Comparative transcriptome analysis to reveal putative genes responsible for high theacrine content in Kucha (Camellia kucha (Chang et Wang) Chang)

Hongjian Li  
Tea research institute, Guangdong Academy of Agricultural Sciences  
https://orcid.org/0000-0002-3372-6655

Kaixing Fang  
Tea Research Institute, Guangdong Academy of Agricultural Sciences

Dandan Qin  
Tea Research Institute, Guangdong Academy of Agricultural Sciences

Xiaohui Jiang  
Tea Research Institute, Guangdong Academy of Agricultural Sciences

Qiushuang Wang  
Tea Research Institute, Guangdong Academy of Agricultural Sciences

Chendong Pan  
Tea Research Institute, Guangdong Academy of Agricultural Sciences

Bo Li  
Tea Research Institute, Guangdong Academy of Agricultural Sciences

Qing Wang  
Tea Research Institute, Guangdong Academy of Agricultural Sciences

Hualing Wu (✉ wuhualing@163.com)

Research article

Keywords: Tea plant, Theacrine, Transcriptome, Biosynthesis

DOI: https://doi.org/10.21203/rs.3.rs-56307/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** The purine alkaloid theacrine, synthesized in a high amount in Kucha (*Camellia kucha* (Chang et Wang) Chang), has shown multiple pharmacological effects such as anti-depression, dehydration and hypnotic activities. It is critical to better understand the theacrine metabolism at the molecular level for future breeding programs.

**Results:** In this study, we carried out comparative transcriptome analysis of Kucha (Kucha-6 and Kucha-11) and conventional varieties (Yinghong 9 and Qingxin 1). HPLC showed that Kucha synthesized more than 27-fold theacrine in comparison to conventional varieties. A total of ~ 671.61 million high quality clean reads were yield, and 71.42 ~ 76.87% of these were aligned to the reference genome for each library. Among the 19,095 differentially expressed genes (DEGs) identified by a pairwise comparison approach, a pool of 1,142 common DEGs between Kucha and conventional varieties were screened. Functional annotation allowed the identified of 20 oxidoreductase (OR) genes and seven methyltransferase (MT) genes with enhanced expression in Kucha, which might be associated with theacrine diversity in the examined materials. Particularly, two N-methyltransferase-encoding genes (TEA010054 and TEA022559) may catalyze the final methylation step during theacrine synthesis in Kucha.

**Conclusion:** The current study focused on screening genes of theacrine formation in tea plant, which contributes to the elucidation of candidate genes and breeding tea plant with high theacrine content.

**Background**

Purine alkaloids (PAs), occurred in about 100 species in the plant kingdom, are secondary metabolites formed from purine nucleotides [1]. As one of the most important non-alcoholic beverages, tea plant synthesized diverse high-value pharmaceutically important PAs including fatigue resistance caffeine, antimicrobial agent theobromine and potential asthma drug theophylline. The theacrine, only accumulated in high amounts in Kucha, also displayed several beneficial effects in improving learning and memory, anti-depressive and hypnotic activities [2-4].

It is generally accepted that PAs biosynthesis begins with substrates from four different pathways: 1) inosine-5'-monophosphate from de novo purine biosynthetic pathway (de novo route), 2) adenosine 5'-monophosphate from de novo and salvage pathways (AMP route), 3) adenosine from S-adenosyl-L-methionine cycle (SAM route) and 4) guanine nucleotides from guanosine-5'-monophosphate (GMP route), that yield xanthosine [5]. Subsequent orderly conversion of xanthosine with three methylation and one nucleosidase reactions produced caffeine as follows: xanthosine → 7-methylxanthosine → 7-methylxansine → theobromine → caffeine [6]. Then caffeine appears to convert to theacrine by oxidation at the C8 and methylation at the N9 positions in Kucha [7].

To elucidate the rate-limiting steps of PAs biosynthesis in tea plant, numerous studies have been conducted, and a few genes involved in PAs turnover have been characterized. Among these genes, tea
caffeine synthase 1 (TCS1) was first isolated in tea leaves and shown to control caffeine accumulation by two molecular macular mechanisms: with reduced transcription level or the encoded proteins with only N-3 methylation activity [8, 9]. Recently, the theacrine synthase CkTcs with key residues that are required for the N9-methylation, was confirmed to catalyze the final methylation steps during theacrine synthesis in Kucha, and thus were suggested as rate-limiting enzyme in PAs biosynthesis [10].

