

---

# Insights into the Metabolite Profiles of Two Camellia (Theaceae) Species in Yunnan Province through Metabolomic and Transcriptomic Analysis

---

[Miao Niu](#) , Ranyang Li , Xiongyu Li , [Hongyan Yang](#) , Jianliang Ding , Xianxiu Zhou , [Yuqi He](#) , Yawen Xu , Qian Qu , Zhiwei Liu , [Jiahua Li](#) \*

Posted Date: 29 July 2024

doi: 10.20944/preprints202407.2207.v1

Keywords: tea species evaluation; Camellia sinensis; Camellia taliensis; metabolome; transcriptome



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

# Insights into the Metabolite Profiles of Two *Camellia* (Theaceae) Species in Yunnan Province through Metabolomic and Transcriptomic Analysis

Miao Niu <sup>1</sup>, Ranyang Li <sup>2</sup>, Xiongyu Li <sup>1</sup>, Hongyan Yang <sup>1</sup>, Jianliang Ding <sup>1</sup>, Xianxiu Zhou <sup>1</sup>, Yuqi He <sup>1</sup>, Yawen Xu <sup>3</sup>, Qian Qu <sup>1</sup>, Zhiwei Liu <sup>1</sup> and Jiahua Li <sup>1,\*</sup>

<sup>1</sup> College of Tea Science, Yunnan Agricultural University, Kunming 650201, China; 13577036525@163.com (M.N.); lxy19912992159@163.com (X.L.); 13529725893@163.com (H.Y.); 13937016027@163.com (J.D.); 2023210231@stu.ynau.edu.cn (X.Z.); 13578128508@139.com (Y.H.); 13658334948@163.com (Q.Q.); 2018016@ynau.edu.cn (Z.L.)

<sup>2</sup> College of Horticulture, Hunan Agricultural University, Changsha 410128, China; 13529036009@163.com

<sup>3</sup> College of Pu-Erh Tea, West Yunnan University of Applied Sciences, Puer 665000, China; 1575859889@163.com

\* Correspondence: 15087172337@163.com; Tel.: +86-15087172337

**Abstract:** Tea (*Camellia* L.) falls into the family Theaceae, is a valuable commercial crop, and tea products made from its buds and young leaves are favored by consumers all over the world. The more common *Thea* plant is *Camellia sinensis* (*C. sinensis*), but its most important relative, *Camellia taliensis* (*C. taliensis*), is also utilized by locals in the area of cultivation to manufacture tea. In this investigation, *C. taliensis* (DL) and *C. sinensis* (QJZ) were characterized in terms of their agronomic traits, physicochemical indices, metabolomics, and transcriptomics. The leaf area of DL is larger than that of QJZ; the color of DL's buds and leaves is yellowish-green, while that of QJZ's is green. DL's buds and leaves are more densely velvety than those of QJZ. The HPLC results indicated that the physicochemical contents varied considerably between the two samples, with DL having greater concentrations of EGCG and GABA than QJZ, while QJZ had remarkably higher concentrations of C, CA, and EGC than DL. A total of 2269 metabolites and 362,190,414 genes were positively identified, with the number of DAMs and DEGs being 1001 and 34,026, respectively. The flavonoids, phenolic acids, and alkaloid metabolites were dramatically different between the two tea group plants. Bioinformatics profiling revealed that the DAMs and DEGs of the two tea group plants interacted with each other and were involved in metabolic pathways including "biosynthesis of secondary metabolites", "biosynthesis of amino acids", "biosynthesis of cofactors", "phenylpropanoid biosynthesis", and "flavonoid biosynthesis". Overall, these results provide statistical support for germplasm conservation and production for both *C. taliensis* and *C. sinensis*.

**Keywords:** tea species evaluation; *Camellia sinensis*; *Camellia taliensis*; metabolome; transcriptome

## 1. Introduction

Tea (*Camellia* L.), belonging to the family Theaceae in the section *Thea*, is a type of perennial woody evergreen plant, widespread in tropical and subtropical regions of the world, particularly in Asia [1–3]. The tea plant is thought to have originated in Yunnan Province in southwest China, according to research and evidence [4–6]. More than 50 countries and regions around the world cultivate it as a valuable economic crop [7]. Tea made from the tender buds and young leaves of the tea plant through specific processing procedures is a purely natural, green, healthy beverage. Its rich aroma, warm taste, wonderful flavor, and variety of health benefits are well received by consumers around the world [8–10]. Tea contains many nutrients and is a rich source of essential amino acids, antioxidants, phenols, sugars, vitamins, and minerals that the human body demands [11].

The genus *Camellia* is a large family that consists of more than 110 taxonomic groups, of which *Camellia sinensis* (*C. sinensis*) is the most commonly grown tea tree and typically used commercially as a source of beverage tea [12]. *C. sinensis* produces more mutations with varying leaf sizes, and

cultivated plants usually have smaller leaves than wild native species. There are three variants of *C. sinensis*, including *Camellia sinensis* var. *assamica* (Masters) Kitamura, *Camellia sinensis* var. *publimba* Chang, var. nov., and *Camellia sinensis* var. *waldensae* (S.Y. Hu) Chang. Also, several species of the congeneric genus of *Camellia* section *Thea* are used to make tea locally [13]. *Camellia taliensis* (W.W. Sm.) Melch, named “Dali Cha” in Chinese, known as “wild tea”, is one of the most prominent wild relatives of *C. sinensis* and is an irreplaceable germplasm resource [14]. Because of its abundance of inclusions, its leaves are widely used by locals to make tea [15]. However, very scarce research has been performed on *C. taliensis*; thus, the biosynthesis pathways and related regulatory gene mechanisms of the bioactive components of *C. sinensis* and *C. taliensis* have not yet been clarified.

Metabolomics is a novel discipline following genomics and proteomics, which can directly reflect the biological processes and mechanisms during a specific physiological period [16]. From 2000 to the present time, metabolomics has attracted increasing attention and has been widely used in the tea industry in relation to the cultivation, processing, and classification of varieties, grades, and origins [17–20]. Identifying transcripts with functional information in plant genomes can be achieved through high-throughput transcriptome analysis [21]; this method has been extensively applied in relevant studies of tea germplasm resources [22]. Moreover, the accelerating development of metabolomics and transcriptomics association technology has boosted gene expression profiling involving different metabolite pathways to excavate key functional genes [15].

In this study, we first observed and recorded the phenotypic traits of DL and QJZ, measured the routine physicochemical indices, and combined metabolomic and transcriptomic analysis to explore the molecular processes of metabolite accretion and differences in DL and QJZ. The outcomes of the investigation are intended to provide partial data support for the further exploitation and conservation of relevant germplasm resources, the expansion of production, and the enhancement of the commercial value of tea plants.

## 2. Materials and Methods

### 2.1. Plant Materials

The two tea species (*C. taliensis* and *C. sinensis*) used in this study were obtained from the tea garden of Pu'er Lancang Ancient Tea Co., Ltd., Pu'er, Yunnan province, China (22°47' N, 100°54' E). The tea shoots (one bud with two leaves) of each tea plant were plucked at 10 AM on the 4th of May. In order to analyze the metabolomes and transcriptomes of all shoots, they were refrigerated in liquid nitrogen instantly and preserved at -80 °C.

### 2.2. Observation and Determination of Morphological Characteristics

Based on the *Descriptors and Data Standards for Tea (Camellia spp.)* [23], and the Industry Standard of China NY/T 1312-2007 [24], 13 phenotypic traits were selected for the statistics of DL and QJZ, including plant shape, growth habit, leaf size, leaf shape, bud color, bud pubescence, bud brightness, tenderness, leaf texture, leaf upper surface, leaf margin, density of leaf serration, and sharpness of leaf serration.

### 2.3. Determination of Tea Physicochemical Indices

National Standard of China GB/T 8303-2013 was followed when processing tea samples (drying, grinding, screening)[25]. The aqueous extraction, tea polyphenols, free amino acids, and catechins were measured according to previous reports. Three replications were performed for each experiment. Refer to the subsections below for further details.

#### 2.3.1. Determination of Aqueous Extraction

Based on previous research [26], in a 500 ml conical flask, 3 g (to the accuracy of 0.001 g) of the tea powder was transferred and 300 ml of boiling water was added. Following this, the conical flask was transferred to a boiling water bath for 45 minutes (shaking every 10 minutes). A vacuum filter

was used to filter the hot mixture after extraction. The tea residue was rinsed several times with about 150 ml of boiling water and then the tea residue and a known mass of filter paper were transferred into an aluminum box and placed together in a constant temperature oven. The oven temperature was set at 120 °C. The aluminum box was opened and tilted to one side. 1 hour later, the lid was closed, removed and cooled for 1 hour, then placed the box back in the oven for 1 hour. The aqueous extract content was computed using the formula (1).

$$\text{Aqueous extract content} = \left(1 - \frac{m_1}{m_0 \times \omega}\right) \times 100\% \quad (1)$$

$m_0$ —the weight of the tea powder, g;

$m_1$ —the weight of the dried tea residue, g;

$\omega$ —the dry matter content of the tea sample (mass fraction), %.

