 10.21 | | 5.71 | | 9.7 | |
| CL8055.Contig1\_All | -2.14 | 0.09 | 0.42 | | 0.34 | | 3.93 | | 3.57 | | 1.51 | |
| Unigene42682\_All | -1.07 | 19.94 | 28.82 | | 17.13 | | 133.13 | | 125.47 | | 79.08 | |
| CL2265.Contig1\_All | 4.49 | 2.05 | 0 | | 0.85 | | 0 | | 0 | | 0 | |
| CL6976.Contig1\_All | 2.22 | 10 | 3.37 | | 19.96 | | 9.78 | | 4.12 | | 4.33 | |