

RESEARCH ARTICLE

Open Access



Transcriptomic and phytochemical analysis of the biosynthesis of characteristic constituents in tea (*Camellia sinensis*) compared with oil tea (*Camellia oleifera*)

Yuling Tai^{1†}, Chaoling Wei^{1†}, Hua Yang^{1†}, Liang Zhang¹, Qi Chen¹, Weiwei Deng¹, Shu Wei¹, Jing Zhang¹, Congbing Fang¹, Chitang Ho² and Xiaochun Wan^{1*}

Abstract

Background: Tea plants (*Camellia sinensis*) are used to produce one of the most important beverages worldwide. The nutritional value and healthful properties of tea are closely related to the large amounts of three major characteristic constituents including polyphenols (mainly catechins), theanine and caffeine. Although oil tea (*Camellia oleifera*) belongs to the genus *Camellia*, this plant lacks these three characteristic constituents. Comparative analysis of tea and oil tea via RNA-Seq would help uncover the genetic components underlying the biosynthesis of characteristic metabolites in tea.

Results: We found that 3,787 and 3,359 bud genes, as well as 4,042 and 3,302 leaf genes, were up-regulated in tea and oil tea, respectively. High-performance liquid chromatography (HPLC) analysis revealed high levels of all types of catechins, theanine and caffeine in tea compared to those in oil tea. Activation of the genes involved in the biosynthesis of these characteristic compounds was detected by RNA-Seq analysis. In particular, genes encoding enzymes involved in flavonoid, theanine and caffeine pathways exhibited considerably different expression levels in tea compared to oil tea, which were also confirmed by quantitative RT-PCR (qRT-PCR).

Conclusion: We assembled 81,826 and 78,863 unigenes for tea and oil tea, respectively, based on their differences at the transcriptomic level. A potential connection was observed between gene expression and content variation for catechins, theanine and caffeine in tea and oil tea. The results demonstrated that the metabolism was activated during the accumulation of characteristic metabolites in tea, which were present at low levels in oil tea. From the molecular biological perspective, our comparison of the transcriptomes and related metabolites revealed differential regulatory mechanisms underlying secondary metabolic pathways in tea versus oil tea.

Keywords: Comparative transcriptome, *Camellia sinensis*, *Camellia oleifera*, Chemical constituents, Metabolic pathways

Background

Tea is produced from the plant *Camellia sinensis* (L.) O. Kuntze in the family *Theaceae*. Tea is one of the most popular beverages worldwide, and tea leaves represent an important source of many biologically active metabolites such as flavonoids, theanine and caffeine [1, 2]. Flavonoids mainly comprise flavan-3-ols (catechins), epicatechin

(EC), gallic catechin (GC), epigallocatechin (EGC), catechin (C) and their respective gallate esters, such as epigallocatechin gallate (EGCG) and epicatechin gallate (ECG) [3]. Tea leaves, which contain various secondary metabolites, are usually used as the raw material for tea production. However, the molecular mechanisms that regulate the biosynthesis of catechins, theanine and caffeine in tea remain elusive.

Great effort has focused on elucidating the molecular mechanisms underlying plant growth, development [4, 5] and secondary metabolite production [6] in tea. Most of

* Correspondence: xcwan@ahau.edu.cn

†Equal contributors

¹State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei 230036, People's Republic of China

Full list of author information is available at the end of the article

these studies have focused on characterizing genes related to secondary metabolism, most of which were revealed through EST sequencing [7] and analysis of the transcriptomes from various tissues of tea plants [8, 9] or under different stress conditions [10, 11]. More recently, Shi *et al.* discovered novel candidate genes involved in pathways in tea by analyzing transcriptome data [12]. Liu *et al.* reported the discovery of a novel enzyme involved in galloylated catechin biosynthesis in tea plants [13]. However, the lack of genomic information has become an obstacle to exploring the molecular mechanisms underlying secondary metabolite biosynthesis in tea. Transcriptome sequencing represents an efficient approach to obtaining functional genomic information.

RNA-Seq is a rapid technique for genome-wide gene expression analysis that is widely used to determine gene structures and expression profiles in model organisms. *De novo* assembly of RNA-Seq data makes it possible to conduct gene analysis in the absence of reference genomes [14–16]. Comparative transcriptomic studies have been performed to identify differential gene expression in several organisms [17–20].