### 2.3.2. Determination of Total Tea Polyphenols

Folin-Ciocalteu method was used to measure the total tea polyphenol content [27]. For the extraction, 0.2 g of tea powder was placed in an extraction tube with 5 ml of 70% methanol, which was extracted for 10 minutes at 70°C with shaking every 5 minutes. The methanolic extract was analyzed for tea polyphenols using a Metash (Shanghai, China) X-5 spectrophotometer at 765 nm, and the total phenolic content was calculated from the gallic acid standard curve. The results were expressed as gallic acid milligram equivalent per 100 g dry weight (GAE mg/100 g). The total tea polyphenol content was computed using the formula (2).

$$\text{Tea polyphenols} = \frac{A \times V \times d}{\text{SLOPE}_{\text{std}} \times m \times 10^6 \times \omega} \times 100 \quad (2)$$

$A$ —absorbance of test solution;

$V$ —volume of sample extract, mL;

$d$ —dilution factor;

$\text{SLOPE}_{\text{std}}$ —slope of the standard curve for gallic acid;

$\omega$ —the dry matter content of the tea sample (mass fraction), %

$m$ —the weight of the tea sample, g.

### 2.3.3. Determination of Total Free Amino Acids

The content of total free amino acids was characterized using the National Standard of China GB/T 8304-2013 [28]. In simple terms, the prepared tea infusion (1 mL), phosphate buffer (1/15MNa<sub>2</sub>HPO<sub>4</sub>, 1/15MKH<sub>2</sub>PO<sub>4</sub>, pH 8.0) (0.5 mL), and reaction solution [2% ninhydrin (C<sub>9</sub>H<sub>4</sub>O<sub>3</sub>·H<sub>2</sub>O)] (0.5 mL) were put into a 25 mL volumetric flask in sequence, and then boiled for 15 min. After cooling, the volume was fixed with distilled water. Absorbance values were read at 570 nm using a Metash X-5 spectrophotometer. The total free amino acid content was computed using the formula (3).

$$\text{Total amino acids} = \frac{C/1000 \times V_1/V_2}{m \times w} \times 100\% \quad (3)$$

$C$ —the theanine weight, which could be obtained according to the OD<sub>570</sub> from a standard curve made by theanine as a standard component, using the same method as mentioned above, mg;

$V_1$ —the total volume of the tea infusion, mL;

$V_2$ —the volume of the infusion taken to reaction, mL;

$m$ —the dry weight of the tea sample, g;

$w$ —the dry ratio of the tea sample, %.

### 2.3.4. Quantification of Catechins and Caffeine via HPLC

Catechins and caffeine were determined by high-performance liquid chromatography (HPLC, Agilent Technologies Inc., CA, America), the adjusted program based on previous studies was as follows [29]: Weigh 0.1g of ground tea sample, add 70% methanol (10 mL), refrigerate for 24 h,

centrifuge the solution at 4500 rpm for 10 min, and filter through a 0.45  $\mu\text{m}$  membrane to obtain the solution to be tested. The column used for HPLC analysis was a Poroshell 120 EC-C18 (4.6  $\times$  5 mm, 2.7  $\mu\text{m}$ , Agilent Technologies Inc., CA, America).

The two mobile phases used in the experiments were solvent A (0.261% phosphoric acid, 5% acetonitrile) and solvent B (80% methanol). The elution gradient was 0–16 min, 92% A; 6–22 min, 50% A; 22–25.9 min, 35% A; 25.9–29 min, 0% A; 29–30 min, 0% A; program stopped for 30 min and then run for 6 min at 92% A.

### 2.3.5. Quantification of Amino Acids using HPLC

Amino acids were determined by high-performance liquid chromatography (HPLC, Agilent Technologies Inc., CA, America), the adjusted program based on previous studies was used, as follows [30]: Weigh 1 g of ground tea sample, add 80 mL of boiling water, 80  $^{\circ}\text{C}$  water bath extraction for 1 h; after cooling, filter to 100 mL volumetric flask capacity, take 1 mL of the tea broth and add 200  $\mu\text{L}$  chloroform oscillation extraction; maintain for 10 min, centrifuge at 12,000 rpm for 10 min, collect the supernatant, and filter through a 0.45  $\mu\text{m}$  membrane to obtain the solution to be tested. The column used for HPLC analysis was Venusil AA (4.6  $\times$  250 mm, 5  $\mu\text{m}$ , Agilent Technologies Inc., CA, America). The two mobile phases used in the experiments were solvent A (20  $\mu\text{M}$  sodium acetate, 0.018% triethylamine, 0.3% tetrahydrofuran, 0.004% ethylenediaminetetraacetic acid disodium salt dihydrate, pH 7.2) and solvent B (40% acetonitrile, 40% methanol, 0.3% acetic acid). The elution gradient was 0–2 min, 78% A; 2–12 min, 73% A; 12–35 min, 70% A; 35–39 min, 0% A; 39–39.9 min, 0% A; program was stopped for 40 min and then run for 6 min at 92% A.

## 2.4. Widely Targeted Metabolite Identification and Quantification

### 2.4.1. Dry Sample Extraction

Using vacuum freeze-drying technology, the biological samples were placed in a lyophilizer (Scientz-100F, Scientz, Ningbo, China) and then the samples were ground (30 Hz, 1.5 min) to powder using a grinder (MM 400, Retsch, Verder Shanghai Instruments and Equipment Co., Ltd., Shanghai, China). Next, 50 mg of sample powder was weighed and 1200  $\mu\text{L}$  of  $-20$   $^{\circ}\text{C}$  pre-cooled 70% methanolic aqueous internal standard extract was added (less than 50 mg added at the rate of 1200  $\mu\text{L}$  extractant per 50 mg sample). The sample was vortexed once every 30 min for 30 s, for a total of 6 times. After centrifugation (rotation speed 12,000 rpm, 3 min), the supernatant was aspirated, and the sample was filtered through a microporous membrane (0.22  $\mu\text{m}$  pore size) and stored in an injection vial for UPLC-MS/MS analysis.

### 2.4.2. UPLC Conditions

The sample extracts were analyzed using an UPLC-ESI-MS/MS system (UPLC, ExionLC<sup>TM</sup> AD, <https://sciex.com.cn/> (accessed on 8 August 2023)) and a tandem mass spectrometry system (MS/MS). The analytical conditions were as follows: UPLC column, Agilent SB-C18 (1.8  $\mu\text{m}$ , 2.1 mm  $\times$  100 mm). The mobile phase consisted of solvent A: pure water with 0.1% formic acid, and solvent B: acetonitrile with 0.1% formic acid. The elution gradient was 0–9 min, 5% A and maintained for 1 min; 10–11.1 min, 95% A and maintained for 2.9 min.

### 2.4.3. ESI-Q TRAP-MS/MS

The ESI source operation parameters were as follows: source temperature 500  $^{\circ}\text{C}$ ; ion spray voltage (IS) 5500 V/–4500 V; ion source gas I, gas II, and curtain gas were set at 50, 60, and 25 psi, respectively; the collision-activated dissociation (CAD) was high. Triple-quadrupole (QQQ) scans were acquired as multiple reaction monitoring (MRM) experiments with collision gas (nitrogen) set to medium. Declustering potential (DP) and collision energy (CE) for individual MRM transitions was carried out with further DP and CE optimization. A specific set of MRM transitions was monitored for each period according to the metabolites eluted within this period.

#### 2.4.4. Data Analysis

Based on the local metabolic database, the metabolites of the samples were qualitatively and quantitatively analyzed using mass spectrometry. The MRM metabolite detection multi-peak map shows the substances that can be detected in the sample, with each color-coded mass spectrum peak representing one metabolite detected. Metabolites with VIP (variable importance) >1, fold change  $\geq 2$ , and fold change  $\leq 0.5$  were considered to have undergone significant changes.

#### 2.5. RNA Sequencing (RNA-seq) Analysis and Differentially Expressed Genes (DEGs)

Three shoot samples for each SWC treatment were used for RNA-Seq. 1  $\mu\text{g}$  RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using the NEBNext Ultra™ RNA Library Prep Kit for IlluminaR (NE, America) following the manufacturer's recommendations, and index codes were added to attribute sequences to each sample. The library preparations were sequenced on an Illumina platform and 150 bp paired-end reads were generated.

At this stage, fastp was used to filter the original data, mainly to remove reads with adapters; these paired reads are removed when the N content of any sequencing read exceeded 10% of the number of bases in the read or when the number of low-quality ( $Q \leq 20$ ) bases exceeded 50% of the number of bases in the read. All subsequent analyses were based on clean reads.