Usually, caffeine is the predominant purine alkaloids (2-5%) in young leaves of tea plant, followed by theobromine [11]. In sharp contrast, theacrine is the major purine alkaloid in Kucha, which is a special tea resource first disclosed in Yunnan province [12]. Tracer experiments concluded that theacrine was synthesized from the downstream branch of caffeine metabolism pathway [7]. In this sense, Kucha can be considered as a class of mutants of normal tea and thus are ideal materials to elucidate PAs biosynthesis pathways and identify related candidate genes.

RNA sequencing (RNA-seq) has become a useful tool for identifying and characterizing secondary metabolism genes and their pathways in plants [13-15]. The recent release of complete tea genome sequence and detailed annotations allowed exploratory analysis of DEGs associated with important secondary metabolisms [16]. In this context, we established complete transcriptome of Kucha (Kucha-6 and Kucha1-11) and conventional varieties (YingHong9 and Qingxin1) using an Illumina HiSeq™ 2500 platform. Based on this data, potential candidate genes encoding oxidoreductase and methyltransferase were revealed. This study will provide useful gene resources for the investigations of PAs biosynthesis and genetic improvement of tea plant.

Results

Quantification of purine alkaloids in different genotypes

Purine alkaloids determination showed that the total purine alkaloids content was different in the four genotypes, highest in YH9 (45.57 mg/g) and lowest in K11 (28.87 mg/g). The content of caffeine (11.33-42 mg/g) was relatively abundant in all genotypes, whereas the content of theobromine (2.46-4.56 mg/g) was low. On the other hand, the theacrine presented the greatest difference, with more than 27-fold higher in Kucha (13.16 mg/g in K6 and 14.94 mg/g in K11, respectively) compared with the conventional varieties (0.48 mg/g in YH9 and 0.37 mg/g in QX1, respectively) (Fig. 1 and Table S1).

Transcriptome data analysis

Totally, 12 libraries were constructed and designated as follows: YH9-1 to 3, QX1-1 to 3, K6-1 to 3 and K11-1 to 3. After stringent quality filtering, the transcriptome sequencing produced ~ 671.61 million of high quality (HQ) clean reads (2´150 bp) (average: about 55.97 million reads for each sample). The averaged Q30 and GC content value were 96.67% and 45.50%, respectively (Table 1 and S2).

HQ clean reads, excluding rRNA, were mapped to the tea plant reference genome (cv. Shuchazao). As a result, the total align ratio was ranked from 71.42% to 76.87%, with a uniquely mapped rate of 69.22% to
74.53%. As assembled transcriptome were annotated according to the 33,932 reference genes, a total of 30,369 known genes and 8,696 novel genes were identified in all samples (Table 1 and S2). For the annotation of novel genes, a total of 623 (7.16%) novel genes could be identified in the Nr database. Additionally, KEGG enrichment analysis demonstrated that “Metabolism” and “Genetic Information Processing” were the main categories (Fig. S1).

The pair-wise Pearson’s correlation coefficients were computed to test the reproducibility of the sequencing data. As shown in Fig. 2A, the robust correlation values within each genotype indicated high repeatability between replicates. PCA analysis indicated that the major transcriptomic variation (86.8%) could be explained by the first two principal components (Fig. 2B). The PCA cleanly clustered the samples into four groups, with the assignment of QX1 similar to that of K6.

Identification of differentially expressed genes (DEGs)

The DEGs were determined between Kucha and conventional varieties using a pairwise approach. In detail, 10,732 (4,298 unregulated and 6,434 downregulated) and 11,461 (5,077 upregulated and 6,384 downregulated) DEGs were identified in YH9 vs K6 and YH9 vs K11 comparisons, respectively. The total number of DEGs in QX1 vs K6 and QX1 vs K11 comparisons were 6,032 (3,091 unregulated and 2,941 downregulated) and 9,290 (5,209 upregulated and 4,081 downregulated), respectively. Based on the analysis, the number of downregulated DGEs was higher than the upregulated DEGs in the comparisons of YH9 vs K6 and YH9 vs K11, but lower in QX1 vs K6 and QX1 vs K11 comparisons. Interestingly, the number of DEGs in QX1 vs K6 was far less than other comparisons (Fig. 3A). Moreover, we obtained 1,142 common DEGs from the four comparisons using a Venn diagram (Fig. 3B).