Another widely known member of *Theaceae* is oil tea, *Camellia oleifera* Abel, a tree serving as an important source of edible oil that is grown specifically in China. Oil tea was genetically closely to tea, and they both belonged to genus *Camellia*. Here, we performed RNA-Seq on buds and second leaves of tea and oil tea to characterize differences in gene expression between these two plants. This comparative transcriptomic analysis provides important insights into the molecular mechanisms underlying secondary metabolite biosynthesis in tea, as well as the phytochemical characteristics of its main metabolites.

Results

Analysis of the contents of catechins, theanine and caffeine

HPLC analyses were conducted to determine the contents of catechins, theanine and caffeine, and related intermediates in buds and five leaves of tea and oil tea (Fig. 1). All standard compounds showed good linearity ($R^2 > 0.9991$) in a relatively wide concentration range. Compared to oil tea, most of these metabolites were present at higher concentrations in tea (Fig. 1b). The average contents of three characteristic components (total catechins, theanine and caffeine) in tea leaves were 1.5- to 173-fold higher than those in oil tea leaves. In particular, tea contained over a 180 mg/g of total catechins in its leaves and buds. The only exception is that the anthocyanin content in oil tea leaves was 32-fold higher than that in tea leaves. These results confirm that tea is rich in catechins, theanine and caffeine (Table 1).

Moreover, the contents of these characteristic constituents varied during the period from the appearance of

buds to the appearance of the five leaves. The levels of GC, EGC and EC increased from the bud to the second or third leaves in tea, whereas a general decline in caffeine, total catechins, ECG and EGCG levels was observed in tea leaves.

A steady decrease in theanine levels was observed from the first leaf to the fifth leaf in tea, and the levels of this compound were almost seven-fold greater in buds than in leaves. A similar variation was detected in oil tea, but the absolute contents were much lower. Due to the variation in the contents of most compounds (EC, EGC, GC and ECG) in the three initial leaves, we selected the second leaves and buds of tea and oil tea for RNA-Seq.

De novo assembly and comparative analyses of RNA-Seq data

We utilized Illumina RNA-Seq technology to sequence the buds and second leaves of tea and oil tea. After removing adaptor sequences, duplication sequences, ambiguous reads and low-quality reads, a total of 23.4 Gb of clean reads was generated, with an average of 5.85 Gb clean reads per sample (Table 2).

The final assembly of tea had 81,826 unigenes with an N50 length of 1,265 bp (Table 3). Functional annotation revealed 53,786, 49,174, 34,636, 31,024, 18,748 and 40,838 unigenes with alignments to the NR (Non-redundant protein database), NT (Non-redundant nucleotide database), Swiss-Prot (Annotated protein sequence database), KEGG (Kyoto encyclopedia of genes and genomes), COG (Clusters of orthologous groups of protein) and GO (Gene ontology) databases, respectively. The final assembly of oil tea consisted of 78,863 unigenes with an N50 length of 1,254 bp. Of these, 54,115, 49,009, 34,682, 30,990, 19,126 and 41,325 unigenes were annotated by alignment against the NR, NT, Swiss-Prot, KEGG, COG and GO databases, respectively (Additional file 1). Sequence comparisons revealed that 17,459 genes are shared by both tea and oil tea, 9,725 of which were mapped to KEGG pathways (<http://www.genome.jp/kegg/>) [21]. High amino acid sequence identity was found in the homologous genes between tea and oil tea, as 64 % of the genes shared over 70 % identity. We also detected 64,826 specific transcripts in tea and 61,863 in oil tea.

Analysis of the differentially expressed genes (DEGs)

The DEGs were identified by comparing FPKM (Fragment Per Kilobase of exon model per Million mapped reads) values [22] between different libraries under the thresholds of log₂ (Fold-change) over 1 and FDR less than 0.001 (Fig. 2 and Additional file 2). The results indicated that both tea and oil tea had more genes with higher transcription levels in the second leaves than in buds. Compared with oil tea, tea contained more DEGs

