

OPEN

# Novel insight into theacrine metabolism revealed by transcriptome analysis in bitter tea (*Kucha, Camellia sinensis*)

Songlin Wang, Jiedan Chen, Jianqiang Ma, Jiqiang Jin, Liang Chen\* & Mingzhe Yao\*

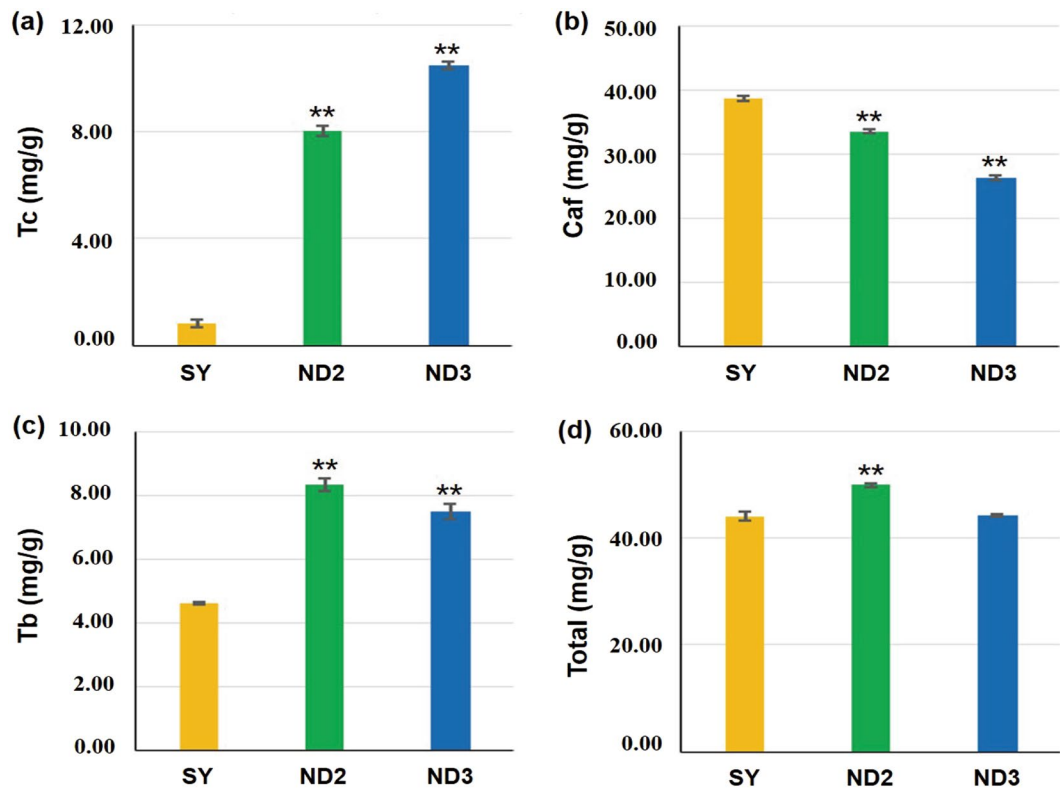
*Kucha (Camellia sinensis)* is a kind of unique wild tea resources in southwest China, containing sizeable amounts of theacrine (1,3,7,9-tetramethyluric acid) and having a special bitter taste both in fresh leaves and made tea. Theacrine has good healthy function locally. But the molecular mechanism of theacrine metabolism in *Kucha* was still unclear. In order to illuminate the biosynthesis and catabolism of theacrine in *Kucha* plants, three tea cultivars, *C. sinensis* 'Shangyou Zhongye' (SY) with low-theacrine, 'Niedu *Kucha* 2' (ND2) with middle-theacrine and, 'Niedu *Kucha* 3' (ND3) with high-theacrine, were used for our research. Purine alkaloid analysis and transcriptome of those samples were performed by High Performance Liquid Chromatography (HPLC) and RNA-Seq, respectively. The related gene expression levels of purine alkaloid were correlated with the content of purine alkaloid, and the results of quantitative real-time (qRT) PCR were also confirmed the reliability of transcriptome. Based on the data, we found that theacrine biosynthesis is a relatively complex process, *N*-methyltransferase (*NMT*) encoded by TEA024443 may catalyze the methylation at 9-*N* position in *Kucha* plant. Our finding will assist to reveal the molecular mechanism of theacrine biosynthesis, and be applied to selection and breeding of *Kucha* tea cultivars in the future.

*Kucha*, a unique kind of tea varieties (*Camellia sinensis*) with bitter taste in both fresh leaves and made tea, was firstly found in Yunnan province, most of which located in the adjoin regions between Guangxi, Guangdong, Hunan and Jiangxi provinces in China<sup>1,2</sup>. *Kucha* plants are usually semi-arbor trees with large leaves, which are difficult to be discriminated from cultivated tea varieties based on morphology observation. However, the leaves of *Kucha* are remarkably characterized with strongly bitter taste and special aroma compared to other tea cultivars. In the *Kucha* growing regions, local residents popularly use its leaves as a medicine to treat wound, inflammatory and diarrhea in daily life.

The tea plant has rich amount of purine alkaloids, such as caffeine and theobromine. Most of tea plants generally contain 25.0 mg/g ~ 45.0 mg/g caffeine, and theobromine has a widely variation range from 0.5 mg/g ~ 46.7 mg/g<sup>3</sup>. A special purine alkaloid, 1,3,7,9-tetramethyluric acid, exists in the *Kucha* leaves, which was significantly related to the formation of bitter taste<sup>4,5</sup>. Li *et al.* found 2.86% (w/w) of theacrine in dry *Kucha* leaves, while no detection in tea leaves of *Cococa* tea (*C. ptilophylla*) and Longjing tea (*C. sinensis*)<sup>6</sup>. Jin *et al.* identified 36 mg/g of theacrine from a *Kucha* plant, while null of theacrine in other tea plants such as *C. crassicolumna*, *C. sinensis* and *C. sinensis* var. *pubilimba*<sup>3</sup>. At present, high purity of theacrine could be separated by HPLC and LC-MS, and obtained by ultrasonic reflux extraction<sup>7,8</sup>. As a unique, naturally occurring purine alkaloid presented in *Kucha*, more and more studies have been focused on its pharmacology and health function. A few of research showed that theacrine had the benefit effects on anti-inflammation, detoxifying, analgesic and promote sleep<sup>9-11</sup>.