The reference genome and its annotation files were downloaded from the designated website, and HISAT was used to construct the index and compare the clean reads to the reference genome.

DESeq2 was used to analyze the differential expression between the two groups, and the  $p$ -value was corrected using the Benjamini–Hochberg method. The corrected  $p$ -value and  $\log_2$ foldchange were used as the threshold for significant difference expression. Functional enrichment was analyzed using Gene Ontology (GO) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

### 3. Results

#### 3.1. Morphological Characteristics

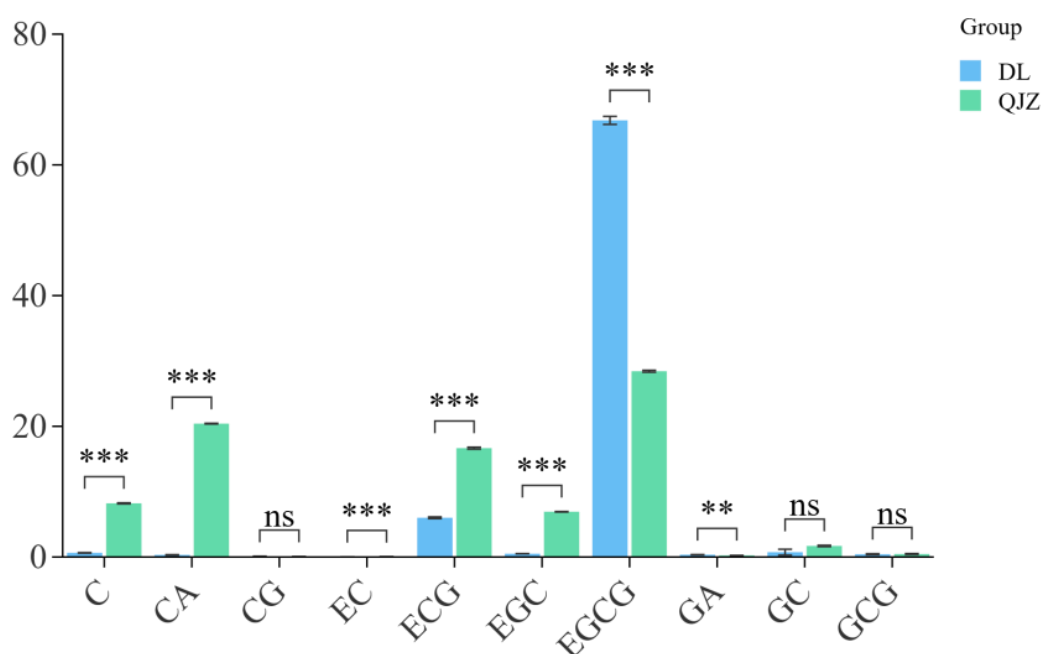
For the systematic analysis of the morphological characteristics of *C. taliensis* (DL) and *C. sinensis* (QJZ), agronomic traits such as plant shape and leaf shape were investigated. The results are as follows. The plant shape for both plants was the semi-arbor type, but DL spread horizontally and QJZ was semi-erect. The leaf size was large and medium for DL and QJZ, respectively. The leaf shape for both DL and QJZ was elliptic. The buds of DL were yellowish-green, while the buds of QJZ were green. The bud pubescence of DL was dense, while that of QJZ, on the contrary, was sparse. Both DL and QJZ had shiny buds and leaves. The tenderness of DL was stronger than that of QJZ. QJZ had a harder leaf texture compared to DL. DL exhibited a rugose leaf upper surface and wavy leaf margins, while QJZ had a slightly rugose leaf upper surface and flat leaf margins. Both plants exhibited sparse and flat leaf serration, with the difference being that the leaf serration of DL was obtuse and the sharpness of the leaf serration of QJZ was medium (Figure 1A,B).

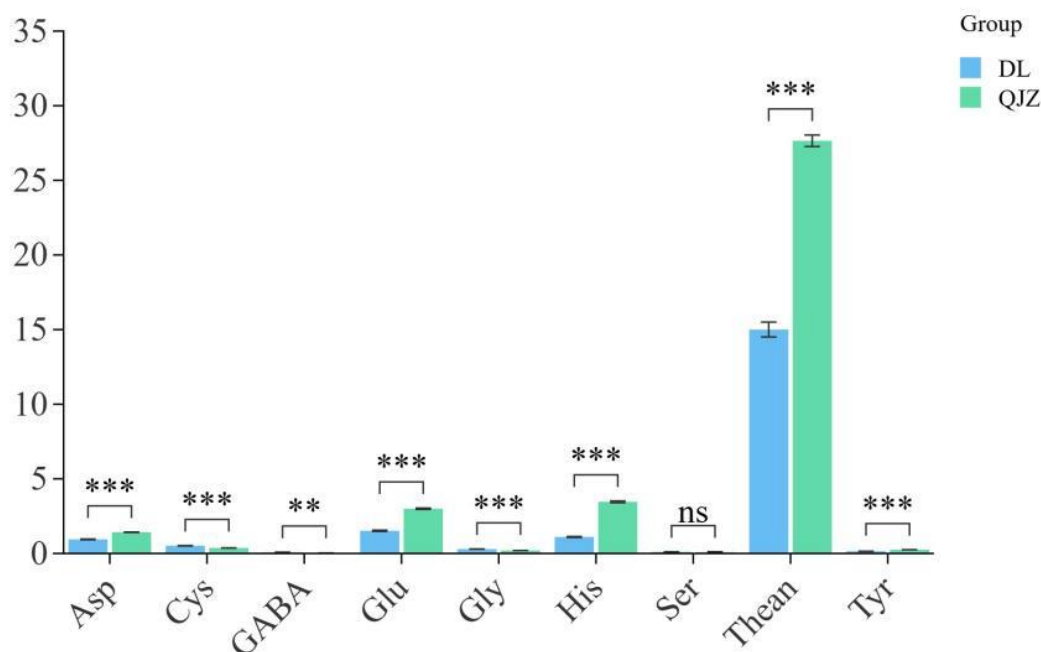


**Figure 1.** Agronomic traits of a single bud, one bud with two leaves, and one bud with three leaves from (A) *C. taliensis* (DL) and (B) *C. sinensis* (QJZ).

### 3.2. Physicochemical Indices in DL and QJZ

To identify the major physicochemical indices of taste in DL and QJZ, we determined the aqueous extraction, tea polyphenols, free amino acids, and catechins in the two tea samples. The outcomes suggested that there were no apparent variance in terms of aqueous extraction or tea polyphenols between DL and QJZ. The quantitative analysis of catechins, caffeine, and amino acids in the tea samples were performed using the HPLC method. As shown in Figure 2, there were differences in individual components between the two tea samples. Compared with QJZ, the content of C, CA, EC, ECG, EGC, GC, Rutin, Glu, His, Phe, Thean, Thr, and Tyr were lower in DL. Notably, the content of C, CA, EGC, EGCG, and GABA showed significant differences between DL and QJZ. The content of C, CA, and EGC in QJZ were 13.03-, 64.81-, and 13.59-fold higher than those in DL, respectively. However, in DL, the EGCG and GABA contents were 2.35 times and 2.66 times higher than in QJZ (Figure 2). The discrepancy in these physicochemical indices was more dramatic in DL and QJZ. Accordingly, we initially speculated that the two species of tea have certain differences in biochemical and taste quality, which paved the way for further research to investigate the possible mechanisms of tea taste formation.

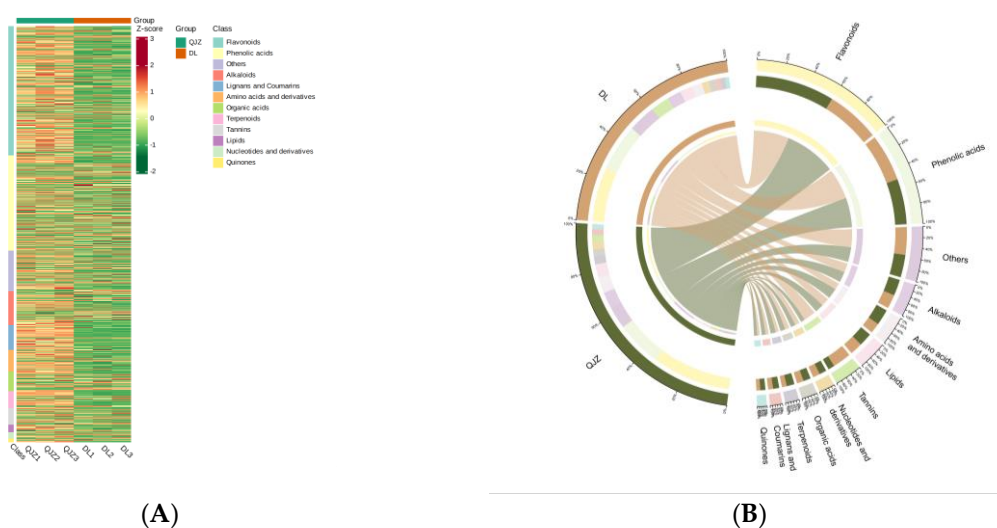




**Figure 2.** Physicochemical indices of *C. taliensis* and *C. sinensis*. \*\* indicates  $p > 0.001$ ; \*\*\* indicates  $p < 0.001$ ; “ns” indicates  $p > 0.05$ ; C: catechin; CA: caffeine; CG: catechin gallate; EC: epicatechin; ECG: epicatechin gallate; EGC: epigallocatechin; EGCG: epigallocatechin gallate; GA: gallic acid; GC: gallic acid; GCG: gallic acid; Asp: aspartic acid; Cys: cysteine; GABA:  $\gamma$ -aminobutyric acid; Glu: glutamic acid; Gly: glycine; His: histidine; Ser: serine; Thean: L-Theanine; Tyr: tyrosine.