Functional characterization of DEGs

To understand the functions of the obtained DEGs, GO subcategory and KEGG pathway analyses were performed. Among these DEGs, 6,606 annotated genes were assigned into one or more GO terms under three domain categories (Fig. 3C). Within the biological process category, the top-enriched GO terms were metabolic process (3,110, 47.08%), cellular process (2,893, 43.79%) and single-organism process (2338, 35.39%). Within the molecular function, the terms related to catalytic activity (3,732, 56.49%) and binding (2,776, 42.02%) were the most enriched. For the cellular component, the dominant terms were membrane (2,342, 35.45%) and cell (2,105, 31.86%), followed by cell part (2,078, 31.46%). A total of 5,054 DEGs were distributed into 134 KEGG pathways of five main categories, among which the largest categories were metabolism and genetic information processing, followed by organismal systems, environmental information processing and cellular processes. (Fig. 3D). The most represented pathways were “Metabolic pathway” (1972, 39.02%), “Biosynthesis of secondary metabolites” (1298, 25.68%), “Plant-pathogen interaction” (467, 9.24%), “Biosynthesis of antibiotics” (440, 8.71%) and “Microbial metabolism in diverse environments” (362, 7.16%). Additionally, “Purine metabolism” pathway accounted for a considerable portion of the DEGs (125, 2.47%) (Table S3).

Putative genes involved in the theacrine biosynthesis
The putative pathway of theacrine biosynthesis was proposed according to the KEGG database (Fig. 4). In brief, the formation of theacrine is initiated by the assembly of xanthosine, which act as methyl acceptor for caffeine biosynthesis, then conversion of caffeine to theacrine occurs by sequential oxidation and methylation steps with 1,3,7-trimethyluric acid acting as the intermediate. Based on this, we grouped genes that are related to theacrine accumulation into three categories: synthesis of xanthosine, synthesis of caffeine and synthesis of theacrine.

A total of 38 genes related to enzymes of xanthosine synthesis were identified, including four genes encoding adenosine kinase (ADK), 12 genes encoding adenine phosphoribosyltransferase (ARPT), seven genes encoding AMP deaminase (AMPD), two genes encoding IMP dehydrogenase (IMPDH), 11 genes encoding 5'-Nase and two genes encoding guanine deaminase (GAD) (Table S4). Expression pattern analysis showed that these genes were expressed quite differently among different genotypes (Fig. 5A). In particular, three genes assigned as common DEGs, including one ADK gene (TEA022322), one AMPD gene (TEA007189) and one 5'-Nase gene (XLOC_057370), were highly expressed in Kucha.

For the 13 caffeine synthase (TCS) genes involved in caffeine synthesis, TEA015791 and TEA028050 exhibited very high expression level in all experimental materials (FPKM > 700), while two common DEGs (TEA010054 and TEA022559) were highly expressed in Kucha (Fig. 5B and Table S5). Although most of the five genes encoding S-adenosylmethionine synthetase (SAMS) in the methyl donor synthesis exhibited diverse expression patterns, none was classified as common DEGs (Fig 5C and Table S5).

In the case of the conversion of caffeine to theacrine, genes related to enzymes of oxidoreductase (OR) and methyltransferase (MT) might be bottlenecks, and it would require up-regulated in high theacrine content genotypes. Thus, we focused on common DEGs encoding the enzymes mentioned above. The data showed that 44 genes annotated as OR, of which 20 genes were highly expressed in Kucha (Fig. 5D and Table S6). Meanwhile, a total of 21 genes annotated as MTs were identified and seven of them showed enhanced expression levels in Kucha (Fig 5E and Table S7). Furthermore, four of the above-mentioned up-regulated MTs were grouped into N-methyltransferase (TEA009098, TEA010054, TEA022559, XLOC_035717), two into O-methyltransferase (TEA030958, XLOC_047820) and one into C-methyltransferase (TEA023802).

Validation of gene expression using qRT-PCR

To validate the quantification of the RNA-seq analysis in this study, eight genes including one IMPDH gene (TEA024175), three TCS genes (TEA010054, TEA022559 and TEA030024), two MT genes (TEA023802 and XLOC_047820) and two OR genes (TEA007843 and XLOC_026296) were selected for RT-qPCR analysis. As shown in Fig. 6, all the genes, except for TEA007843, shown a similar expression pattern in qRT-PCR analysis as obtained in the RNA-seq results, indicating the reliability of our RNA-seq data.