The typical route of caffeine biosynthesis in tea leaves is a 'xanthosine → 7-methylxanthosine → theobromine → caffeine' pathway<sup>12,13</sup>. Theacrine is an important purine alkaloid converted from caffeine by hydration, oxidation and methylation. In usual, large quantities of purine alkaloids such as caffeine and theobromine, and little theacrine are synthesized in *C. sinensis* plants. Zheng *et al.* used radioactive <sup>14</sup>C isotope trace method to explore the metabolic pathway of theacrine, which indicate that theacrine was synthesized from caffeine in what

Tea Research Institute of Chinese Academy of Agricultural Sciences, Key Laboratory of Tea Biology and Resources Utilization, Ministry of Agriculture and Rural Affairs, 9 South Meiling Road, Hangzhou, Zhejiang, 310008, China.  
\*email: [liangchen@tricaas.com](mailto:liangchen@tricaas.com); [yaomz@tricaas.com](mailto:yaomz@tricaas.com)



**Figure 1.** The content of purine alkaloids in SY, ND2 and ND3. (a) Tc: theacrine content, (b) Caf: caffeine content, (c) Tb: Theobromine content, (d) Total: total purine alkaloid content. \*The significance of the difference compounds between two comparison groups (SY vs ND2, SY vs ND3) is shown as \*\* $P < 0.01$ .

is probably a three-step pathway with 1,3,7-methyluric acid acting an intermediate<sup>14</sup>. This was the first demonstration about theacrine biosynthesis pathway in Kucha. However, why the theacrine were highly synthesized and concentrated in Kucha plant? And it still remains unclear about the molecular mechanism of theacrine metabolism in Kucha plant.

In order to determine the molecular mechanism of theacrine metabolism, RNA-Seq was performed using three tea cultivars with different theacrine content. They are Shangyou Zhongye (SY) with minor theacrine, Niedu Kucha 2 (ND2) with medium theacrine and Niedu Kucha 3 (ND3) with high theacrine content. In this study, genes related to theacrine metabolism were explored and identified based on transcriptomes analysis. It will assist to explain the theacrine mechanisms in Kucha plant.

## Results

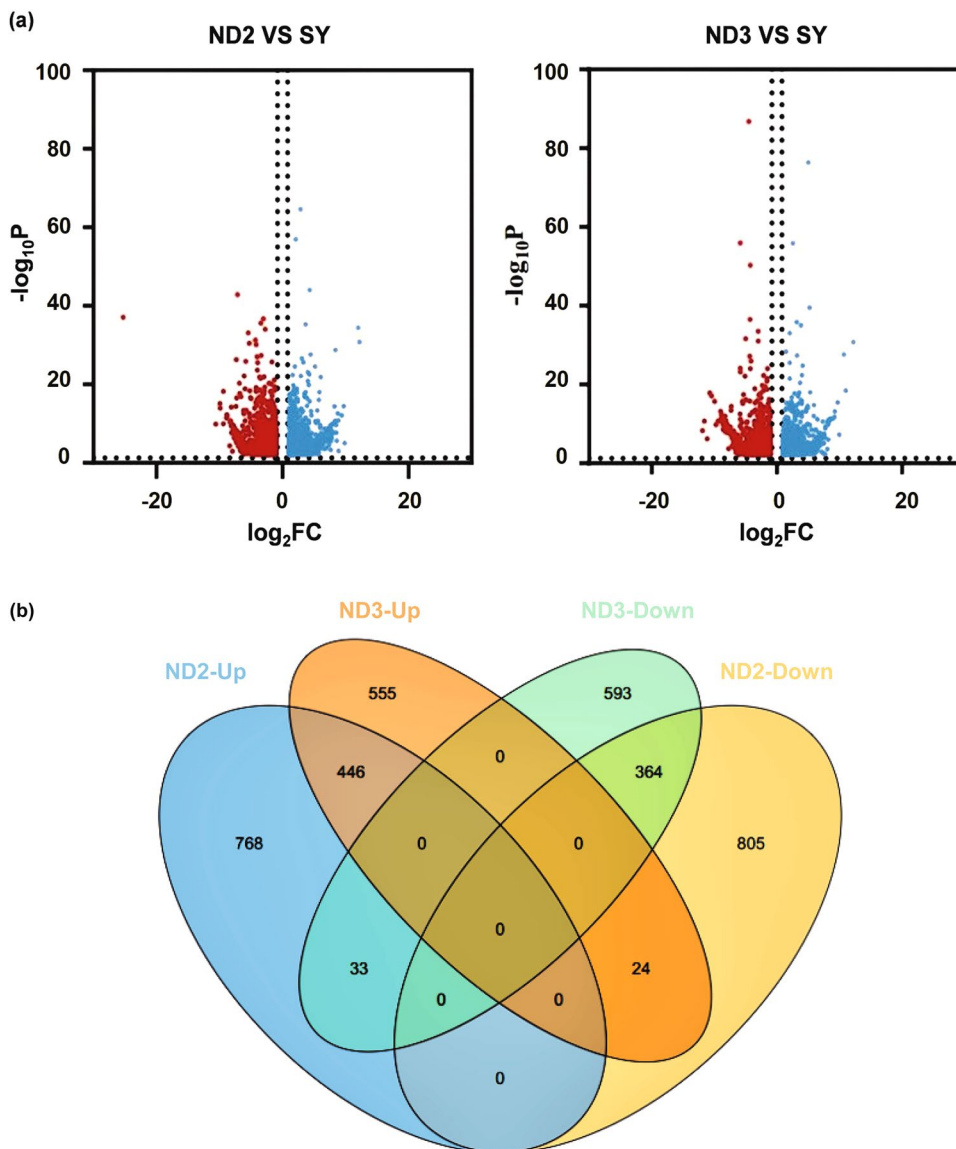
**Analysis of purine alkaloid content.** The total purine alkaloids were relatively stable in tested samples, they were no significant difference between SY and ND3, especially. However, the significant differences of theacrine and caffeine were found between them. The theacrine content were 8.04 mg/g and 10.45 mg/g in ND2 and ND3 respectively, significant higher than SY (0.83 mg/g). While the variation of caffeine content was opposite to theacrine between the tested samples (Fig. 1). The caffeine was estimated 38.62 mg/g, 33.47 mg/g in ND2 and ND3 respectively, they were considerably lower than SY (38.62 mg/g). It may due to caffeine was the precursor of the theacrine synthesis, which was partly converted to theacrine in Kucha. Moreover, the content of theobromine in ND2 (8.34 mg/g) and in ND3 (7.49 mg/g) was considerably higher than those in SY (4.60 mg/g).

**Sequencing analysis and reference genome alignment.** Three cDNA libraries of ND2, ND3 and SY were constructed and sequenced. Totally, 18.87 Gb high-quality reads were obtained from paired-end reads, with a single read length of ~150 bp. A total of 41.28, 42.84 and 41.33 million high-quality reads were generated from ND2, ND3 and SY, respectively. Q20 percentage values were over 96.65%, while the Q30 percentage value were more than 92.56%, and the GC content was ranged from 49.10% ~ 49.29% (Table 1). The average mapped ratio of three samples to the genome was 92.28%. These results showed that the obtained transcriptomic data with relatively high comparison to the reference genome could be used for the further research.