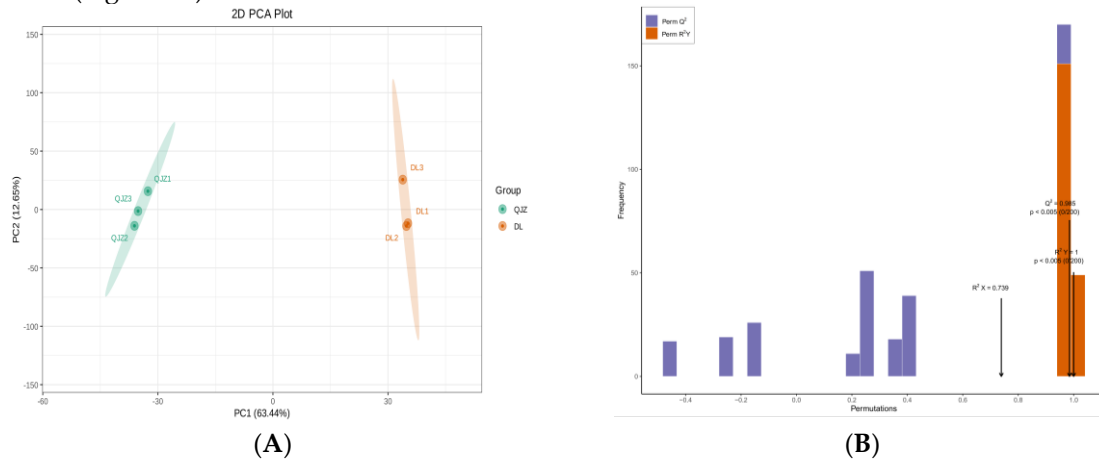
### 3.3. Metabolome Profiling

To obtain the metabolite profiling within *C. taliensis* and *C. sinensis*, metabolomics was performed. 2269 individual metabolites were positively identified in these two tea samples (Figure 3A), and grouped to 12 main categories (Figure 3B): alkaloids (8.73%), amino acids and derivatives (7.89%), flavonoids (22.52%), lignans and coumarins (5.38%), lipids (6.96%), nucleotides and derivatives (3.57%), organic acids (4.89%), others (13.18%), phenolic acids (17.85%), quinones (0.84%), tannins (3.13%) and terpenoids (5.07%). The most plentiful metabolites were flavonoids, with phenolic acids, others, alkaloids, amino acids and their derivatives being the next in order.



**Figure 3.** Metabolomic profiling of *C. taliensis* and *C. sinensis*. (A) Heat map of metabolites. Sample names are shown horizontally and metabolite information vertically. (B) Metabolite classification Circos chart.

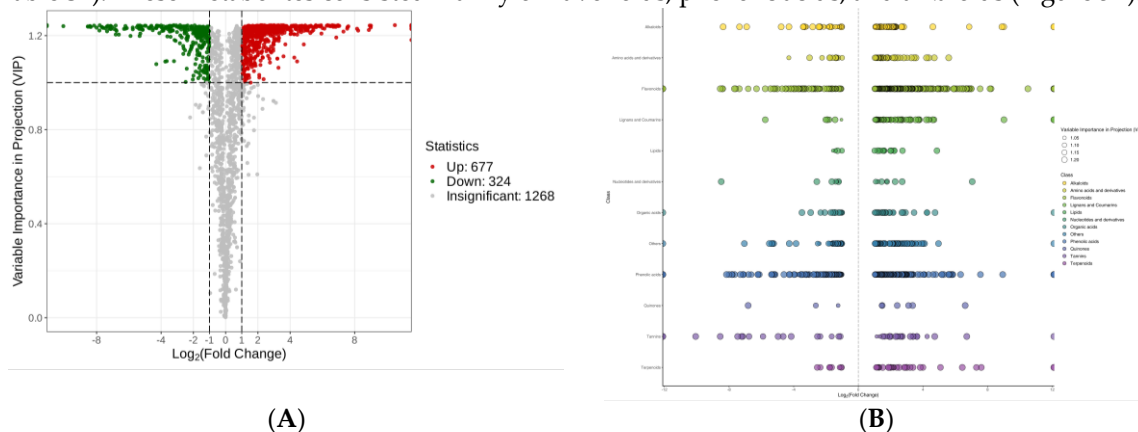
The PCA (Principal Component Analysis) illustrated that PC1 and PC2 accounted for 63.44% and 12.65% of the total variance, respectively, and distinguish between the two tea samples clearly with different colors (Figure 4A). Larger value of PC1 signified a stronger incidence of genetic variation between species. The OPLS-DA (Orthogonal Partial Least Squares-Discrimination Analysis) score plot (Figure S1) also revealed significant separateness among the judgment classes of DL and QJZ. The values for  $R^2Y$  and  $Q^2$  were 1 and 0.985, demonstrating that the metabolomic data were credible (Figure 4B).



**Figure 4.** Multivariate statistical analysis of metabolites from *C. taliensis* and *C. sinensis*. (A) PCA score plot. (B) The permutation analysis of the OPLS-DA model.

### 3.4. Differentially Accumulated Metabolites (DAMs) Analysis of Two Tea Species

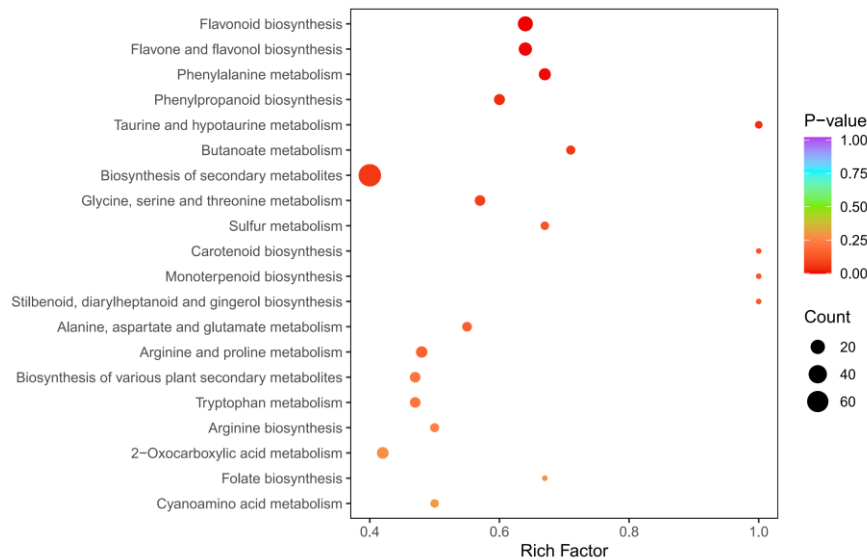
To further investigate the possible metabolic pathways involved, a volcano plot was mapped to elucidate the DAMs (Figure 5A). Under the threshold of  $VIP \geq 1$ , fold change (FC)  $\geq 2$ , or  $FC \leq 0.5$ , a total of 1001 DAMs were identified, including 677 upregulated DAMs and 324 downregulated DAMs (Table S1). These metabolites consisted mainly of flavonoids, phenolic acids, and alkaloids (Figure 5B).



**Figure 5.** Differentially accumulated metabolite (DAM) profiles of *C. taliensis* and *C. sinensis*. (A) Volcano plot. (B) Scatter plot.