Discussion
Kucha, a kind of special germplasm resource found in individual areas of China, is often used as traditional medicine by local resident [17]. Though it is difficult to discern them from conventional varieties because of the morphological similarity, the remarkably bitter taste in both fresh leaves and made tea made Kucha very distinctive. The theacrine, in spite of showing varied content in different sources of Kucha, is defined as a characteristic substance and account for the bitterness in this unique germplasm [10, 12, 18-21]. In recent years, theacrine has attracted increasing attention for its potential therapeutic applications in the pharmaceutical industry, and thus motivated research efforts to unveil the molecular basis underlying theacrine biosynthesis. For instance, two RNA-seq based studies have reported that MT genes were involved in the final step of theacrine formation [10, 20]. However, oxidoreductase-encoding genes that catalyze the critical step before methylation have not been mentioned yet.

Guangdong province, an important tea producing region in China, is rich in diverse tea germplasms. Earlier, we collected Kucha germplasms throughout this region and found that K6 and K11 displayed the highest concentration of theacrine. Hence, we performed transcriptome analysis with interest in identifying possible genes involved in theacrine biosynthesis. A total of 671.61 million of HQ clean reads were generated in this study, which was adequate for accurate quantification of gene expression pattern [22]. The alignment rate to the reference genome was quite comparable between Kucha and conventional varieties, and the percentage of annotated genes was similar to that recently reported in tea plant [23, 24], suggesting that the relatively conserved functions and our project has captured the majority of the tea transcriptome. Comparative transcriptome profiling of Kucha and conventional varieties identified 19,095 DEGs and 1,142 of them was assigned as common DEGs. Surprisingly, the number of DEGs in QX1 vs K6 was nearly half that of other comparisons. Besides, the PCA analysis indicated that K6 was relatively close to QX1 rather than K11. These results illustrated the fact that tea has not undergone long-term artificial directional selection and frequent genetic exchange may occur between K6 and parental sources of QX1 due to the geographical proximity [25].

The current research on the metabolic pathway of theacrine in tea plant is relatively clean [1, 5, 7]. We divided the entire pathway into three modules, i.e. xanthosine synthesis pathway, caffeine synthesis pathway and theacrine synthesis pathway, according to the formations of key intermediates and the final product (Fig. 4). In this paper, the expression pattern of xanthosine biosynthesis-related genes that encoding ADK, ARPT, AMPD, 5’-Nase and GAD were investigated. Considering the variation of total purine alkaloids content among the tested materials, we proposed that the dynamic expression level of these genes in different routes may influence an adequate supply of xanthosine for the downstream pathway. In particular, IMPDH was previously reported to be implicated as an inhibitor of the enzyme reduced the rate of caffeine biosynthesis in tea and coffee leaves [26]. Our data showed that the expression pattern of an IMPDH gene (TEA024175) was consistent with the total purine alkaloids content among the four genotypes. TCS is the most important enzyme that catalyzed the methylation of xanthosine to caffeine [8, 9, 27]. The RNA-seq data revealed that two TCS genes (TEA015791 and TEA028050) showed strong expression levels (FPKM > 700) in all samples, implying their important and conserved role in the formation of caffeine. SAM is the methyl donor for the three methylation steps in the caffeine
biosynthetic pathway [28]. However, the expression pattern of SAMS genes did not show a consistent trend with purine alkaloids in our study, which might be due to the availability of SAM is not a principal factor in the control of caffeine biosynthesis in our four genotypes.

Caffeine converts to theacrine by oxidation at the C8 and methylation at N9 positions, and genes related to enzymes of OR and MT might control these steps [7]. Based on the annotation results, we screened 20 OR genes and seven MTs genes that were significantly enhanced in Kucha. A former RNA-seq analysis of Niedu Kucha from Jiangxi province showed that an NMT gene (TEA024443) was highly expressed in high-theacrine sample and therefore was suggested to be involved in the final synthesis process of theacrine [20]. Unfortunately, this gene displayed consistent moderate expression levels in K6, K11 and QX1 (FPKM: 5.00, 8.89, 8.72, respectively), but relatively high in YH9 (FPKM: 14.92) in our study. Recently, the TCS gene (CkTcs) encoding N9-methyltransferase was proved to catalyze the last step of theacrine formation [10]. Among the four common DEGs encoding N-methyltransferase in the current research, two genes (TEA010054 and TEA022559) showing homology to CkTcs seems most likely to be involved in the last step leading to theacrine biosynthesis from 1,3,7-trimethyluric acid. Thus, further functional characterization of these genes and their substrate specificity assay is required to determine whether they are genotype-specific in catalyzing the rate-limiting steps of theacrine.