**Analysis of different expression genes (DEGs).** A total of 3,588 DEGs were detected among three Kucha cultivars. In ND2 vs SY, 2,440 DEGs were noted, of which 1,247 were upregulated and 1,193 were downregulated. And 2,015 DEGs were identified in ND3 vs SY, of which 1025 were upregulated and 990 were downregulated. While 810 differentially co-expressed genes (446 up-regulated and 364 down-regulated) were identified

Sample	High-quality reads	Clean base (Gb)	Q20 (%)	Q30 (%)	Error (%)	GC content (%)	Total mapped reads	Total Mapped ratio (%)
ND2	41,278,692	6.19	96.65	92.56	0.012	49.24	36,056,312	91.42
ND3	42,839,490	6.42	97.75	94.40	0.006	49.10	31,241,850	91.66
SY	41,331,604	6.20	97.65	94.17	0.005	49.29	30,857,847	93.77

**Table 1.** Quality evaluation of sequencing data in ND2, ND3, SY.

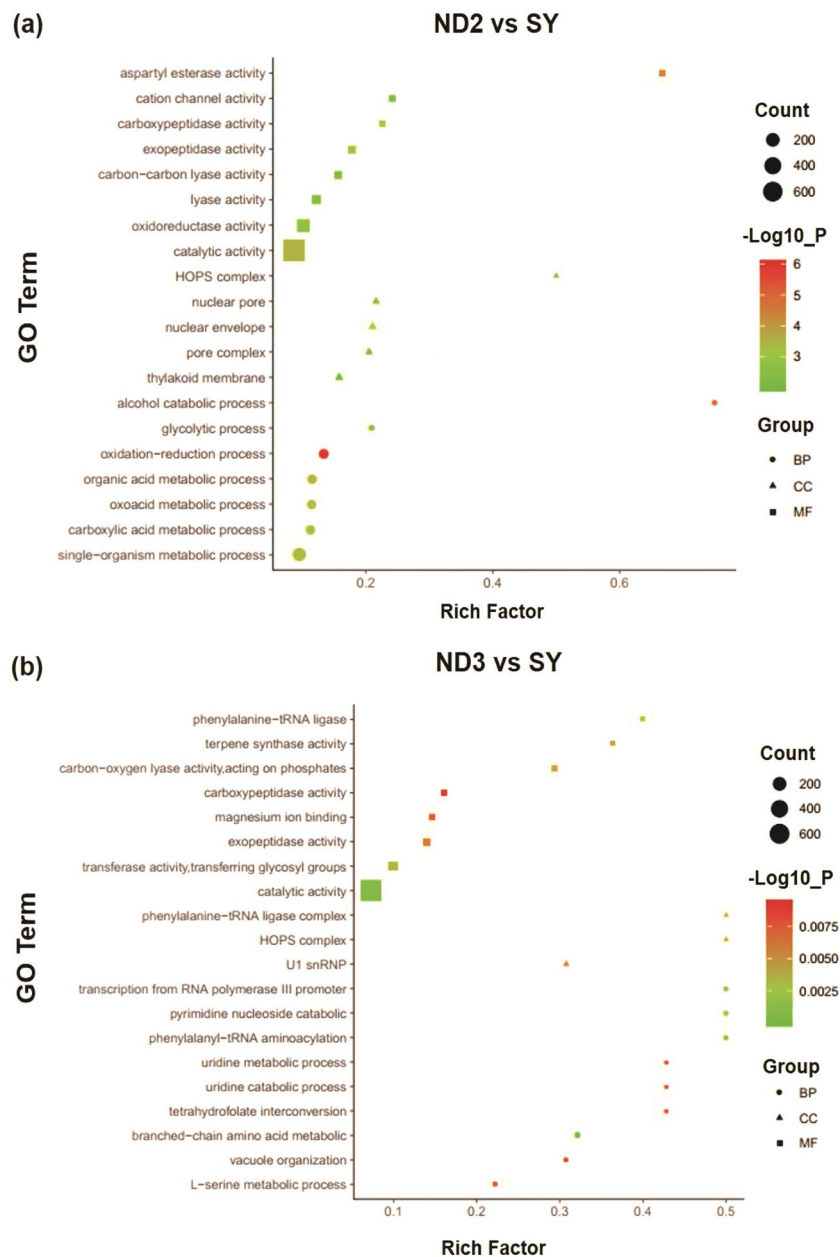


**Figure 2.** (a) Volcano plots of DEGs. (b) Venn diagram of DEGs: up-regulated DEGs and down-regulated DEGs.

between two group of ND2 vs SY and ND3 vs SY (Fig. 2a,b). These genes may reveal the difference in metabolic pathway of purine alkaloid between high- and low- theacrine tea plant.

**GO enrichment analysis of DEGs.** GO (Gene Ontology) is an international standard database for gene functional classification, which contains three categories, molecular function (MF), cellular component (CC) and biological process (BP)<sup>15</sup>. In the MF category, the molecular function, the catalytic activity and the binding were the top three enriched GO terms of two groups. In the CC category, DEGs were enriched in terms of the cellular component, the cell and the cell part. And in the BP category, the biological process, the metabolic process and the cellular process were the top terms.

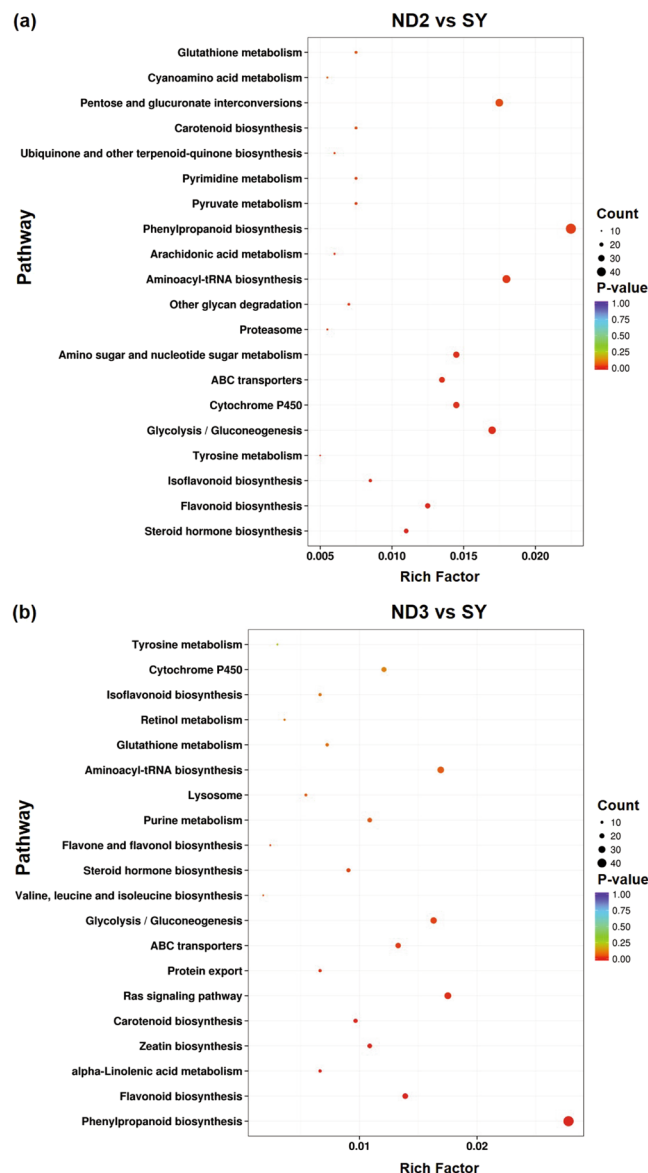
In ND2 vs SY, 1327, 150, and 1465 DEGs were significant enriched into 106, 15 and 48 subcategories of BP, CC and MF, respectively. And in ND3 vs SY, 492, 295, and 840 DEGs were significant enriched