The KEGG enrichment analysis was conducted to establish the primary pathways of DAMs in *C. taliensis* and *C. sinensis*. The findings pointed that the DAMs in DL and QJZ were predominantly enriched in the “biosynthesis of secondary metabolites (ko01110)”, “flavonoid biosynthesis (ko00941)”, “flavone and flavonol biosynthesis (ko00944)”, “phenylpropanoid biosynthesis (ko00940)”, “phenylalanine metabolism (ko00360)”, “glycine, serine and threonine metabolism (ko00260)”, “taurine and hypotaurine metabolism (ko00430)”, and “Butanoate metabolism (ko00650)” pathways (Figure 6), with “biosynthesis of secondary metabolites”, “flavonoid biosynthesis”, and “flavone and flavonol biosynthesis” being the three most prominent pathways for

enrichment. The results demonstrated that these pathways and their pertinent metabolites are pivotal elements impacting the biochemical components and sensory quality of tea.

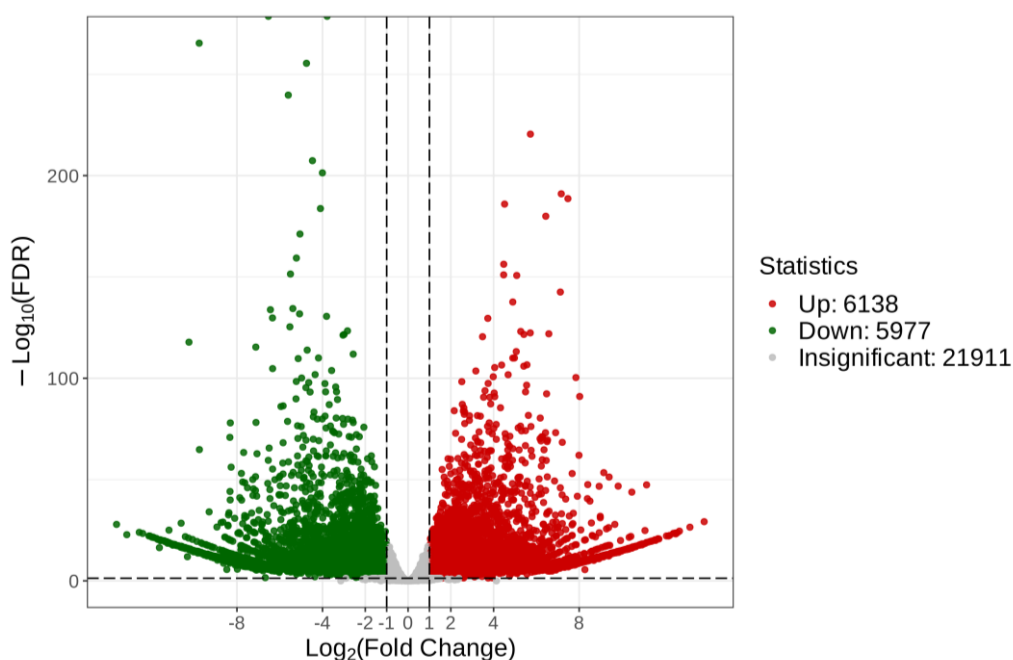


**Figure 6.** Pathway enrichment analysis of DAMs of *C. taliensis* and *C. sinensis*.

### 3.5. Comparative Transcriptome Analysis of Two Tea Species

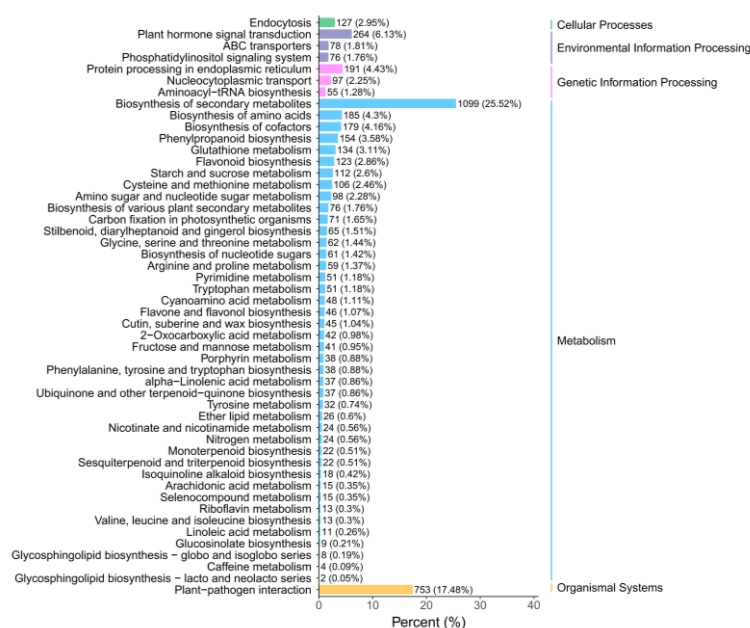
In order to characterize the variability of performance features of *C. taliensis* and *C. sinensis* and their formation mechanisms, transcriptome sequencing and analysis were performed in this study. 362,190,414 clean reads were acquired after removal of low-quality reads. The percentages of Q30 and GC were 94.4–95.09% and 45.12–45.52%, respectively, demonstrating the high standard of quality of the transcriptome sequencing data.

To reveal the unique genes that were differentially expressed between *C. taliensis* and *C. sinensis*, we performed comparative analysis of transcriptomic data. With the thresholds of FDR < 0.05 and  $\log_2 FC \geq 1$ , 34,026 DEGs were recognized (Table S2), of which 6138 and 5977 were up- and down-regulated, respectively (Figure 7).

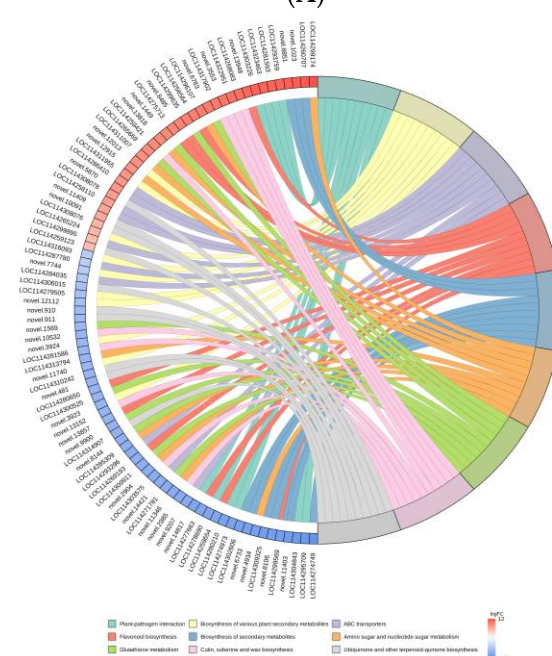


**Figure 7.** Volcano plot of DEGs of *C. taliensis* and *C. sinensis*.

In this study, we collected the top 50 most enriched pathways and plotted them as bar graphs (Figure 8A). The enrichment chord diagram was then plotted using the nine pathways with the minimum q-value (Figure 8B). KEGG pathway investigations pointed that a majority of the DEGs were enriched in the “biosynthesis of secondary metabolites (ko01110)”, “biosynthesis of various plant secondary metabolites (ko00999)”, “plant–pathogen interaction (ko04626)”, “flavonoid biosynthesis (ko00941)”, “glutathione metabolism (ko00480)” and “amino sugar and nucleotide sugar metabolism (ko00520)” pathways. In parallel, we discovered that DEGs were remarkably enriched in the “ABC transporters (ko02010)”, “protein processing in endoplasmic reticulum (ko04141)”, and “phenylpropanoid biosynthesis (ko00940)” pathways as well.



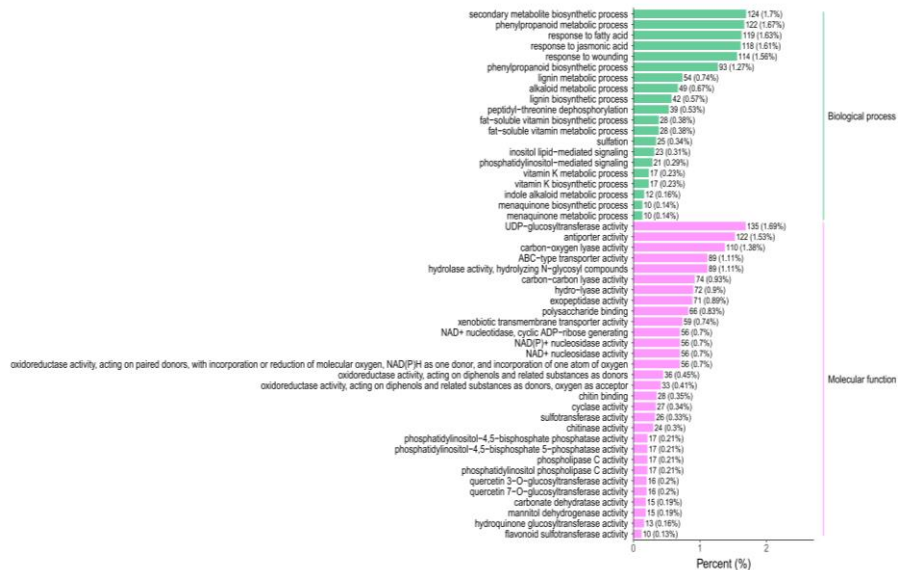
(A)