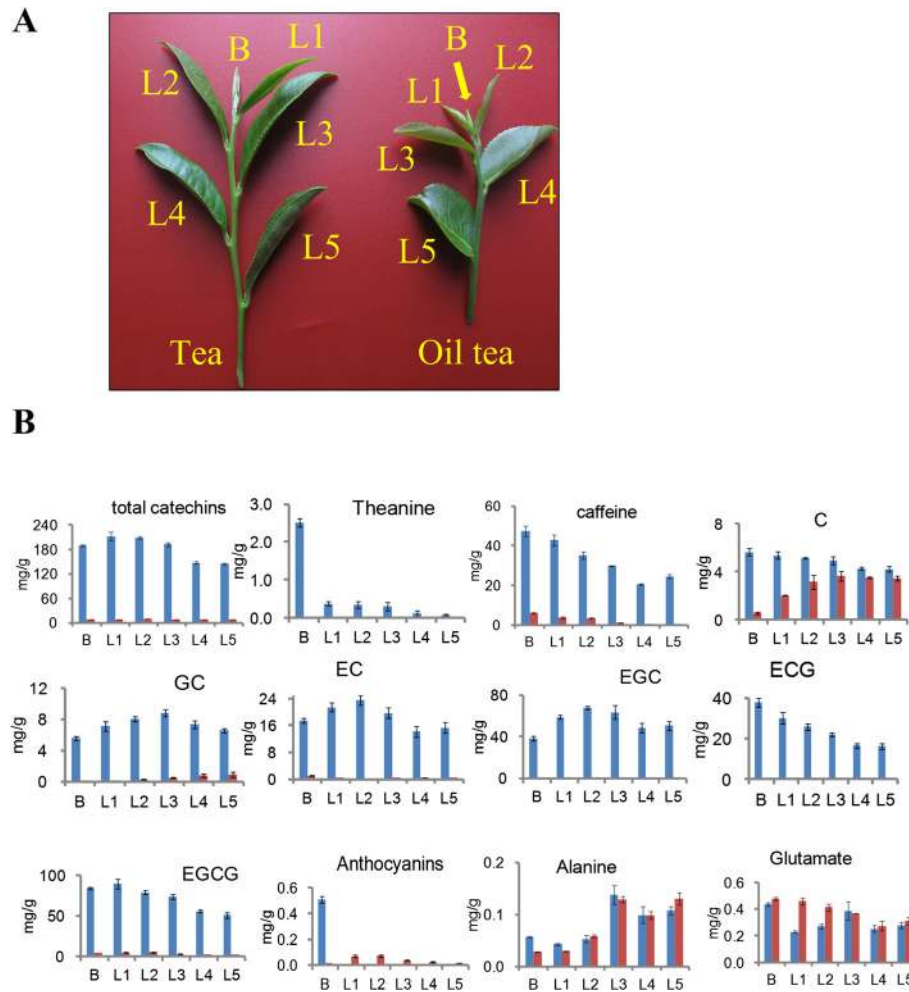


Fig. 1 Determined contents of compounds involved in three main secondary metabolic pathways. **a** Samples examined. Buds and five initial leaves were collected from tea and oil tea. Bud, L1, L2, L3, L4 and L5 indicate the bud, first, second, third, fourth and fifth leaf, respectively. **b** Variation in the contents of compounds biosynthesized by catechin-, theanine- and caffeine-related metabolic pathways

(3,787 in buds and 4,042 in leaves) with increased expression in both buds and leaves than oil tea (3,359 in buds and 3,302 in leaves). Next, we analyzed the DEGs using KEGG pathway analysis, which assigned 4,226 DEGs derived from tea buds versus oil tea buds (TBvsOTB), 4,174 from tea buds versus tea leaves (TBvsTL), 4,334 from tea leaves versus oil tea leaves (TLvsOTL) and 3,418 from oil tea buds versus leaves (OTBvsOTL). High proportions of these DEGs are involved in secondary metabolite pathways, including 483 DEGs (11.43 %) from TBvsOTB, 503 (11.61 %) from TLvsOTL, 594 (14.23 %) from TBvsTL and 482 (14.1 %) from OTBvsOTL, respectively. The estimated rich factors (number of DEGs mapped to a certain pathway/total number of genes mapped to this pathway) of secondary metabolism were 0.4–0.7 in TBvsOTB and TLvsOTL (Fig. 3a and b), whereas they were 0.1–0.3 in TBvsTL and OTBvsOTL (Fig. 3c and d). The DEGs

identified through comparisons between tea and oil tea were clustered in the pathway secondary metabolism, suggesting that there are different secondary metabolism pathways in these two species. A lower rich factor between two stages for either of two species implies that steady metabolism occurs during this period (Additional file 3).