Conclusions

Our study comprehensively reported transcriptomic resource from Kucha and conventional varieties. A total of 19,095 DEGs were obtained, of which 6,606 and 5,054 genes were annotated based on the GO and KEGG databases, respectively. A Venn diagram of the DEGs showed that 1,142 genes were shared between all pairwise comparisons. Genes involved in putative theacrine biosynthesis pathway were identified. In particularly, candidate genes encoding oxidoreductase and methyltransferase that might catalyze the last two steps of theacrine formation should provide insight into the molecular mechanisms underlying the high level of theacrine in Kucha. Taken together, the target genes produced in this study can be good candidates for genetic improvement and metabolic engineering studies in the future.

Methods

Plant materials

Two Kucha genotypes (Kucha-6 (K6) and Kucha-11 (K11)) and two conventional varieties (Yinghong 9 (YH9) and Qingxin 1 (QX1)) with extremely contrasting theacrine content were used in this study. Both genotypes were grown in Germplasm Bank of Tea Plant in Guangdong Province (Yingde, Guangdong, China, 24°18′N;113°23′E), and details on agro-techniques (fertilization and irrigation) are available upon request. The first leaves of each genotype were harvested on May 2018. Part of the materials were frozen immediately in liquid nitrogen and stored an -80 °C for RNA isolation. The remaining samples were subject to purine alkaloids (caffeine, theobromine and theacrine) analysis. Tea Research Institute,
Guangdong Academy of Agricultural Sciences owned the land/materials and had approved the study. No specific permissions were required for these locations/activities.

**Extraction and HPLC determination of purine alkaloids**

The harvested leaves were freeze-dried and ground finely. Purine alkaloids was extracted with H₂O (1/100, w/v) in a water bath at 100 °C for 30 minutes and centrifuged at 5000×g for 5min. Then the obtained supernatant was metered and filtered through a 0.22 μm membrane prior to HPLC analysis. Purine alkaloids were determined using Agilent 1200 HPLC system (Agilent Technologies, USA). The HPLC was made to run with a column temperature of 35 °C, a wavelength of 231nm, an injection volume of 10 μL, and flow rate of 1.0 mL/min. 5 μm Zorbax SB-C18 column (Agilent) with solvent A (0.1% formic acid) and solvent B (acetonitrile) used as mobile phase. The gradient programs performed as follows: 0-15 min, linear gradient from 4% to 6% A; 15-30min, linear gradient from 6% to 12% A; 30-55min, linear gradient from 12% to 18% A; 55-58min, linear gradient from 18% to 4% A. All samples were analyzed using three biological replicates.

**RNA extraction, library construction and RNA-Seq analysis**

Total RNA was extracted using a TRIzol Kit following the manufacturer's protocol (Promega, Beijing, China). DNase I (TaKaRa, Dalian, China) was used to remove residual gDNA. RNA quality was assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). RNA samples with 260 nm/280 nm ratios between 1.8 and 2.2 were used for further analysis. Next, mRNA was enriched by Oligo (dT) beads, and sequencing libraries were created using an Illumina Truseq™ RNA Sample Prep Kit (Illumina, SD, USA) according to the standard high-throughput protocol. In total, 12 libraries (i.e., four genotypes, with three biological replicates) were sequenced on the Illumina HiSeq™ 2500 platform by Genedenovo Biotechnology Co., Ltd (Guangzhou, China).

The original reads were filtered with the following parameters: 1) reads containing adaptors, 2) reads with N (unknown base) ratio more than 10%, 3) reads containing more than 50% of low quality (Q-value ≤ 20) bases (Jiang et al., 2019). Then, the obtained clean reads were aligned to rRNA database to remove the remaining rRNA reads using Bowtie2 [29]. The high quality clean reads were mapped to reference genome by TopHat2 (version 2.0.3.12) with maximum read mismatch of 2, distance between mate-pair reads of 50 bp and error of distance between mate-pair reads of ± 80 bp [16, 30]. We adopted Cufflinks software to perform reconstruction of transcripts and the reference annotation based transcripts (RABT) program was preferred [31]. Genes with classcode “u” (the transcripts were unknown or in the intergenic spacer region) with transcript length longer than 200 bp and exon number more than 2 were defined as reliable novel genes. Novel genes were further annotated through alignment with the Nr and KEGG databases.

Gene abundances were quantified by software RSEM [32]. The gene expression level was normalized by using FPKM (Fragments Per Kilobase of transcript per Million mapped reads) method, which is able to eliminate the influence of different gene lengths and sequencing data amount on the calculation of gene
expression. Pearson correlation coefficient and PCA on all samples were performed with R package gmodels (http://www.r-project.org/) in this experience.