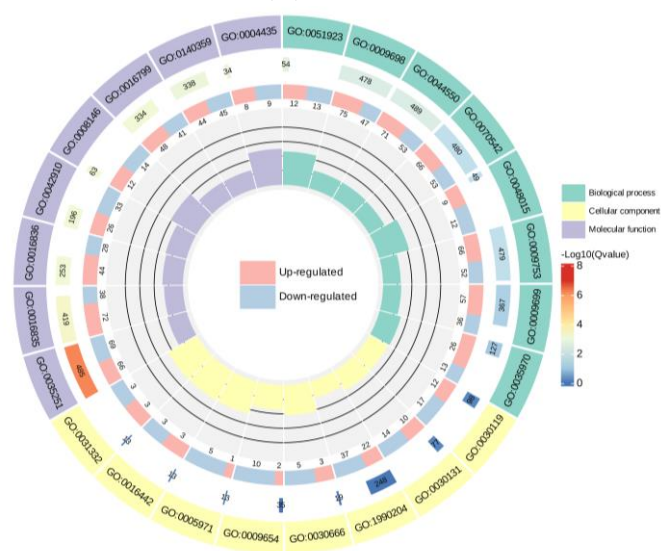
(B)

**Figure 8.** KEGG enrichment profiling of DEGs of *C. taliensis* and *C. sinensis*. (A) KEGG enrichment bar plot. (B) KEGG enrichment chord plot. On the left side of the graph are the 10 genes with the largest logFC in each category, on the right side of the plot was the 9 pathways with the most significant enrichment, and the central line represented the counterpart of the pathways and genes. The legend of the heatmap on the lower right side indicates the logFC values of the genes.

To make a further determination of the function of the DEGs, enrichment of DEGs in the GO was analyzed. In this study, the GO enrichment bar plot was plotted using the 50 GO terms with the minimum q-value (Figure 9A). Following this, the eight GO terms with minimal q-value in biological processes, cellular components, and molecular function were selected to map the GO enrichment Circos plot, respectively (Figure 9B). The results of the GO enrichment dot plot (Figure S2) demonstrated that the nine most dramatically enriched GO terms of the DEGs were “UDP-glucosyltransferase activity”, “carbon-oxygen lyase activity”, “phenylpropanoid metabolic process”, “secondary metabolite biosynthetic process”, “ABC-type transporter activity”, “antiporter activity”, “response to jasmonic acid”, “hydrolase activity, hydrolyzing N-glycosyl compounds”, and “response to fatty acid”.



(A)

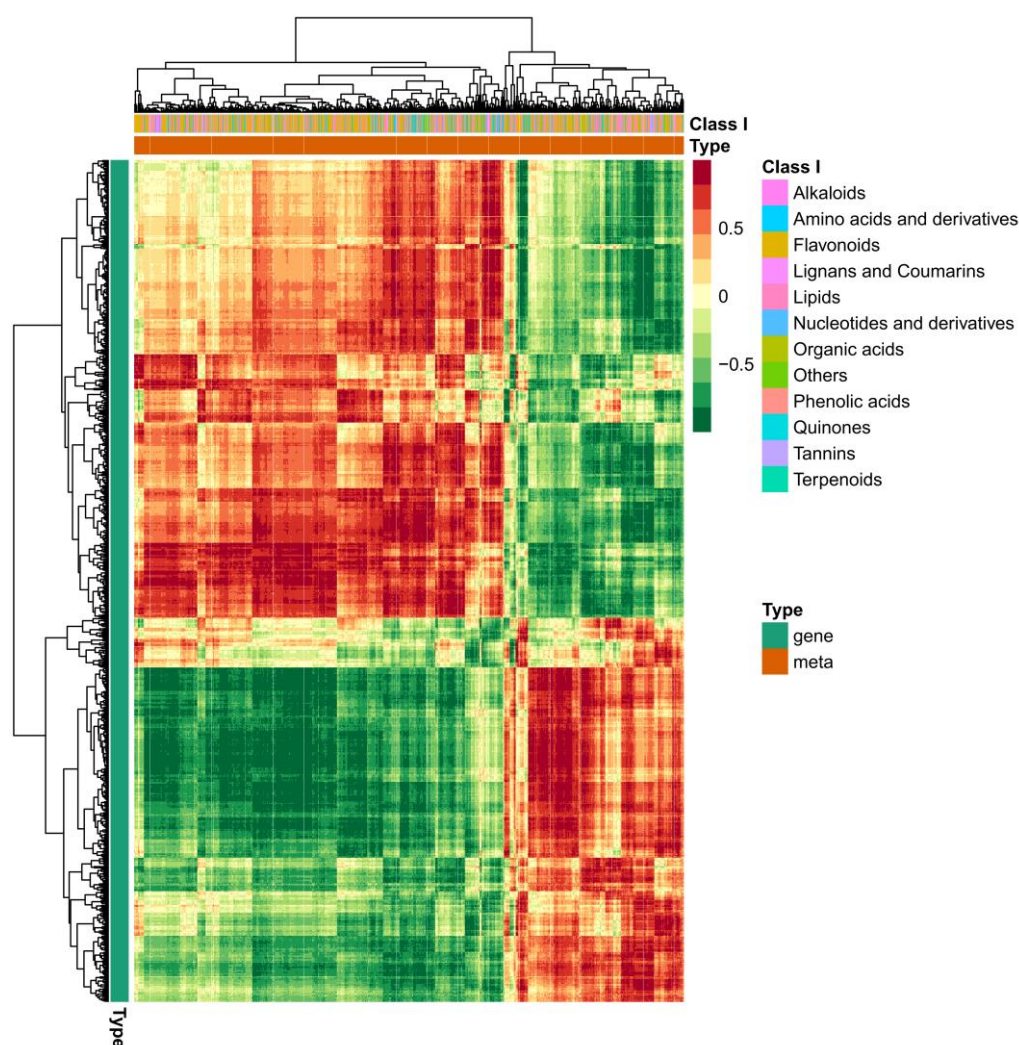


(B)

**Figure 9.** GO enrichment analysis of DEGs of *C. taliensis* and *C. sinensis*. (A) GO enrichment bar plot. Numbers in the plot represent the amount of differential genes annotated to the GO term, and the labels on the far right represent the classification to which the GO term is affiliated. (B) GO enrichment Circos plot. In the figure, from outside to inside, the first circle denotes the GO terms; the second circle indicates the amount of background genes in that classification and q-value by bar length and color; the third circle is the up- and down-regulated gene ratio bar. Specific values are shown at the bottom; with the fourth circle representing the Rich Factor value for each classification, and all cells on the supporting line in the background correspond to 0.2.

### 3.6. Conjoint Analysis of Transcriptomics and Metabolomics Data

To gain additional insight into the associations between genes and metabolites, we conducted a joint DEG and DAM analysis. Comprehensive categorization of the DEGs and DAMs was achieved by drawing a correlation heat map and categorizing the metabolites into different groups (Figure 10A). Of these, the DAMs were mainly concentrated in phenolic acids, flavonoids, alkaloids, amino acids and derivatives, and organic acids. The findings indicated significant dissimilarity between QJZ and DL. A nine-quadrant plot was drawn in accordance with the  $PCC > 0.8$  and  $p\text{-value} < 0.05$  (Figure 10B). Diverse genes/metabolites were characterized by distinct colors and quadrants as well as positive and negative regulatory interactions. Quadrant 5 indicated that neither genes nor metabolites were differentially expressed; quadrants 3 and 7 represented a consistent pattern of gene and metabolite expression, while the contrary was observed in quadrants 1 and 9. The analysis of the DAMs and DEGs in conjunction with KEGG (Figure 10C) revealed 25 co-enriched pathways in the DL and QJZ, with “biosynthesis of secondary metabolites (ko01110)”, “biosynthesis of amino acids (ko01230)”, “biosynthesis of cofactors (ko01240)”, “phenylpropanoid biosynthesis (ko00940)” and “flavonoid biosynthesis (ko00941)” being the metabolic pathways that were significantly different between the DAMs and DEGs. We also constructed correlation network diagrams by selecting DAMs and DEGs with a PCC greater than 0.80 and a p-value less than 0.05 in each pathway, further demonstrating the relationship between the DAMs in DL and QJZ and the DEGs enriched in the same KEGG pathway. In the biosynthesis of amino acids, 764 DAMs including 68 alkaloids, 309 amino acids and their derivatives, 281 organic acids, 55 others, and 51 phenolic acids were correlated strongly with DEGs (Figure 10D, Table S3).