**Figure 3.** The top 20 enriched GO terms of ND2 vs SY (a) and ND3 vs SY (b). The rich factor is the ratio of the number of DEGs and all annotated genes. BP: Biological process, CC: Cellular component, MF: Molecular function.

into 41, 8 and 25 subcategories of BP, CC and MF, respectively. Moreover, the terms of the methylation, the *S*-adenosylmethionine-dependent methyltransferase activity and the *N*-methyltransferase activity were enriched in 23, 11 and 7 genes in two groups, which may play a key role in purine alkaloid biosynthesis. The top 20 enriched terms of GO classification are shown in Fig. 3.

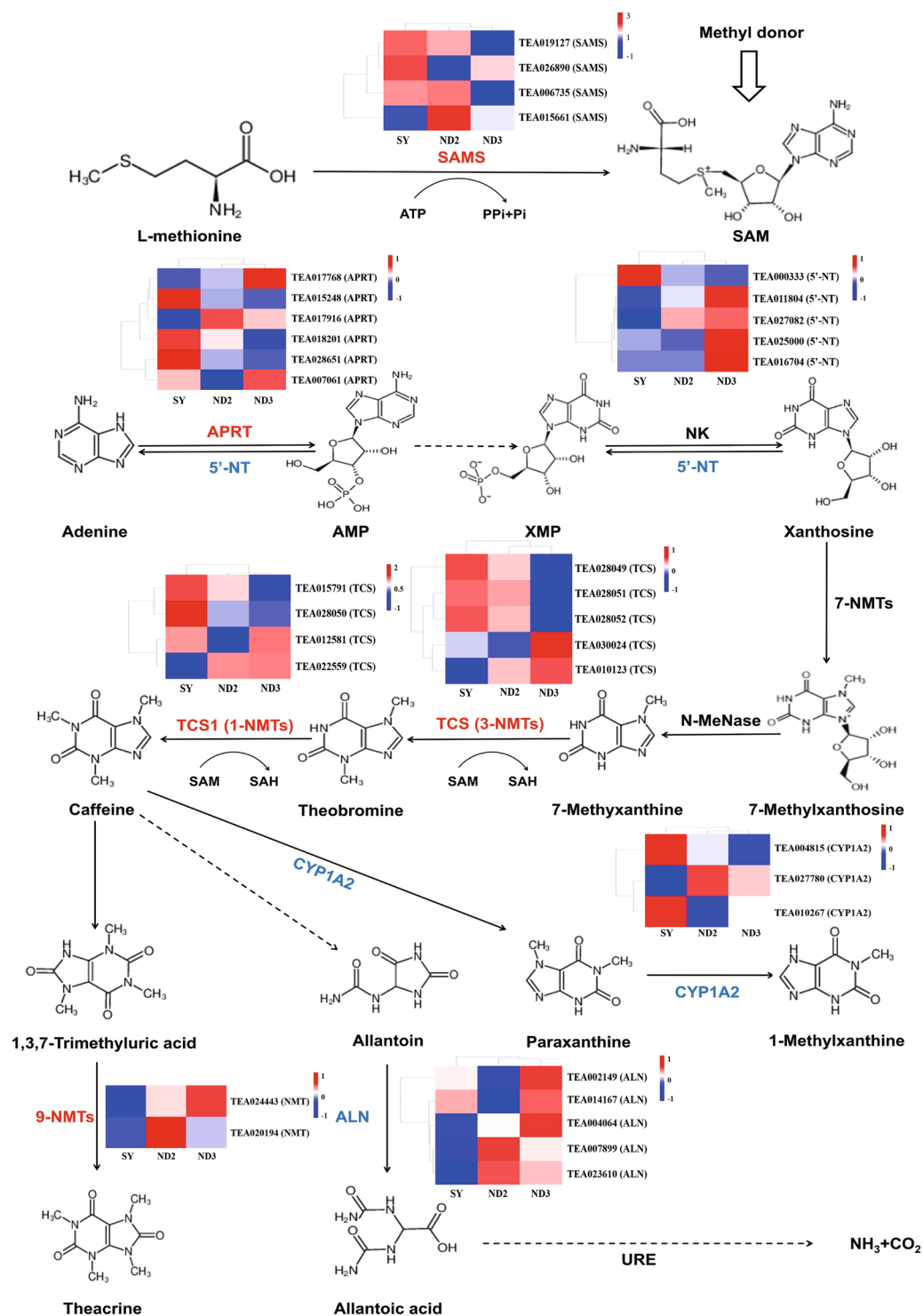
**KEGG enrichment analysis of DEGs.** To explore whether the related genes of metabolism in Kucha were enriched, we constructed the KEGG pathway analyses by the obtained unigenes which corresponding reference pathway<sup>16</sup>. Results showed that 192 and 182 metabolic pathways in ND2 vs SY and ND3 vs SY group, respectively. These pathways contain flavonoid biosynthesis, phenylpropanoid biosynthesis, carotenoid biosynthesis, ABC transporters and steroid hormone biosynthesis were significant enriched in both group (Fig. 4). Moreover, a certain amount of DEGs were noted in secondary metabolite biosynthesis pathway, such as purine metabolism, flavone and flavonol biosynthesis, tyrosine metabolism, tropane, piperidine and pyridine alkaloid biosynthesis and indole alkaloid biosynthesis.



**Figure 4.** The top 20 KEGG pathways of ND2 vs SY (a) and ND3 vs SY (b). The rich factor is the ratio of the number of DEGs and all annotated genes.

**Gene expressions in purine alkaloid metabolism.** Caffeine metabolism was the main pathway of purine alkaloids in *C. sinensis*<sup>17</sup>. Theobromine is the intermediate product in caffeine synthesis process, while theacrine and theophylline are in caffeine degradation process. The pathway of “7-Methyxanthosine → 7-Methyxanthine → Theobromine → Caffeine → 1,3,7-Trimethyluric acid → Theacrine” was the only process for theacrine synthesis at present. The purine salvage process is “Adenosine → AMP → XMP → Xanthosine”. In addition to the process of “caffeine → theacrine”, we also annotated other three caffeine degradation pathway in KEGG database, the final degradation products in this pathway are “NH<sub>3</sub> + CO<sub>2</sub>”, “7-Methyluric acid” and “1-Methyluric acid”, respectively (Fig. 5). The identification of those pathway can help us to further explain the metabolism of purine alkaloids in tea plant.