**Identification of differentially expressed genes (DEGs)**

To screen out differentially expressed genes (DEGs), the edgeR package (http://www.rproject.org/) was used. We determined genes with a fold change \( \geq 2 \) (i.e. \(|\log_2 FC| \geq 1\)) and a false discovery rate (FDR) < 0.05 in a comparison as significant DEGs. For Gene Ontology (GO) term enrichment analysis, DEGs were then mapped to GO database (http://www.geneontology.org/) with FDR \( \leq 0.05 \) [33]. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis (FDR \( \leq 0.05 \)) was performed using the KEGG database (http://www.genome.jp/kegg/) [34].

**Quantitative real-time PCR (qRT-PCR) validation**

The qRT-PCR was performed to validate the transcriptomic data from the extracted total RAN. Eight genes related to theacrine synthesis were selected and analyzed. The corresponding primers were designed with Primer Premier 5.0 and listed in Table S8. The cDNA was synthesized using the PrimeScript\textsuperscript{TM} RT Master Mix kit (Takara, Dalian, China). PCR amplification was performed using Applied Biosystems 7500 Real-Time PCR System according to the manufacturer's instructions. We chose \textit{GAPDH} gene as the endogenous control [35]. All reactions were run in three technical and three biological replicates. Statistical analysis was performed using the \( 2^{-\Delta \Delta CT} \) method [36].

**Abbreviations**

DEGs, differentially expressed genes; OR, oxidoreductase; MT, methyltransferase; PAs, purine alkaloids; Anase, adenosine nucleosidase; ADK, adenosine kinase; ARPT, adenine phosphoribosyltransferase; AMPD, AMP deaminase; IMPDH, IMP dehydrogenase; 5′-Nase, 5′-nucleotidase; GAD, 7-NMT, 7-Methylxanthosine; guanine deaminase; MXN, methylxanthosine nucleotidase; TCS, caffeine synthase; CkTcS, theacrine synthase; FPKM, fragments Per Kilobase of transcript per Million mapped reads.

**Declarations**

**Acknowledgements**

Not Applicable.

**Authors’ contributions**

HJL conceived and designed the project; KXF and DDQ contributed the cultivation and the collection of samples; XHJ and QSW provided bioinformatics support; CDP, BL and QW performed the experiments; HJL, and HLW wrote the manuscript.

**Funding**
This work was financially supported by Key-Area Research and Development Program of Guangdong Province [2020B020220004], the National Modern Agriculture Technology System [CARS-19], Found of President of Guangdong Academy of Agricultural Sciences [201830], Special Funds for Scientific Innovation Strategy Construction of High Level Academy of Agriculture Science [R2016YJ-YB3002, R2017PY-QY009, R2018QD-99, R2018QD-100], and Discipline Team Building Projects of Guangdong Academy of Agricultural Sciences in the 13th Five-Year Period [201619TD].

Availability of data and materials

All data generated or analyzed during this study are included in this published article and the supplementary information files. The sequence data was deposited in the NCBI Sequence Read Archive (SRA) with accession number SRP271859.

Ethics approval and consent to participate

Not Applicable.