(A)





The primary objective of metabolomics analysis is to examine and sift out biologically and statistically significant metabolites from samples and to illustrate mechanisms of metabolic processes and transformation in organisms. [45]. In the context of the present study, a metabolomics profiling of DL and QJZ was undertaken. The experimental outcomes demonstrated that among the 2269 metabolites detected, flavonoids, phenolic acids, others, alkaloids, and amino acids and their derivatives accounted for the majority of the metabolites, with a total proportion of more than 70% of the aggregate of metabolites, indicating that they are the principal constituents within the metabolites of tea and make a remarkable contribution to the quality of the tea leaves. The biosynthesis of secondary metabolites, flavonoid biosynthesis, and flavone and flavonol biosynthesis were the three pathways most notably enriched in the KEGG analysis.

Transcriptome broadly refers to the collection of all transcriptome products in a cell under a particular physiological condition, and the object of study refers to the combination of a specific cell's all RNAs that can be transcribed under a particular functional state [46]. The results suggested that most of the enriched terms belonged to molecular functions. DEGs are mainly enriched in the access pathways of “secondary metabolite biosynthesis”, “plant–pathogen interactions” and “flavonoid biosynthesis”.

Biological processes are complex and holistic, and it is difficult to analyze the macro-developmental processes of biological systems from single-omics data [47]. Integrating multi-omics data for analysis can reduce the false positives associated with single-omics analysis and provide a panoramic window into biological activity [48]. The metabolomics and transcriptomics KEGG co-enrichment analysis showed a significant enrichment of DAMs and DEGs in the biosynthesis of secondary metabolites, biosynthesis of amino acids, biosynthesis of amino acids, and biosynthesis of cofactors, which indicated that the activities of these pathways differed dramatically between the DL and the QJZ.

Tea quality is the consequence of the combined action of sub-metabolites including alkaloids, tea polyphenols and amino acids [49]. Differences in tea species lead to variances in the variety and content of inclusions. Species diversity leads to differences in endogenous metabolites in the fresh leaves, which can have a dramatic effect on the flavor of the tea products, even under the same processing conditions. Precisely for this reason, the study of different tea species is the basis for a proper understanding of the suitability of the tea plant.

## 5. Conclusions

In this study, *C. taliensis* and *C. sinensis* were used as research materials to characterize and compare their phenotypic traits, biochemical components, metabolomes, and transcriptomes, revealing significant differences between them. By combining data from metabolomic and transcriptomic analysis, we found that the metabolites responsible for the differences are mainly flavonoids, phenolic acids, and alkaloids, which have an impact on the quality of tea products. These results help us to gain a deeper comprehension of the biosynthetic pathways of quality-related components in *C. taliensis* and *C. sinensis*, and provide partial data support for further studies.

**Supplementary Materials:** The following supporting information can be downloaded at: [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1), Figure S1: OPLS-DA score plot; Figure S2: GO enrichment dot plot; Table S1: The DAMs between DL and QJZ; Table S2: The DEGs between DL and QJZ; Table S3: Cornetwork file of ko01230.

**Author Contributions:** Conceptualization, Z.L. and J.L.; methodology, R.L.; software, R.L. and X.L.; validation, M.N., H.Y. and X.Z.; formal analysis, M.N., J.D. and Y.H.; investigation, M.N., H.Y., Q.Q.; data curation, M.N.; writing—original draft preparation, M.N.; writing—review and editing, Z.L.; visualization, J.L.; supervision, Y.X.; project administration, M.N.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China, grant number 32160726.

**Institutional Review Board Statement:** No applicable.

**Informed Consent Statement:** No applicable.

**Data Availability Statement:** The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Li, J.-W.; Li, H.; Liu, Z.-W.; Wang, Y.-X.; Chen, Y.; Yang, N.; Hu, Z.-H.; Li, T.; Zhuang, J. Molecular markers in tea plant (*Camellia sinensis*): Applications to evolution, genetic identification, and molecular breeding. *Plant Physiol. Biochem.* **2023**, *198*, 107704.
2. Hashimoto, M.; Takasi, S. Morphological studies on the origin of the tea plant V, a proposal of one place of origin by cluster analysis. *Jpn J. Trop. Agric.* **1978**, *21*, 93–101.
3. Shen, J.; Wang, S.; Sun, L.; Wang, Y.; Fan, K.; Li, C.; Wang, H.; Bi, C.; Zhang, F.; Ding, Z. Dynamic changes in metabolic and lipidomic profiles of tea plants during drought stress and re-watering. *Front. Plant Sci.* **2022**, *13*, 978531. <https://doi.org/10.3389/fpls.2022.978531>
4. Chen, L.; Yu, F.L.; Tong, Q.Q. Discussions on phylogenetic classification and evolution of section *Thea*. *J. Tea Sci.* **2000**, *20*, 89–94.
5. Chang, H.T. *Thea*—A section of beverage tea trees of the genus *Camellia*. *Acta Sci. Nat. Univ. Sunyats.* **1981**, *1*, 87–99.
6. Yu, F.L. Discussion on the originating place and the originating center of tea plants. *J. Tea Sci.* **1986**, *6*, 1–8.
7. Wambulwa, M.C.; Meegahakumbura, M.K.; Kamunya, S.; Wachira, F.N. From the Wild to the Cup: Tracking Footprints of the Tea Species in Time and Space. *Front. Nutr.* **2021**, *8*, 706770. <https://doi.org/10.3389/fnut.2021.706770>
8. Liu, S.; An, Y.; Li, F.; Li, S.; Liu, L.; Zhou, Q.; Zhao, S.; Wei, C. Genome-wide identification of simple sequence repeats and development of polymorphic SSR markers for genetic studies in tea plant (*Camellia sinensis*). *Mol. Breed.* **2018**, *38*, 1–13.
9. Sueoka, N.; Sukanuma, M.; Sueoka, E.; Okabe, S.; Matsuyama, S.; Imai, K.; Nakachi, K.; Fujiki, H. A new function of green tea: Prevention of lifestyle-related diseases. *Ann. N. Y. Acad. Sci.* **2001**, *928*, 274–280.
10. Zaveri, N.T. Green tea and its polyphenolic catechins: Medicinal uses in cancer and noncancer applications. *Life Sci.* **2006**, *78*, 2073–2080.
11. Wang, F.; Cheng, X.; Cheng, S.; Li, W.; Huang, X. Genetic diversity of the wild ancient tea tree (*Camellia taliensis*) populations at different altitudes in Qianjiazhai. *PLoS ONE* **2023**, *18*, e0283189.
12. Ming, T.L. A revision of *Camellia* Section *Thea*. *Acta Bot. Yunnanica* **1992**, *14*, 116–132.
13. Zhu, L.F.; Xu, M.; Zhu, H.-T.; Wang, D.; Yang, S.-X.; Yang, C.-R.; Zhang, Y.-J. New Flavan-3-ol Dimer from Green Tea Produced from *Camellia taliensis* in the Ai-Lao Mountains of Southwest China. *J. Agric. Food Chem.* **2012**, *60*, 12170–12176.
14. Zhao, D.W.; Yang, J.-B.; Yang, S.-X.; Kato, K.; Luo, J.-P. Genetic diversity and domestication origin of tea plant *Camellia taliensis* (Theaceae) as revealed by microsatellite markers. *BMC Plant Biol.* **2014**, *14*, 14.
15. Huang, R.; Wang, Z.; Wen, W.; Yao, M.; Liu, H.; Li, F.; Zhang, S.; Ni, D.; Chen, L. Comprehensive dissection of variation and accumulation of free amino acids in tea accessions. *Hortic. Res.* **2023**, *11*, uhad263. <https://doi.org/10.1093/hr/uhad263>
16. Utpott, M.; Rodrigues, E.; de Oliveira Rios, A.; Mercali, G.D.; Flôres, S.H. Metabolomics: An analytical technique for food processing evaluation. *Food Chem.* **2022**, *366*, 130685. <https://doi.org/10.1016/j.foodchem.2021.130685>
17. Ji, H.G.; Lee, Y.R.; Lee, M.S.; Hwang, K.H.; Park, C.Y.; Kim, E.H.; Hong, Y.S. Diverse metabolite variations in tea (*Camellia sinensis* L.) leaves grown under various shade conditions revisited: A metabolomics study. *J. Agric. Food Chem.* **2018**, *66*, 1889–1897. <https://doi.org/10.1021/acs.jafc.7b04768>
18. Xu, Q.; He, Y.; Yan, X.; Zhao, S.; Zhu, J.; Wei, C. Unraveling a crosstalk regulatory network of temporal aroma accumulation in tea plant (*Camellia sinensis*) leaves by integration of metabolomics and transcriptomics. *Environ. Exp. Bot.* **2018**, *149*, 81–94. <https://doi.org/10.1016/j.envexpbot.2018.02.005>
19. Farag, M.A.; Elmetwally, F.; Elghanam, R.; Kamal, N.; Hellal, K.; Hamezah, H.S.; Zhao, C.; Mediani, A. Metabolomics in tea products; a compile of applications for enhancing agricultural traits and quality control analysis of *Camellia sinensis*. *Food Chem.* **2023**, *404* (Pt. B), 134628. <https://doi.org/10.1016/j.foodchem.2022.134628>