The expression patterns of key genes in theacrine synthesis such as *TCS*, *SAMS*, *APRT*, *NMT* were conducted to ascertain the reason of high concentration of theacrine in Kucha plant (Fig. 5). TEA015248 and TEA028651 (encoded adenine phosphoribosyltransferase, *APRT*) in the adenine donor synthesis, TEA015791, TEA028050, and TEA028052 (encoded tea caffeine synthase, *TCS*) were highly expressed in SY with high caffeine concentration. TEA006735 and TEA015661 (encoded S-adenosylmethionine synthetase, *SAMS*) in the methyl donor synthesis were up-regulated in ND2 which contains the most purine alkaloid among three samples. The expression levels of TEA024443 (encoded N-methyltransferase, *NMT*) were up-regulated in ND2 and ND3 which had a highly content of theacrine compare with SY. We also identified the key genes of purine alkaloid degradation, TEA027082 and TEA011804 (encoded 5'-nucleotidase, 5'-*NT*) which involved the degradation pathway of “AMP → Adenosine” and “Guanosine → GMP” were up-regulated in ND2 and ND3 compared with SY.



**Figure 5.** Purine alkaloid metabolic pathway. Dotted line indicates that some metabolites have been omitted. The column chart shown that the gene expression of three tea samples. The red genes play a synthetic role and the blue represent degradation. \*SAMS, S-adenosylmethionine synthetase; APRT, adenine phosphoribosyltransferase; NK, nucleoside kinase; 5'-NT, 5'-nucleotidase; 3-NMT and 1-NMT, tea caffeine synthase (TCS); N-MeNase, N-methyl nucleosidase; CYP1A2, cytochrome P450 family 1 subfamily apolypeptide 2; ALN, allantoinase; URE, urease.

TEA002149 (encoded allantoinase, *ALN*) was highly expressed in ND3 with high theacrine concentration. From the KEGG enrichment analysis, 5 and 6 DEGs which annotated in cytochrome P450 family were obtained in SY vs ND2, and SY vs ND3, respectively. TEA010267 and TEA004815 (encoded cytochrome P450 family 1 subfamily A polypeptide 2, *CYP1A2*) were highly expressed in high caffeine concentration sample.

Gene name	Gene ID	Correlation Coefficient		
		TC	CAF	TB
<i>NMT</i>	TEA024443	0.983	-0.986	0.799
<i>NMT</i>	TEA020194	0.224	-0.548	-0.232
<i>SAMS</i>	TEA019127	-0.860	0.984	-0.544
<i>SAMS</i>	TEA006735	-0.648	0.872	-0.243
<i>SAMS</i>	TEA026890	-0.581	0.262	-0.882
<i>SAMS</i>	TEA015661	0.339	0.009	0.721
<i>TCS</i>	TEA028049	-1.000*	0.947	-0.885
<i>TCS</i>	TEA028050	-0.994	0.971	-0.840
<i>TCS</i>	TEA028051	-0.931	1.000*	-0.672
<i>TCS</i>	TEA028052	-0.924	0.999*	-0.659
<i>TCS</i>	TEA030024	0.992	-0.887	0.945
<i>TCS</i>	TEA010123	0.933	-1.000**	0.677
<i>TCS</i>	TEA012581	0.885	-0.992	0.587
<i>TCS</i>	TEA015791	-0.906	0.997	-0.625
<i>TCS</i>	TEA022559	0.987	-0.871	0.956
<i>APRT</i>	TEA017916	0.749	-0.473	0.965
<i>APRT</i>	TEA007061	0.019	-0.365	-0.426
<i>APRT</i>	TEA015248	-0.999*	0.955	-0.871
<i>APRT</i>	TEA028651	-0.973	0.833	-0.974
<i>APRT</i>	TEA018201	-0.936	1.000**	-0.370
<i>APRT</i>	TEA017768	0.837	-0.975	-0.508
<i>5'-NT</i>	TEA011804	0.993	-0.890	0.942
<i>5'-NT</i>	TEA025000	0.668	-0.885	0.269
<i>5'-NT</i>	TEA027082	0.484	-0.151	0.822
<i>5'-NT</i>	TEA016705	0.694	-0.901	0.302
<i>5'NT</i>	TEA000333	-0.999*	-0.916	-0.660
<i>CYP1A2</i>	TEA010267	-0.939	0.762	-0.994
<i>CYP1A2</i>	TEA027780	0.998*	-0.956	0.870
<i>CYP1A2</i>	TEA004815	-0.992	0.975	-0.832
<i>ALN</i>	TEA002149	0.323	-0.631	-0.130
<i>ALN</i>	TEA007899	0.685	-0.390	-0.994
<i>ALN</i>	TEA023610	0.838	-0.597	-0.893
<i>ALN</i>	TEA014167	0.993	-0.303	0.601
<i>ALN</i>	TEA004064	0.972	-0.994	0.285
<i>URE</i>	TEA020308	0.700	-0.409	0.944

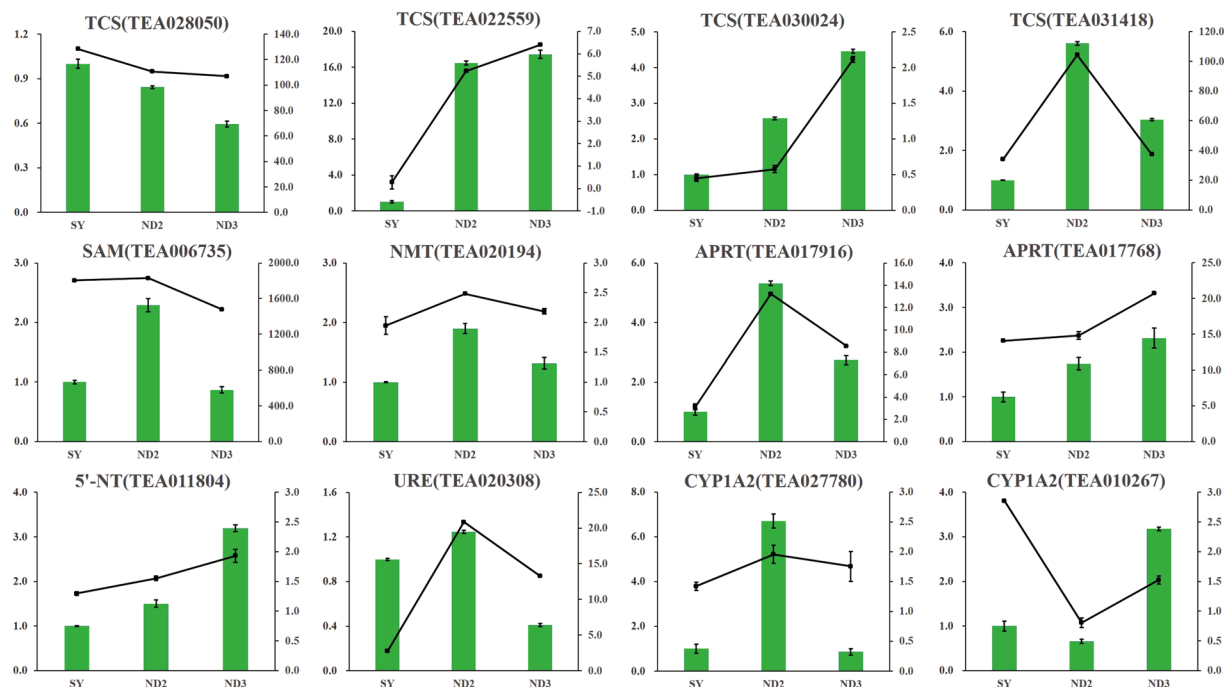
**Table 2.** The correlation coefficient of purine alkaloid and gene expression levels.