Based on alignments against the Swiss-Prot, COG and KEGG databases with an e-value cutoff of less than 1×10^{-30} , 117, 51 and 18 tea genes and 110, 52 and 20 oil tea genes were found to be involved in the biosynthesis of catechins, theanine and caffeine, respectively (Additional file 4). We detected over 200 homologous genes in tea and oil tea encoding enzymes potentially involved in catalyzing these reactions. Tea and oil tea contain a similar number of genes encoding most enzymes in the assembled gene models, but their transcription levels are considerably different (Table 4).

Table 1 Variation in the contents of major compounds in tea and oil tea

Compounds	Content in buds (mg/g)		Tea bud/Oil tea bud	Average content in leaves (mg/g) ^a		Tea leaves/Oil tea leaves
	Tea	Oil tea		Tea	Oil tea	
Total catechins	189.00	7.63	24.8	181.00	8.76	20.7
Theanine	2.50	0.01	252.5	0.25	0.00	86.8
Caffeine	47.50	5.98	7.9	30.50	1.81	16.8
C	5.65	0.60	9.4	4.79	3.18	1.5
GC	5.54	0.00	76,654.6	7.57	0.51	14.8
EC	17.40	1.12	1.1	18.90	0.43	43.8
EGC	38.20	0.38	12.5	57.70	0.33	173.3
ECG	0.52	13.20	2.9	22.10	0.52	42.9
EGCG	4.98	4.98	16.9	69.90	3.75	18.7
Anthocyanins	0.51	0.02	32.5	0.00	0.04	0.0
Alanine	0.06	0.03	2.04	0.08	0.09	0.96
Glutamate	0.44	0.44	1.0	0.29	0.37	0.8

^a Average contents in leaves were calculated using values from the five initial leaves

Identification of DGEs involving in characteristic metabolic pathways in tea

We used qRT-PCR to confirm the differential expression levels of 34 DEGs involved in the biosynthesis of catechins, theanine and caffeine and quantified their maximum transcription levels in tea and oil tea (Fig. 4 and Additional file 5). Of these genes, the data from 25 (74 % of 34) matched the RNA-Seq data. As determined from the published flavonoid pathways [23], catechin biosynthesis occurs via successive enzymatic reactions (Fig. 4a). Interestingly, PAL (phenylalanine ammonia-lyase) and CHI (chalcone isomerase) genes, which are employed in the upstream phenylpropanoid pathway, were more highly expressed in oil tea than in tea. However, in the downstream biosynthetic pathway of catechins, the F3H (flavanone 3-hydroxylase), DFR (dihydroflavonol 4-reductase) and ANR (anthocyanidin reductase) genes were more highly expressed in tea. Notably, the ANR gene encodes an enzyme that catalyzes the transfer of anthocyanidins to 2,3-cis-flavan-3-ol, which is an intermediate in the final step of esterified catechin synthesis. Both RNA-Seq and qRT-PCR analyses revealed considerable activation of the ANR gene in tea but not in oil tea, which is consistent with the data from HPLC analyses of EC, EGC, C and GC contents. The DFR, LAR and ANR genes in tea are responsible for the biosynthesis of nongalloylated catechins [24]. The differential expression levels of

F3H, DFR and ANR genes might be responsible for the differences detected in the levels catechin components between tea and oil tea.

Tea buds and leaves contain theanine at levels as much as 252-fold and 86-fold those of oil tea (Fig. 1), respectively. However, we did not identify genes encoding the enzyme responsible for the final reaction in theanine biosynthesis. The qRT-PCR analysis revealed that the GS (glutamine synthetase) and GDH (glutamate dehydrogenase) genes were more highly expressed in tea than in oil tea (Fig. 4b). Previous studies suggest that theanine is synthesized from glutamic acid and ethylamine by TS (theanine synthetase), which is highly homologous to glutamine GS [25]. Phytochemical analysis revealed a much higher content of theanine in tea buds and leaves than in oil tea, suggesting a potential connection between the activation of GS genes and high theanine levels in tea. In our transcriptomic data, five GS unigenes were found in tea and seven in oil tea. Whether they are functional copies of TS genes remains to be confirmed by further analysis of enzymatic reactions.