20. Cai, H.; Zhong, Z.; Li, Z.; Zhang, X.; Fu, H.; Yang, B.; Zhang, L. Metabolomics in quality formation and characterisation of tea products: A review. *Int. J. Food Sci. Technol.* **2022**, *57*, 4001–4014. <https://doi.org/10.1111/ijfs.15767>
21. Li, X.; Liu, Y.; Hu, W.; Yin, B.; Liang, B.; Li, Z.; Zhang, X.; Xu, J.; Zhou, S. Integrative physiological, metabolomic, and transcriptomic analysis reveals the drought responses of two apple rootstock cultivars. *BMC Plant Biol.* **2024**, *24*, 219. <https://doi.org/10.1186/s12870-024-04902-2>
22. Liu, H.; Duan, L.; Ma, J.; Jin, J.; Huang, R.; Liu, Y.; Chen, S.; Xu, X.; Chen, J.; Yao, M.; et al. CsEXL3 regulate mechanical harvest-related droopy leaves under the transcriptional activation of CsBES1.2 in tea plant. *Hortic. Res.* **2024**, *11*, uhae074. 7 Mar. . <https://doi.org/10.1093/hr/uhae074>
23. Chen, L. Descriptors and Data Standard for Tea (*Camellia spp.*). **2005**.
24. NY/T 2943-2006; Descriptors for Germplasm Resources. Ministry of Agriculture of the People's Republic of China: Beijing, China, 2006.
25. GB/T 8303-2013; Tea — Preparation of Ground Sample and Determination of Dry Matter Content. Standardization Administration of the People's Republic of China: Beijing, China, 2013.
26. Shen, J.; Wang, H.; Sun, L.; Fan, K.; Zhang, X.; Huang, Q.; Ding, S.; Wang, Y.; Ding, Z. Metabolic Variations among Three New Tea Varieties Cultivated in Shandong, China. *Foods* **2023**, *12*, 1299. <https://doi.org/10.3390/foods12061299>
27. Hua, J.; Xu, Q.; Yuan, H.; Wang, J.; Wu, Z.; Li, X.; Jiang, Y. Effects of novel fermentation method on the biochemical components change and quality formation of Congou black tea. *J. Food Compos. Anal.* **2021**, *96*, 103751.
28. GB/T 8314-2013; Tea—Determination of Free Amino Acids Content. Standardization Administration of the People's Republic of China: Beijing, China, 2013.
29. Nian, B.; Chen, L.; Yi, C.; Shi, X.; Jiang, B.; Jiao, W.; Liu, Q.; Lv, C.; Ma, Y.; Zhao, M. A high performance liquid chromatography method for simultaneous detection of 20 bioactive components in tea extracts. *Electrophoresis* **2019**, *40*, 2837–2844.
30. Zhao, M.; Ma, Y.; Dai, L.-L.; Zhang, D.-L.; Li, J.-H.; Yuan, W.-X.; Li, Y.-L.; Zhou, H.-J. A High-Performance Liquid Chromatographic Method for Simultaneous Determination of 21 Free Amino Acids in Tea. *Food Anal. Methods* **2013**, *6*, 69–75.
31. Fang, J. No tea before the Warring States period. *Agric. Hist. China* **1998**, *17*, 6–14, 39.
32. Jackson, J.R. Tea. *Nature* **1870**, *2*, 215–217.
33. Jankun, J.; Selman, S.H.; Swiercz, R.; Skrzypczak-Jankun, E. Why drinking green tea could prevent cancer. *Nature* **1997**, *387*, 561.
34. Paul, S.; Wachira, F.N.; Powell, W.; Waugh, R. Diversity and genetic differentiation among populations of Indian and Kenyan tea [*Camellia sinensis* (L.) O. Kuntze] revealed by AFLP markers. *Theor. Appl. Genet.* **1997**, *94*, 255–263.
35. Yan, Z.M.; Zhong, Y.Z.; Duan, Y.H.; Chen, Q.H.; Li, F.N. Antioxidant mechanism of tea polyphenols and its impact on health benefits. *Anim. Nutr.* **2020**, *6*, 115–123.
36. Fu, Q.Y.; Li, Q.S.; Lin, X.M.; Qiao, R.Y.; Yang, R.; Li, X.M.; Dong, Z.B.; Xiang, L.P.; Zheng, X.Q.; Lu, J.L.; et al. Antidiabetic effects of tea. *Molecules* **2017**, *22*, 849.
37. Li, Q.S.; Wang, Y.Q.; Liang, Y.R.; Lu, J.L. The anti-allergic potential of tea: A review of its components, mechanisms and risks. *Food Funct.* **2021**, *12*, 57–69.
38. Xia, E.-H.; Zhang, H.-B.; Sheng, J.; Li, K.; Zhang, Q.-J.; Kim, C.; Zhang, Y.; Liu, Y.; Zhu, T.; Li, W.; et al. The Tea Tree Genome Provides Insights into Tea Flavor and Independent Evolution of Caffeine Biosynthesis. *Mol. Plant* **2017**, *10*, 866–877.
39. Nehlig, A. Are we dependent upon coffee and caffeine? A review on human and animal data. *Neurosci. Biobehav. Rev.* **1999**, *23*, 563–576.
40. Mi, X.; Yang, C.; Qiao, D.; Tang, M.; Guo, Y.; Liang, S.; Li, Y.; Chen, Z.; Chen, J. De novo full length transcriptome analysis of a naturally caffeine-free tea plant reveals specificity in secondary metabolic regulation. *Sci. Rep.* **2023**, *13*, 6015.
41. Xie, R.; Zhang, H.; Lv, X.; Lin, Q.; Chen, B.-H.; Lai, Y.-W.; Chen, L.; Teng, H.; Cao, H. The evaluation of catechins reducing heterocyclic aromatic amine formation: Structure-activity relationship and mechanism speculation. *Curr. Res. Food Sci.* **2024**, *8*, 100727. <https://doi.org/10.1016/j.crfs.2024.100727>
42. Wu, C.; Xu, H.; Héritier, J.; Andlauer, W. Determination of catechins and flavonol glycosides in Chinese tea varieties. *Food Chem.* **2012**, *132*, 144–149.

43. Jin, J.Q.; Ma, J.-Q.; Ma, C.-L.; Yao, M.-Z.; Chen, L. Determination of Catechin Content in Representative Chinese Tea Germplasms. *J. Agric. Food Chem.* **2014**, *62*, 9436–9441.
44. Yang, G.; Meng, Q.; Shi, J.; Zhou, M.; Zhu, Y.; You, Q.; Xu, P.; Wu, W.; Lin, Z.; Lv, H. Special tea products featuring functional components: Health benefits and processing strategies. *Compr. Rev. Food Sci. Food Saf.* **2023**, *22*, 1686-1721. <https://doi.org/10.1111/1541-4337.13127>
45. Ain, A.; Sarsaiya, S.; Singh, R.; Gong, Q.; Wu, Q.; Shi, J. Omics approaches in understanding the benefits of plant-microbe interactions. *Front. Microbiol.* **2024**, *15*, 1391059. <https://doi.org/10.3389/fmicb.2024.1391059>
46. Zhang, Q.; Folk, R.A.; Mo, Z.-Q.; Ye, H.; Zhang, Z.-Y.; Peng, H.; Zhao, J.-L.; Yang, S.-X.; Yu, X.-Q. Phylotranscriptomic analysis reveal deep gene tree discordance in *Camellia* (Theaceae). *Mol. Phylogenet. Evol.* **2023**, *188*, 107912. <https://doi.org/10.1016/j.ympev.2023.107912>
47. Su, G.; Yu, C.; Liang, S.; Wang, W.; Wang, H. Multi-omics in food safety and authenticity in terms of food components. *Food Chem.* **2024**, *437* (Pt. 2), 137943. <https://doi.org/10.1016/j.foodchem.2023.137943>
48. Dai, X.; Li, S. Advances and Trends in Omics Technology Development. *Front. Med.* **2022**, *9*, 911861. <https://doi.org/10.3389/fmed.2022.911861>
49. Zhang, J.; Xin, W.; Zou, Y.; Yan, J.; Tang, W.; Ji, Y.; Li, W. Dynamic changes and correlation analysis of microorganisms and flavonoids/ amino acids during white tea storage. *Food Chem.* **2024**, *455*, 139932. <https://doi.org/10.1016/j.foodchem.2024.139932>

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.