Pearson correlation coefficient was used to evaluate the correlation between the gene expression level and purine alkaloid content (Table 2). The results shown that the expression patterns of TEA028049 (*TCS*), TEA015248 (*APRT*) and TEA000333 (*5'-NT*) were negatively correlated with the content of theacrine ( $P < 0.05$ ). Meanwhile, TEA028050 (*TCS*) and TEA028052 (*TCS*) were positively correlate with caffeine content ( $P < 0.05$ ), especially TEA018201 (*APRT*) presents the significantly positive correlation ( $P < 0.01$ ). It was speculated that the proteins encoded by these genes were involved in the regulation of theacrine and caffeine content in the Kucha plants.

**qRT-PCR analysis of the DEGs.** To further confirm the reliability of the gene expression levels, 12 genes which were functionally related to purine alkaloid metabolism were selected for verification by quantitative real-time (qRT) PCR analysis, and their primers are listed in Table S1. Results shown that, 12 genes, except for TEA027780 (*CYP1A2*) and TEA010267 (*CYP1A2*), were generally consistent with RNA-Seq which further indicated the reliability of the transcriptome data (Fig. 6).

## Discussion

As a kind of special tea germplasm, the classification of Kucha still remains being under discussion. In Zhang Hongda classification method, Kucha was defined as *C. sinensis* var. *kucha* Zhang et Wang<sup>18</sup>; Du *et al.* defined Kucha as *C. sinensis* var. *kuiea* Du et Li based on methods of chemical classification and digital taxonomy<sup>19</sup>; Min studied the morphological characteristics of Kucha cultivars, indicated that Kucha should be classified into *C. sinensis* var. *assamica*, not be defined as a new tea variety<sup>20</sup>; Chen *et al.* considered that Kucha should be classified into *C. sinensis* var. *assamica*, according to the systematic study of tea resources and the comprehensive investigation of tea plants in original place<sup>21</sup>; Wang *et al.* determined biochemical components of 24 Kucha resources,



**Figure 6.** qRT-PCR verification of the selected DEGs involved in caffeine and theacrine metabolic pathway.

\*The histogram represent the relative expression of the genes in three sample (ND2, ND3 and SY) by qRT-PCR, and the expression pattern (FPKM) of the genes obtained from RNA-Seq was indicated by the trend line.

indicated that Kucha germplasms were similar to *C. sinensis* var. *assamica* in terms of catechin composition which was reflected the degree of tea plant evolution<sup>22</sup>. For Kucha classification, further study should be focused on genomic level by comparing Kucha plant with other tea species.

Meanwhile, the exploitation and utilization of bitter tea resources are also an important problem faced by breeders. Jianghua Kucha (a bitter tea) has been studied and utilized deeply which was widely promoted in the area of broken black tea at Hunan province, because of its suitable for black tea<sup>23–25</sup>. In our study, High level of purine alkaloids in Kucha cultivars is conducive to the processing of high-quality black tea, and it also provides a new way to develop product of natural purine alkaloid and multi-purpose exploitation of Kucha.

Theacrine is an advantage purine alkaloid second only to caffeine in Kucha plant generally, which affect quality and flavor of Kucha products significantly. Previous studies indicated that the theacrine varied in different Kucha cultivars, and less affected by environments<sup>22</sup>. Ye *et al.* found that the content of theacrine was stable with season change in young leaves of Kucha. While the theacrine was significantly higher in immature leaves than that in mature leaves and aged leaves<sup>4</sup>. At present, metabolic pathway of purine alkaloid and the related genes of biosynthesis have been basically determined which lay the foundation for our research of theacrine metabolism<sup>26,27</sup>. In our study, RNA-Seq was used to discovered the transcriptome of three tea cultivars, DEGs involved in purine alkaloid metabolism were discovered in two comparison groups ND2 vs SY and ND3 vs SY. Contents of theacrine and caffeine among three cultivars were obviously different, theacrine in ND2 and ND3 was higher than SY, while the caffeine content was lower.

KEGG pathway analysis indicated that purine alkaloid metabolism pathway plays significant roles in Kucha metabolism, and a number of key genes related to theacrine and caffeine biosynthesis were selected from these DEGs. Theacrine biosynthesis is catalyzed by a series of enzyme, including *S*-adenosylmethionine synthetase (SAMS), tea caffeine synthase (TCS), *N*-methyltransferase (NMTs) and Adenine phosphoribosyltransferase (APRT). In general, SAMS is the key enzyme in purine alkaloid metabolism pathway, of which *S*-adenosylmethionine act as methyl donor in theacrine and other purine alkaloids biosynthesis<sup>28</sup>. Our results show that the gene expression levels of SAMS were up-regulated in two groups, which the total amount of purine alkaloids was higher than that of control variety. TCS is a class of *N*-methyltransferase that catalyzes the methylation of *N*-3 (theobromine synthase) and *N*-1 (caffeine synthase) in tea plant. Previous studies point that *TCS1* plays a crucial role for methylation of xanthosine at 1-*N* position in caffeine biosynthesis, while *TCS1* alleles with low transcription level or encoded proteins with no TCS activity<sup>29</sup>. TEA015791, encoded *TCS1*, was down-regulated in ND2 and ND3, it may be the reason that why the lower content of caffeine in two Kucha samples. But no research indicated that TCS has the ability to catalyzed methylation at 9-*N* position. It still needs the further research that whether TCS involved in the final step of theacrine biosynthesis. TEA024443, encoded NMTs, was highly expression in high-theacrine sample which was positively correlated with the content. *NMT* may play a significant role in theacrine synthesis which dominant the pathway of “1,3,7-trimethyl xanthine (caffeine) → 1,3,7-trimethyluric acid → 1,3,7,9-tetramethyluric acid (theacrine)”, we speculated that *N*-methyltransferase encoded by this gene responsible for catalyze the methylation at 9-*N* position in theacrine



synthesis. Meanwhile, the expressions of related gene in purine alkaloid degradation pathway was probably correlated with the decreased of purine alkaloid in Kucha samples. TEA007899 and TEA023610, encoded allantoinase (ALN), and TEA020308, encoded urease (URE), were all highly expression in ND2 and ND3 with low-caffeine Kucha plant, which involved the final degradation step of “Allantoin $\rightarrow$ Allantoic acid $\rightarrow$  $\rightarrow$ NH<sub>3</sub>+CO<sub>2</sub>”.