There are three key enzymes in the caffeine biosynthesis pathway: TCS (tea caffeine synthase), IMPDH (inosine-5'-monophosphate dehydrogenase) and SAMS (S-adenosyl-methionine synthetase) [26]. We detected homologous genes that are involved in four steps of the caffeine pathway. TCS catalyzes the final step in caffeine biosynthesis.

Table 2 Statistics from the generated RNA-Seq reads

	Sample	Total raw reads	Total clean reads	Total clean nucleotides (nt)	Q20 percentage	N percentage	GC percentage
Tea	buds	78,077,028	66,059,720	5,945,374,800	97.50 %	0.01 %	46.71 %
	leaves	71,399,954	65,258,822	5,873,293,980	97.36 %	0.01 %	46.34 %
Oil tea	buds	86,743,714	65,259,264	5,873,333,760	97.44 %	0.01 %	46.57 %
	leaves	73,366,916	63,928,844	5,753,595,960	97.50 %	0.01 %	46.76 %

Table 3 Summary of assemblies of RNA-Seq data

Species	Sample	Total number	Total length(nt)	Mean length(nt)	N50 length(nt)	Total consensus sequences	Distinct clusters	Distinct singletons	
Tea	Contig	buds	157,832	50,217,182	318	508	-	-	
		2nd leaves	151,557	47,978,363	317	507	-	-	
Oil tea	Unigene	buds	89,155	55,909,930	627	1103	89,155	30,384	
		2nd leaves	83,415	83,415	53,056,344	636	1099	83,415	28,571
Tea	All unigenes	buds	81,826	64,132,659	784	1265	81,826	33,079	48,747
		2nd leaves	81,826	64,132,659	784	1265	81,826	33,079	48,747
Oil tea	Contig	buds	153,407	47,519,556	310	491	-	-	
		2nd leaves	148,760	47,208,814	317	508	-	-	
Oil tea	Unigene	buds	84,244	52,523,964	623	1073	84,244	30,379	
		2nd leaves	83,679	53,084,367	634	1097	83,679	29,963	
Oil tea	All unigenes	buds	78,863	62,035,893	787	1254	78,863	34,041	44,822
		2nd leaves	78,863	62,035,893	787	1254	78,863	34,041	44,822

The TCS gene was much more highly expressed in tea buds and leaves (by over 45-fold) than in oil tea, although the genes responsible for the upstream reactions had higher transcription levels in oil tea, which was confirmed by qRT-PCR (Fig. 4c).

Taken together, our investigation of gene expression in tea revealed the activation of related metabolic pathways compared to oil tea. Most genes exhibited slightly higher expression levels in buds than in leaves (Table 4). These findings are potentially related to the differences in metabolic components revealed by HPLC.

Discussion

In this study, we observed differences in the contents and gene expression patterns of the characteristic compounds in tea compared to oil tea. We found that tea contains more beneficial nutrients, such as catechins, theanine and caffeine, in its buds and leaves because the pathways related to these metabolites were considerably more active in tea than in oil tea. Theanine is a unique

non-protein amino acid that was first discovered in tea. There are trace amounts of this compound in two other *Camellia* species (*C. japonica* and *C. sasanqua*) and in one species of mushroom (*Xerocomus badius*) [27].

Of the phenolic compounds, high flavonoid levels are present in oil tea, as revealed by HPLC (140.06 mg/g dry material) [28]. Flavonoids are a class of important secondary metabolites including flavanones, flavones, dihydroflavonols, flavonols and flavan-3-ols (catechins). These compounds are important for tea quality and are beneficial for human health (especially catechins) [29]. Catechins, theanine and caffeine are the main characteristic compounds in tea, and the results of our analysis of these compounds are in accordance with recent reports [30, 31]. Oil tea is genetically closely to tea, but no theanine and caffeine were reported except flavonoids in oil tea leaves in previous study [32, 33]. We chose tea and oil tea buds and leaves of plants from the same environment for analysis to reveal the mechanism behind the high levels catechins, theanine and caffeine in tea. Our results indicated