Consent for publication

Not Applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Ashihara H, Crozier A. Biosynthesis and metabolism of caffeine and related purine alkaloids in plants. Adv. Bot. Res. 1999; 30, 117-205. https://doi.org/10.1016/S0065-2296(08)60228-1.
2. Li YF, Chen M, Wang C, Li XX, Ouyang SH. et al. Theacrine, a purine alkaloid derived from Camellia assamica kucha, ameliorates impairments in learning and memory caused by restraint-induced central fatigue. Journal of Functional Foods. 2015; 16, 472-483. https://doi.org/10.1016/j.jff.2015.05.003.
3. Guo X, Wu MZ, Huang YR, et al. Experimental study of theacrine on antidepressant effects. Chinese Pharmacological Bulletin. 2009; 25(9), 1160-1163.
4. Qiao H, Ye X, Bai X, He J, Li T, et al. Theacrine: a purine alkaloid from Camellia assamica kucha with a hypnotic property via the adenosine system. Neurosci. Lett. 2017; 659, 48-53. DOI: 10.1016/j.neulet.2017.08.063.
5. Ashihara H, Kato M, Crozier A. Distribution, biosynthesis and catabolism of Methylxanthines in Plants. Handbook of experimental pharmacology. 2011; 11-31.
6. Kato M, Kanehara T, Shimizu H, Suzuki T, et al. Caffeine biosynthesis in young leaves of Camellia sinensis In vitro studies on N-methyltransferase activity involved in the conversion of xanthosine to
caffeine. Physiol. Plant. 1996, 98 (3), 629-636. https://doi.org/10.1111/j.1399-3054.1996.tb05720.x
7. Zheng XQ, Ye CX, Kato M, Crozier A, Ashihara H. Theacrine (1, 3, 7, 9-tetramethyluric acid) synthesis in leaves of a Chinese tea, kucha (Camellia assamakuchaa). Phytochemistry. 2002; 60(2), 129-134. DOI: 10.1016/s0031-9422(02)00086-9.
8. Kato M, Mizuno K, Crozier A, et al. Caffeine synthase gene from tea leaves. Nature. 2000; 406, 956-957. https://doi.org/10.1038/35023072.
9. Jin JQ, Yao MZ, Ma CL, Ma JQ, Chen L. Natural allelic variations of TCS1 play a crucial role in caffeine biosynthesis of tea plant and its related species. Plant Physiology and Biochemistry. 2016; 18-26. DOI: 1016/j.plaphy.2015.12.020.
10. Zhang YH, Li YF, Wang YJ, et al. Identification and characterization of N9-methyltransferase involved in converting caffeine into non-stimulatory theacrine in tea. Nat. Commun. 2020; 11, 1473. https://doi.org/10.1038/s41467-020-15324-7.
11. Wan XC. Tea biochemistry. China Agriculture Press, 2013.
12. Ye CX, Lin YC, Su JY, et al. Purine Alkaloids in Camellia assamakuchaa Chang et Wang. Acta scientiarum naturalium Universitatis Sunyatseni. 1999; 38(05):82-86.
13. Zhang YZ, Wei K, Li HL, Wang LY, et al. Identification of key genes involved in catechin metabolism in tea seedlings based on transcriptomic and HPLC analysis. Plant Physiology and Biochemistry. 2018; 133, 107-115. DOI: 10.1016/j.plaphy.2018.10.029.
14. Guo YQ, Chang XJ, Zhu C, Zhang ST, Li XZ. De novo transcriptome combined with spectrophotometry and gas chromatography-mass spectrometer (GC-MS) reveals differentially expressed genes during accumulation of secondary metabolites in purple-leaf tea (Camellia sinensis cv Hongyafoshou), The Journal of Horticultural Science and Biotechnology. 2019; 94:3, 349-367. DOI: 1080/14620316.2018.1521708.
15. Zhou T, Luo XJ, Yu CN, Zhang CC, et al. Transcriptome analyses provide insights into the expression pattern and sequence similarity of several taxol biosynthesis-related genes in three Taxus BMC Plant Biol. 2019; 19, 33. https://doi.org/10.1186/s12870-019-1645-x.
16. Wei C, Yang H, Wang S, Zhao J, Liu C, Gao L, Xia E, Lu Y, Tai Y, et al. Draft genome sequence of Camellia sinensisinensis provides insights into the evolution of the tea genome and tea quality. Proc. Natl. Acad. Sci, USA. 2018; 115(18):201719622. https://doi.org/10.1073/pnas.1719622115.
17. Wang XC, Yao MZ, Ma CL, et al. Analysis and evaluation of biochemical components in bitter tea plant germplasms. Chinese Agricultural Science Bulletin. 2008; 24(6):65-69.
18. Li KK, Shi XG, Yang XR, et al. Antioxidative activities and the chemical constituents of two Chinese teas, Camellia kucha and ptilophylla. International Journal of Food Science and Technology. 2012; 47, 1063-1071. https://doi.org/10.1111/j.1365-2621.2012.02942.x
19. Jin JQ, Zhou CY, Ma CL, et al. Identification on purine alkaloids of representative tea germplasm in China. Journal of Plant Genetic Resources. 2014; 15(2): 279-285.
20. Wang SL, Chen JD, Ma JQ, et al. Novel insight into theacrine metabolism revealed by transcriptome analysis in bitter tea (Kucha, Camellia sinensis). Sci Rep. 2020; 10(1):6286. Doi: 10.1038/s41598-
21. Yang C, Ma CQ, Zhou BX, et al. Characteristic components research in bitter tea. Journal of Kunming University. 2014; 36(3):75-78.

22. Vijay N, Poelstra JW, Künstner A, Wolf JB. Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive in silico assessment of RNA-seq experiments. Molecular Ecology. 2013; 22(3), 620-634. DOI: 1111/mec.12014.