Caffeine is seemed as a precursor of theacrine, its biosynthesis pathway is very clear. And the key genes involved in caffeine metabolism have been identified, which can provide reference for the further study of theacrine<sup>29–31</sup>. However, until now few studies have been reported on molecular mechanism of theacrine biosynthesis in Kucha cultivars. In this study, we proposed an NMT, encoded by TEA024443, was likely the key enzyme involving in the final synthesis process of theacrine. Next, the gene of TEA024443 should be focused on verification and testing its function in theacrine biosynthesis. Further, we will construct the tea genetic population with Kucha cultivars as parents, key genes involved in theacrine metabolism were identified and analyzed by density genetic map and QTL mapping. It will make it possible to breeding new tea cultivar which is rich in theacrine yet contains low caffeine.

To sum up, our study provided a new insight to illuminating theacrine molecular metabolism in Kucha plants, and it may help to the breeding improvement of Kucha plant in subsequent studies.

## Methods

**Plant materials.** Three tea cultivars were used as plant materials in this study, including high-theacrine cultivar ‘Niedu Kucha 3’ (ND3), medium-theacrine ‘Niedu Kucha 2’ (ND2) and low-theacrine ‘Shangyou Zhongye’ (SY) as a control. The former two are Kucha accessions from the same local population of tea landrace named ‘Niedu Kucha’, the latter one is local cultivated tea clone. All of them were collected from the same region located in the southwest part of Jiangxi province. And they are currently preserved in National Germplasm Hangzhou Tea Repository, in the Tea Research Institute of the Chinese Academy of Agricultural Sciences at Hangzhou city, Zhejiang province in China. Leaves of two Kucha plants are tasted very bitter, while that of the controlled cultivar is not bitter. Fresh healthy leaves were picked on the two terminal leaf position from ND2, ND3 and SY respectively, quickly frozen in liquid nitrogen firstly, and then stored at  $-80^{\circ}\text{C}$  for RNA extraction. And the young shoots with “two leaves and a bud” collected from three cultivars were fixed with hot air at  $120^{\circ}\text{C}$  for 5 min and then dried to constant weight at  $80^{\circ}\text{C}$ . These dried samples were used to detect content of purine alkaloid such as theacrine, caffeine and theobromine using HPLC method. All experiments had three biological replicates.

**Identification of purine alkaloids.** Purine alkaloids such as theacrine, caffeine and theobromine were detected by HPLC methods described in a previous reported method<sup>32</sup>. A Waters High Performance Liquid Chromatography (HPLC) instrument was used to detect the compounds in this experiment. A Waters C<sub>18</sub> column (5  $\mu\text{m}$ , 4.6 cm  $\times$  250 mm) was performed at a flow rate of 1.0 mL/min with the column temperature of  $35^{\circ}\text{C}$ . 1% formic acid was used as mobile phase A, 100% acetonitrile as mobile phase B, 10  $\mu\text{L}$  sample liquid was injected and analyzed at 280 nm, and the elution time was 45 min.

**RNA extraction and library construction.** EASY-spin Plus Complex Plant RNA kit (Aidlab Biotechnologies Co., Beijing, China) was used to extract total RNA. The integrity and quantity of total RNA were estimated by NanoDrop ultraviolet spectrophotometer (Thermo, Waltham, MA) and agarose gel electrophoresis. cDNA libraries established by the standard mRNA-Seq Library Prep Kit, and the library quality was assessed by using the Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA). The cDNA libraries were sequenced by the Illumina HiSeq platform, performed by Personal Biotechnology Co., Ltd. (Beijing, China).

**Gene expression analysis.** High-quality reads were evaluated and acquired from the raw data, which need to filter out the low-quality sequence, adaptor sequences, duplicated sequences and ambiguous (with the ratio of  $N > 5\%$ ). The clean reads from all the samples were mapped to *C. sinensis* ‘Shuchazao’ genome by HISAT software with default parameters and counted by feature counts<sup>33,34</sup>.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation were downloaded from <http://tpia.teaplant.org/>. The DEGs (differentially expressed genes) among tea samples (three biological replicates per group) were identified by DESeq. 2. Genes with a FDR (false discovery rate)  $\leq 0.05$  and an absolute fold change value  $\geq 1.0$  were identified as DEGs<sup>35,36</sup>.

**Validation of quantitative real-time PCR.** To confirm the accuracy of RNA-Seq and DEG analyses, 12 DEGs associated with purine alkaloid biosynthesis pathway were selected for qRT-PCR. The primers of specific genes for qRT-PCR were designed by Primer Premier 5, and the primer pair sequences were listed in Table S1. First-strand cDNA (10-fold dilution) synthesis from total RNA was operated by using FastKing RT kit (with gDNase) (TianGen Biotech Co, Beijing). qRT-PCR was performed in ABI7500 (Applied Biosystems Inc, US), and the GAPDH (glyceraldehyde 3-phosphate dehydrogenase) gene were utilized as the internal reference. The total volume of the qRT-PCR reaction is 20  $\mu\text{L}$  which contained 7.4  $\mu\text{L}$  distilled H<sub>2</sub>O, 0.6  $\mu\text{L}$  primer each pair, 10  $\mu\text{L}$  2x FastFire qPCR PreMix, 0.4  $\mu\text{L}$  50x Rox Reference Dye $\Delta$  and 1  $\mu\text{L}$  diluted cDNA. The cycling profile was  $95^{\circ}\text{C}$ , 60 s; 40 cycles of  $95^{\circ}\text{C}$ , 5 s, and  $60^{\circ}\text{C}$ , 32 s in 96-well optical reaction plates. The relative gene expression was calculated though the  $2^{-\Delta\Delta\text{Ct}}$  method<sup>37</sup>. Each sample was examined in three technical replicates.

**Statistical analysis.** Three biological replicates of sequencing data and purine alkaloid components were used for calculating the mean value and the standard deviation (SD) based the Duncan’s multiple range tests. SPSS 18.0 was used to analyze the correlations of gene expression and purine alkaloid concentrations via Pearson correlation. The volcano figure was produced with Prism 7.0 software. The double coordinate figures and heat maps were draw with R package.

