Supporting Information

Succinct Title: Metabolic and Transcriptome Analysis Reveals Metabolite Variation in Fresh Shoots of Tea (*Camellia sinensis* 'Lingtou Dancong') Under Nitrogen-Deficient Conditions

Running Title: Tea Shoot Response to Nitrogen Deficiency

Zihao Qiu [1], Ansheng Li [1], Xinyuan Lin [1], Jiyuan Yao [1], Yimeng Yang [1], Renjian Liu [1], Binmei Sun [1,2], Shaoqun Liu [1,2]*, Peng Zheng [1,2]*

[1] College of Horticulture, South China Agricultural University, Guangzhou 510642, China

[2]Guangdong Provincial Engineering Technology Research Center for Southern Specialty Tea

** Corresponding authors, E-mail: scauok@scau.edu.cn; zhengp@scau.edu.cn*

Supplementary method 1. Supplementary method to this study.

Determination of Chlorophyll and Carotenoid.

Chlorophyll content was determined using the method outlined by Lichtenthaler [14]. For extraction, 0.1 g of fresh shoot tissue was ground into a fine powder in liquid nitrogen. The powdered samples were then incubated in 10 mL of a mixed extraction solution (acetone/ethanol/water = 4.5:4.5:1) in the dark for 24 hours, following the protocol of [15].. After extraction, the absorbance of the solution was measured at wavelengths of A645, A663, and A470 using a UV-visible spectrophotometer (U-5100, Hitachi, Japan). Each sample was analyzed in triplicate for both biological and technical replicates.

Measurement of Carbon, Hydrogen, Nitrogen, and Sulfur (CHNS) Element Content.

30 mg of dried samples, passed through a 100-mesh sieve, were used to determine nitrogen (N) content using the Elementar Rapid N Exceed analyzer [9], with a protein conversion factor of 6.25 [16]. Additionally, 5 mg of dried samples, also passed through a 100-mesh sieve, were analyzed using the Elementar Vario EL Cube to verify nitrogen content and measure carbon (C), hydrogen (H), and sulfur (S) contents [17]. Four biological replicates were performed for each sample group.

Measurement of Water Extract, Total Tea Polyphenols (TP), and Total Free Amino Acids (FAA) Content.

The determination of water extract content was conducted according to national standard GB/T 8305-2013 for water extract (WE). Total tea polyphenols were measured following national standard GB/T 8313-2018. The total free amino acids were determined

using the method outlined in GB/T 8314-2013. All experiments included four biological replicates.

Measurement of Theanine, Caffeine, and Catechin Monomer Content.

Following the method of Chen et al. [18], caffeine, theanine, and catechins were identified using HPLC-UV/Vis (Waters Alliance 2695, 2489 UV/Vis; Waters Technologies, Milford, MA, USA) by comparing the retention times of the samples with those of the standards. A calibration curve was used for quantitative analysis (Table S1). The determination of caffeine, theanine, and catechins was performed in triplicate with three biological replicates.

Measurement of malondialdehyde (MDA) and Soluble Sugar Content.

MDA content was determined using the MDA content detection kit (BC0020, Beijing Solarbio Technology Co., Ltd.), with four biological replicates per group. Soluble sugar content was measured using the plant soluble sugar detection kit (BC0030, Beijing Solarbio Technology Co., Ltd.), with three biological replicates per group (Table S1).

Measurement of Volatile Compounds.

Volatile compounds were detected following the method described by Chen et al [19]. each sample group was analyzed in triplicate with three biological replicates. 0.2 g of dried sample and 8.64 μg of decanoic acid ethyl ester were added to a 20 mL headspace vial, which was then sealed with aluminum foil. An SPME fiber was inserted into the vial, and the mixture was extracted for 40 minutes in a metal bath at 80°C. After extraction, analysis was performed using a GC-MS system (Agilent 7890B-5977A, Agilent, Santa Clara, CA, USA) with a temperature hold of 3 minutes at 250°C. The carrier gas was high-purity helium (≥ 99.99%), with a flow rate of 1.0 mL/min. The chromatographic column had a solvent delay of 4 minutes. The temperature program was as follows: initial temperature of 50°C held for 1 minute, then ramped at 5°C/min to 220°C, held for 34 minutes, and finally maintained at 220°C for 5 minutes. The electron impact (EI) ionization source was set to 70 eV, and the mass spectrum was scanned from 30 to 400 amu. The GC column used was HP-5ms (30 m × 0.25 mm × 0.25 μm). Volatile compounds were identified by comparing their retention index (RI) with data from the National Institute of Standards and Technology (NIST) MS database (https://webbook.nist.gov/). A compound was considered matched when the calculated RI was below 15 or the mass spectrum match factor exceeded 90. Quantification of volatile compounds was performed using the internal standard method [5].

Calculation of Odor Activity Values (OAVs).

Odor Activity Values (OAVs) represent the ratio of a flavor compound's concentration to its threshold, indicating its contribution to the overall flavor. The OAV of each volatile compound was calculated as OAVi = Ci/Ti, where Ci is the concentration of compound i (μg/kg) and Ti is the threshold value of compound i (μg/kg). The relative concentrations of the released volatiles were calculated using the following formula [20], Internal standard concentration (CIS):8.64 ug/mL.

$$\text{Relative concentration}(\mu g/kg)=\frac{(\text{Peak area of target/Peak area of IS})\times C_{IS}\times 10uL}{\text{Amount of sample (2.0 g)}}$$

Measurement of Non-volatile Metabolites.