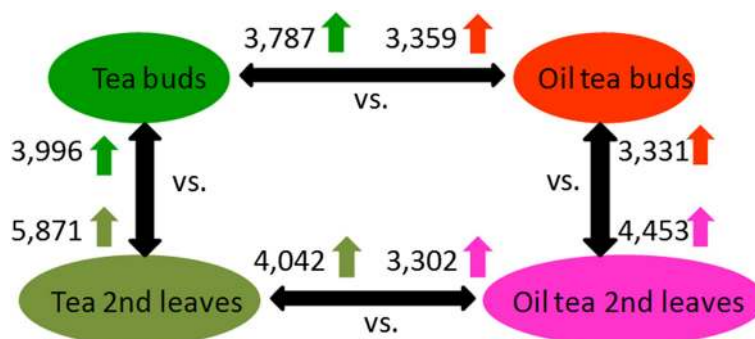
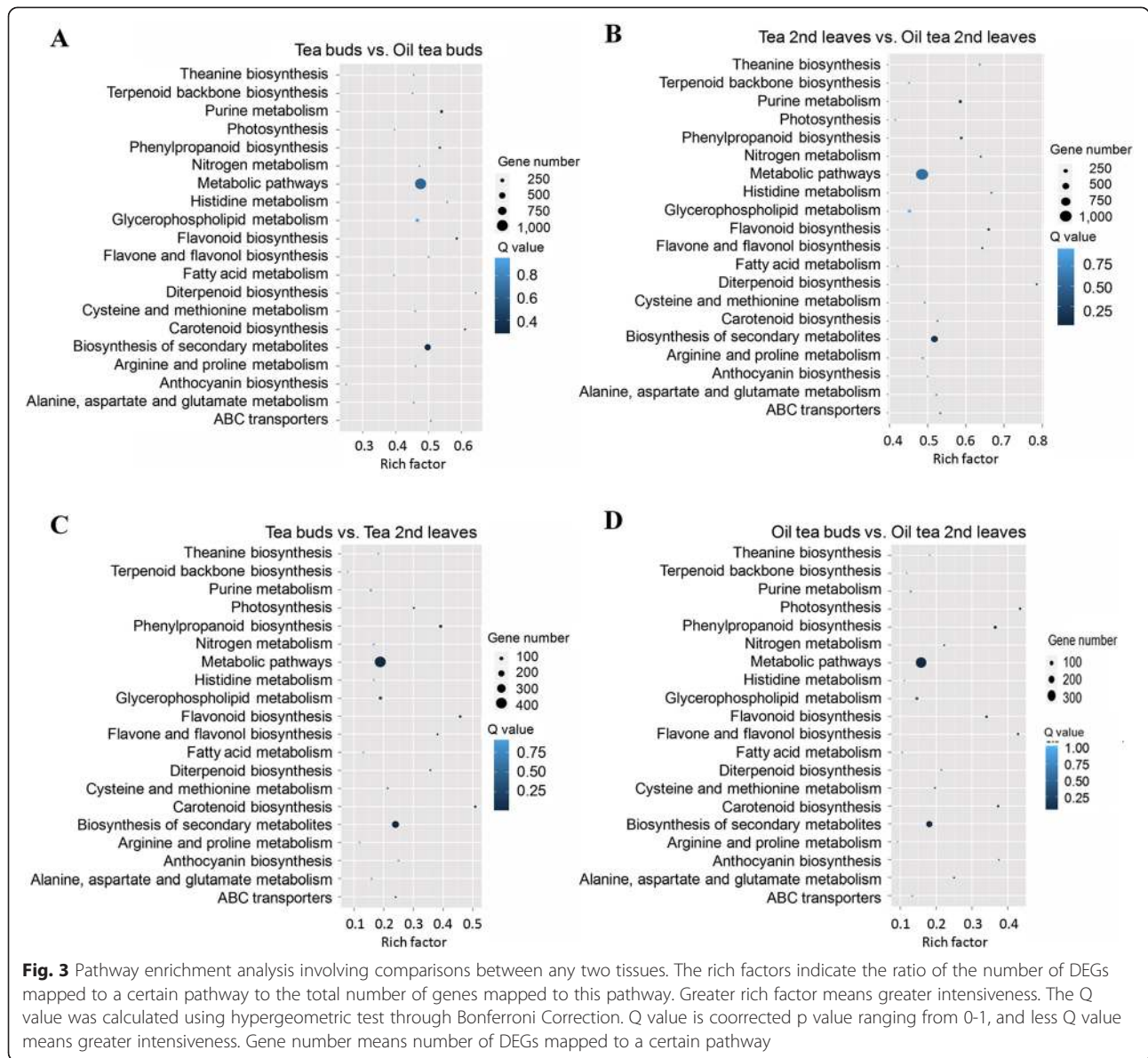


Fig. 2 Number of identified DEGs identified by comparing gene expression levels between any two tissues. The numbers of significantly up-regulated genes (log2(Fold-change) > 1; FDR < 0.001) between two tissues (indicated by double-headed arrows). Arrows indicate up-regulated genes associated with each number



that the catechins, theanine and caffeine in tea were also present in oil tea, but in much lower amounts. We detected increased expression of some key genes in these three metabolic pathways in tea compared to oil tea, which might lead to the differences in their contents.