Received: 25 February 2020; Accepted: 19 March 2020;

Published online: 14 April 2020

## References

- Chen, L., Yang, Y. J. & Yu, F. L. Tea germplasm research in China: recent progresses and prospects. *Journal of Plant Genetic Resources* **5**, 389–392 (2004).
- Shi, X. G., Zheng, X. Q., Song, X. H., Wang, Y. Y. & Ye, C. X. A new combination of kucha. *Acta Scientiarum Naturalium Universitatis Sunyatseni* **47**, 129–131 (2008).
- Jin, J. Q. *et al.* Identification on purine alkaloids of representative tea germplasms in China. *Journal of Plant Genetic Resources* **15**, 279–285 (2014).
- Ye, C. X. *et al.* New discovery of pattern purine alkaloids in wild tea trees. *Acta Scientiarum Naturalium Universitatis Sunyatseni* **42**, 62–65 (2003).
- Jing, J. *et al.* Research progress in theacrine. *China Pharmacist* **19**, 344–346 (2016).
- Li, K. K. *et al.* Antioxidative activities and the chemical constituents of two Chinese teas, *Camellia kucha* and *C. ptilophylla*. *International Journal of Food Science and Technology* **47**, 1063–1071 (2012).
- Yang, X. R., Ye, C. X., Xu, J. K. & Jiang, Y. M. Simultaneous analysis of purine alkaloids and catechins in *Camellia sinensis*, *Camellia ptilophylla* and *Camellia assamica* var. *kucha* by HPLC. *Food Chemistry* **100**, 1132–1136 (2007).
- Peng, L., Song, X. R., Shi, X. G. & Li, J. X. An improved HPLC method for simultaneous determination of phenolic compounds, purine alkaloids and theanine in *Camellia* species. *Journal of Food Composition and Analysis* **21**, 559–563 (2008).
- Wang, Y. Y. *et al.* Theacrine, a purine alkaloid with anti-inflammatory and analgesic activities. *Fitoterapia* **81**, 627–631 (2010).
- Xie, G., He, R. R., Feng, X. & Yan, T. The hypoglycemic effects of *Camellia assamica* var. *kucha* extract. *Journal of the Agricultural Chemical Society of Japan* **74**, 3 (2010).
- Qiao, H. Y. *et al.* Theacrine: A purine alkaloid from *Camellia assamica* var. *kucha* with a hypnotic property via the adenosine system. *Neuroscience Letters* **659**, 48–53 (2017).
- Kato, M. *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. *Physiologia Plant* **98**, 629–636 (1996).
- Zhou, C. Y., Jin, J. Q., Yao, M. Z. & Chen, L. Progress on purine alkaloids metabolism in tea and other plants. *Journal of Tea Science* **31**, 87–94 (2011).
- Zheng, X. Q., Ye, C. X., Kato, M., Crozier, A. & Ashihara, H. Theacrine (1,3,7,9-tetramethyluric acid) synthesis in leaves of a Chinese tea, *kucha* (*Camellia assamica* var. *kucha*). *Phytochemistry* **60**, 129–134 (2002).
- Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Gene* **25**, 25–29 (2000).
- Minoru, K., Susumu, G., Yoko, S., Miho, F. & Mao, T. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research* **40**, 109–114 (2012).
- Zhu, B. Y. *et al.* 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* **67**, 3400–3411 (2019).
- Zhang, H. D. A revision on the tea resources plants. *Acta Scientiarum Naturalium Universitatis Sunyatseni* **1**, 3–14 (1984).
- Du, Q. Z., Li, M. J. & Liu, W. H. Chemical and numerical taxonomies of plants *Thea* section plants. *Journal of Tea Science* **10**, 1–12 (1990).
- Min, T. L. A revision of *camellia* sect. *Thea*. *Acta Botanica Yunnanica* **14**, 115–132 (1992).
- Chen, L., Yu, F. L. & Tong, Q. Q. Discussions on phylogenetic classification and evolution of sect. *Thea*. *Journal of Tea Science* **20**, 89–94 (2000).
- Wang, X. C., Yao, M. Z., Ma, C. L. & Chen, L. Analysis and evaluation of biochemical component in bitter tea plant germplasm. *Chinese Agricultural Science Bulletin* **24**, 65–69 (2008).
- Li, D., Li, D. S., Yang, C., Wang, Q. & Luo, J. W. Genetic diversity and relationship of tea germplasm resources *Camellia sinensis* var. *assamica* cv. *Jianghua* revealed by ISSR markers. *Journal of Tea Science* **32**, 135–142 (2012).
- Li, D. & Luo, J. W. Relationship of *Tea assamica* cv. *Jianghua* germplasm resource and its status in the process of tea evolution. *Hunan Agricultural Sciences* **1**, 10–12 (2011).
- Yang, C. *et al.* Analysis of seasonal variation of basic biochemical composition of *Camellia sinensis* var. *assamica* cv. *Jianghua*. *Southwest China Journal of Agricultural Sciences* **26**, 1402–1405 (2013).
- Ashihara, H., Yokota, T. & Crozier, A. Biosynthesis and catabolism of purine alkaloids. *Advances in Botanical Research* **68**, 111–138 (2013).
- Mazzafera, P. Catabolism of caffeine in plants and microorganisms. *Frontiers in Bioscience* **9**, 1348–1359 (2004).
- Koshihishia, C., Kato, A., Yama, S., Crozier, A. & Ashihara, H. A new caffeine biosynthetic pathway in tea leaves: utilisation of adenosine released from the S-adenosyl-L-methionine cycle. *FEBS Letters* **499**, 50–54 (2001).
- Jin, J. Q., Yao, M. Z., Ma, C. L., Ma, J. Q. & 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* **100**, 18–26 (2016).
- Jin, J. Q., Yao, M. Z., Ma, C. L., Ma, J. Q. & Chen, L. Association mapping of caffeine content with TCS1 in tea plant and its related species. *Plant Physiology and Biochemistry* **105**, 251–259 (2016).
- Li, Y., Qgita, S., Keya, C. A. & Ashihara, H. Expression of caffeine biosynthesis genes in tea (*Camellia sinensis*). *Zeitschrift für Naturforschung C* **63**, 267–270 (2008).
- Jin, J. Q. *et al.* Hongyacha, a naturally caffeine-free tea plant from Fujian, China. *Journal of Agriculture and Food Chemistry* **66**, 11311–11319 (2018).
- Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12**, 357–360 (2015).
- Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq. 2. *Genome Biology* **15**, 550 (2014).
- Wei, C. L. *et al.* Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality. *Proc Natl Acad Sci USA* **115**, 201719622 (2018).
- Dewey, C. N. & Bo, L. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323–323 (2011).
- Wu, Z. J., Tian, T., Jiang, Q. & Li, X. H. Selection of suitable reference genes for qRT-PCR normalization during leaf development and hormonal stimuli in tea plant (*Camellia sinensis*). *Scientific Reports* **6**, 19748 (2016).

## Acknowledgements

The research was supported by the Earmarked Fund for China Agriculture Research System (CARS-019), the Chinese Academy of Agricultural Sciences through the Agricultural Science and Technology Innovation Program (CAAS-ASTIP-2017-TRICAAS) to L.C., the Fundamental Research Fund for Tea Research Institute of Chinese Academy of Agricultural Sciences (1610212016002) to M.Z.Y.

### Author contributions

M.Z.Y. and L.C. conceived and designed the research. J.Q.J. provided valuable research design insights. S.L.W. performed the experiments. J.D.C. analyzed the data. J.Q.M. contributed materials, and transcriptome analysis. S.L.W. wrote the paper. M.Z.Y., L.C. and J.D.C. revised the paper.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-62859-2>.

**Correspondence** and requests for materials should be addressed to L.C. or M.Y.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020