The widely targeted detection of non-volatile metabolites was assisted by Wuhan MetWare Biotechnology Co., Ltd. The detection protocol followed Wang et al [21]. In brief, 1200 μL of 70% methanol was used to dissolve 50 mg of sample powder, followed by extraction for 3 hours with vortexing for 30 seconds every 30 minutes. The supernatant was then collected by centrifugation and filtered through a 0.22 μm membrane into the injection vial. Metabolite analysis was performed using an ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) system. The chromatographic column was an Agilent SB-C18 (1.8 μm, 2.1 mm × 100 mm) set at 40°C, with a 2 μL injection volume. Gradient elution was applied at a flow rate of 0.35 mL/min. From 0 to 9 minutes, the B phase increased linearly from 5% to 95%; from 9 to 10 minutes, the B phase was maintained at 95%; and from 10 to 14 minutes, the B phase was reduced back to 5%. Mobile phase A consisted of 0.1% formic acid, and mobile phase B was acetonitrile containing 0.1% formic acid. The mass spectrometry analysis employed an electrospray ionization (ESI) source in both positive and negative ion modes, with spray voltages set to 5500 V and -4500 V, respectively. The ion source temperature was maintained at 550°C. The gas parameters were set as follows: gas source I (GSI) at 50 psi, gas source II (GSII) at 60 psi, and curtain gas at 25 psi. Quantitative analysis was performed using multiple reaction monitoring (MRM) mode, with collision-induced dissociation (CID) set to high intensity and collision gas (nitrogen) at moderate strength. The declustering potential (DP) and collision energy (CE) were optimized to accurately monitor the target metabolites. The metabolomics analysis was performed with three biological replicates for each group. In this study, we employed a 'widely targeted' metabolomic approach for the relative quantification of metabolites. This method focuses primarily on identifying differential abundance metabolites (DAMs) through relative changes rather than providing absolute quantification values. However, it is important to note that the quantification provided is relative, and therefore, does not reflect precise concentrations of individual metabolites.

Differentially Accumulated Metabolites (DAMs) Screening.

The samples were qualitatively analyzed using mass spectrometry in conjunction with the MVDB database developed by Wuhan MetWare Biotechnology Co., Ltd. Metabolite identification was performed using mass spectrometry based on the MetWare database (MWDB). Identification was validated through multi-parameter matching, including accurate precursor ion (Q1), product ion (Q3), retention time (RT), and isotope distribution. MS/MS spectra of sample metabolites were compared with reference spectra in the database using an intelligent spectral matching algorithm, with mass tolerances set at 2 ppm for MS and 5 ppm for MS/MS, and an RT tolerance of 0.2 min. All identifications were classified and confirmed according to a three-tier confidence level system. Furthermore, a hybrid MS strategy combining high-resolution and triple quadrupole mass spectrometry

was applied to enhance identification reliability. The raw data were log2-transformed and mean-centered. For comparison between two groups, Differentially Accumulated Metabolites (DAMs) were selected based on VIP > 1 and |Log2FC| ≥ 1.0. For multi-group comparisons, DAMs were selected based on VIP > 1 and $P < 0.05$. To avoid model overfitting, all OPLS-DA analyses underwent 200 permutation tests.

Transcriptome Sequencing and Differentially Expressed Genes (DEGs) Screening.

Transcriptome sequencing and differentially expressed genes (DEGs) screening were performed according to the procedure described by Lin et al. [22]. Total RNA from the samples was extracted using the Trizol reagent kit (TransGenBiotech) and quality control and sequencing were performed at Wuhan MetWare Biotechnology Co., Ltd. Transcriptome library construction, sequencing, and analysis were carried out using the Illumina HiSeq X Ten platform (Illumina Inc, CA, USA). mRNA was first enriched using Oligo(dT) magnetic beads and rRNA was removed. The RNA samples were then randomly fragmented, followed by reverse transcription using random primers. Additionally, the 3′ and 5′ ends were repaired, an A tail was added, and sequencing adapters were ligated. DNA fragments were amplified and sequenced using paired-end sequencing on the Illumina platform. The raw sequences underwent quality control to remove low-quality sequences and adapter contamination, ensuring high-quality data. Base calling was performed using CASAVA software, converting the sequences into raw data. Adapter-contaminated reads, low-quality reads, and reads with more than 5% N content were removed using fastp software. Sequence alignment was performed with Tophat software, using Bowtie2 as the alignment tool, and visualized with Integrative Genomics Viewer. Finally, HTseq-count software was used to calculate the read count for each gene, representing gene expression levels. Filtered reads were aligned to the (*Camellia sinensis* cv. Shuchazao2) reference genome using HISAT v2.1.0. Gene expression levels were quantified as FPKM (Fragments Per Kilobase of exon per Million mapped fragments) with featureCounts v1.6.4 and StringTie v1.3.4. Differentially expressed mRNAs were analyzed using the DESeq2 package. Differentially expressed genes (DEGs) were selected based on a false discovery rate (FDR) < 0.05 and |log2 fold change| ≥ 1. Enrichment of metabolic pathways and gene functions was annotated using the KEGG database. The transcriptome analysis was performed with three biological replicates for each group.