23. Zhu BY, Chen LB, Lu MQ, Zhang J, Han JY, Deng WW, Zhang ZZ. Caffeine Content and Related Gene Expression: Novel Insight into Caffeine Metabolism in Camellia Plants Containing Low, Normal, and High Caffeine Concentrations. Journal of Agricultural and Food Chemistry. 2019; 67(12), 3400-3411. https://doi.org/10.1021/acs.jafc.9b00240.

24. Zhou CB, Mei X, Rothenberg DO, Yang ZB, et al. Metabolome and transcriptome analysis reveals putative genes involved in anthocyanin accumulation and coloration in white and pink tea (Camellia sinensis) flower. Molecules. 2020; 25(1), 40. https://doi.org/10.3390/molecules25010190.

25. Zhang WY, Zhang YJ, Qiu HJ, Guo YF, Wan HL, et al. Genome assembly of wild tea tree DASZ reveals pedigree and selection history of tea varieties. Nat Commun. 2020; 11, 3719. https://doi.org/10.1038/s41467-020-17498-6.

26. Keya CA, Crozier A, Ashihara H. Inhibition of caffeine biosynthesis in tea (Camellia sinensis) and coffee (Coffea arabica) plants by ribavirin. FEBS Lett. 2003; 554, 473–477. https://doi.org/10.1016/S0014-5793(03)01213-4.

27. Jin JQ, Chai YF, Liu YF, Zhang J, Yao MZ, Chen L. Hongyacha, a Naturally Caffeine-free Tea Plant from Fujian, China. Journal of Agricultural and Food Chemistry. 2018; 66(43), 11311-11319. https://doi.org/10.1021/acs.jafc.8b03433.

28. Ashihara H, Sano H, Crozier A. Caffeine and related purine alkaloids: biosynthesis, catabolism, function and genetic engineering. Phytochemistry. 2008; 69(4), 841-856. DOI: 10.1016/j.phytochem.2007.10.029.

29. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 2012; 9, 357–359. https://doi.org/10.1038/nmeth.1923.

30. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14, R36. https://doi.org/10.1186/gb-2013-14-4-r36.

31. Trapnell C, Roberts A, Goff LA, Pertea G, Kim D, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols. 2012; 7(3): 562-578. https://doi.org/10.1038/nprot.2012.016.

32. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12(1): 323-323. https://doi.org/10.1186/1471-2105-12-323.

33. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 2008; 4, 44–57.
34. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, et al. KEGG for linking genomes to life and the environment. Nucleic. Acids Res. 2007; 36:480–484. https://doi.org/10.1093/nar/gkm882

35. Lin J, Wilson IW, Ge G, et al. Whole transcriptome analysis of three leaf stages in two cultivars and one of their F₁ hybrid of *Camellia sinensis* with differing EGCG content. Tree Genetics & Genomes. 2017; 13, 13. https://doi.org/10.1007/s11295-016-1095-59.

36. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻ΔΔCT Methods. 2001; 25:402-8. https://doi.org/10.1006/meth.2001.1262.

### Tables

Table 1: Summary of the RNA-Seq data.

|                        | Maximum         | Minimum        | Average       |
|------------------------|-----------------|----------------|---------------|
| HQ clean data No.      | 120,817,998     | 39,462,950     | 55,103,541    |
| Q30 (%)                | 96.98           | 95.05          | 96.67         |
| GC (%)                 | 45.71           | 44.98          | 45.50         |
| Unique Mapped Reads (%)| 74.53           | 69.22          | 71.28         |
| Multiple Mapped reads (%) | 2.42          | 2.13           | 2.28          |
| Mapping Ratio (%)      | 76.87           | 71.42          | 73.56         |
| Total Gene             | 39,065          |                |               |

### Figures
Figure 1

Purine alkaloids content in Kucha (YH9 and QX1) and conventional varieties (K6 and K11).
Figure 2

General transcription patterns of all samples.
**Figure 3**

Identification and functional analysis of differentially expressed genes (DEGs) between Kucha and conventional varieties.
Figure 4

Theacrine biosynthesis pathway.
Figure 5

Heat map of key genes related to theacrine accumulation in the four genotypes.
Figure 6

qRT-PCR validation of eight selected genes putatively involved in theacrine biosynthesis.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS8.xlsx
- TableS7.xlsx
- TableS6.xlsx
- TableS5.xlsx
- TableS4.xlsx
- TableS3.xlsx
- TableS2.xlsx
- Fig.S1.docx
- TableS1.xlsx