Our results indicated that the genes encoding F3H, DFR and ANR in the flavonoid pathway were more highly expressed in tea than in oil tea. On the contrary, the expression levels of PAL and CHI genes were lower in tea than in oil tea. These observations were consistent with previous results [34]. High PAL activity was associated with the accumulation of flavonoids and other phenolic compounds [35, 36], and DFR, ANR and LAR played an important role in the formation of catechins [3]. Xiong *et al.* found that stable expression

of F3H insured the formation of dihydrokaempferol, the precursor of individual catechins [37]. In the current study, we did not observe a difference in the expression levels of the C4H gene between tea and oil tea.

Our analysis of the DEGs related to flavonoid, theanine and caffeine metabolism in tea and oil tea suggests that these two species share common pathways, but the expression levels of some key genes in these pathways might result in differential biosynthesis of catechins, theanine and caffeine.

Since tea is self-incompatible and recalcitrant to genetic manipulation, little genetic or genomic information is currently available for this species. Therefore, instead of providing a comprehensive in-depth investigation of the tea transcriptome, our experiment was designed to generate a quick view of the landscape. Moreover, since there were

Table 4 Transcription levels of genes involved in the biosynthesis of catechins, theanine and caffeine

Pathway	Enzyme	Gene #	Tea buds		Tea leaves		Gene #		Oil tea buds		Oil tea leaves		Max.Log10(Tea FPKM)-Max. Log10 (Oil Tea FPKM)	
			Description	Abbr.	Max. Log10 (FPKM)	Max. Log10 (FPKM)	Gene #	Gene #	Max. Log10 (FPKM)	Max. Log10 (FPKM)	Bud	2nd leaf		
Catechines	phenylalanine ammonia-lyase	PAL	21	6.12	5.71	20	5.22	5.01	0.90	0.70				
	cinnamic acid 4-hydroxylase	C4H	-	-	-	-	-	-	-	-	-	-	-	-
	4-coumarate-CoA ligase	4CL	16	5.60	5.61	17	5.60	5.82	0.00	-0.21				
	chalcone synthase	CHS	37	7.27	6.90	36	7.59	7.22	-0.32	-0.32				
	chalcone isomerase	CHI	1	5.34	4.83	1	5.68	5.27	-0.34	-0.44				
	flavanone 3-hydroxylase	F3H	5	7.18	6.61	6	5.70	5.66	1.48	0.95				
	flavonoid 3',5'-hydroxylase	F3'5'H	5	6.10	5.75	4	5.40	4.90	0.70	0.85				
	flavonoid 3'-hydroxylase	F3'H	-	-	-	-	-	-	-	-	-	-	-	-
	leucoanthocyanidin reductase	LAR	4	6.93	7.15	6	5.10	6.33	1.83	0.82				
	anthocyanidin synthase	ANS	-	-	-	-	-	-	-	-	-	-	-	-
	anthocyanidin reductase	ANR	5	6.97	6.46	2	6.55	6.09	0.42	0.37				
	flavone synthase	FNS	-	-	-	-	-	-	-	-	-	-	-	-
	dihydroflavonol 4-reductase	DFR	11	4.10	3.79	10	3.68	3.52	0.42	0.27				
flavonol synthase	FLS	12	3.43	3.14	8	4.08	4.23	-0.65	-1.09					
Theanine	glutamate synthase	GOGAT	36	3.43	3.73	34	3.95	3.98	-0.52	-0.25				
	glutamate dehydrogenase	GDH	9	2.83	2.55	8	2.87	3.12	-0.04	-0.57				
	alanine aminotransferase	ALT	1	3.47	3.05	3	0.00	0.00	3.47	3.05				
	glutamine synthetase	GS	5	6.54	5.76	7	5.72	6.18	0.82	-0.42				
	theanine synthetase	TS	-	-	-	-	-	-	-	-	-	-	-	-
Caffeine	5'-nucleotidase	5'-Nase	4	3.95	3.60	5	4.12	3.76	-0.17	-0.16				
	IMP dehydrogenase	IMPDH	1	5.26	4.95	1	3.83	3.69	1.43	1.26				
	ribokinase	RBK	11	4.08	4.24	13	5.18	5.02	-1.10	-0.78				
	caffeine synthase	TCS	2	6.63	6.34	1	2.56	2.49	4.07	3.85				

significant differences in the contents of the major components from one bud and five leaves of tea versus oil tea, we used the transcriptome data to search for key genes in these metabolic pathways and to uncover the factors underlying this divergence. The quality of tea in large part depends on its metabolic profiles. We therefore performed additional analyses of catechin, theanine and caffeine biosynthesis. We were able to detect almost all genes in these metabolic pathways. Many of these genes appeared to form multigene families, implying that the tea genome, like the genomes of many other higher plants, had undergone one or more rounds of genome duplication during evolution [38], which might explain why higher

levels of gene expression did not always lead to higher enzyme activity in the present study. In our annotated tea and oil tea transcriptome dataset, multiple transcripts encoding all DEGs involved in flavonoid, theanine and caffeine biosynthesis pathways were identified.

Using a reciprocal best hit (RBH) method with relatively strict filters, 13,025 putative ortholog pairs were identified between tea and oil tea. We calculated their K_a (non-synonymous) / K_s (synonymous) ratios to estimate the rate of gene evolution [39, 40]. Of these ortholog pairs, 12,400 (95.2 % of 13,025) had a K_a/K_s value of 1 or less than 1, while 625 (4.8 % of 13,025) had a K_a/K_s value of over 1 (Additional file 6), suggesting that they were under positive selection (PS).

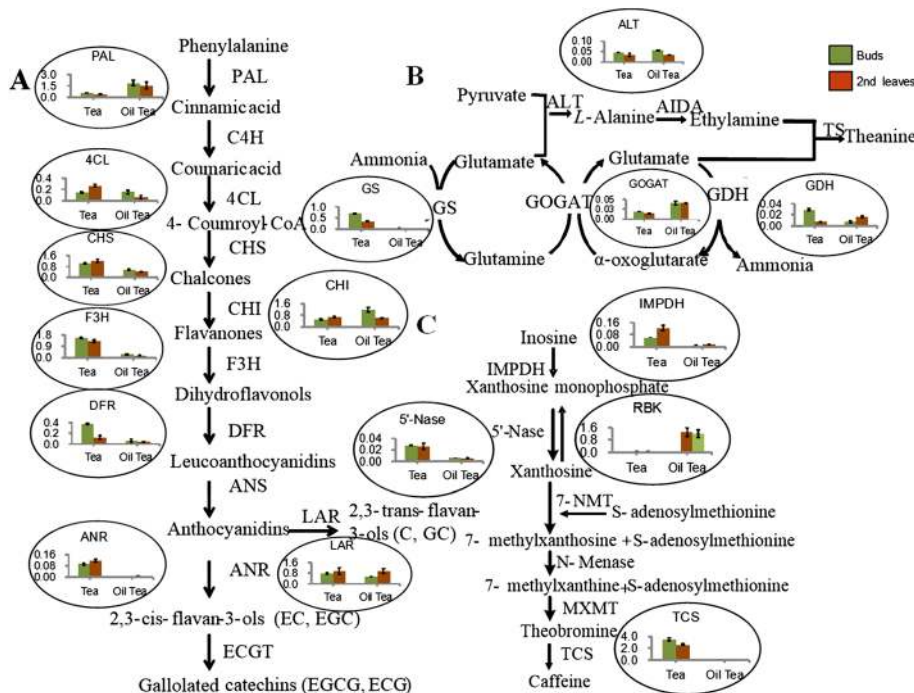


Fig. 4 Pathways of the three main secondary metabolites in tea and oil tea. **a** Core reactions of flavonoid biosynthesis involved in the biosynthesis of catechins. Compound names are shown below each arrow. Abbreviations beside the arrows indicate the enzymes catalyzing the transfer. The gene expression levels detected by qRT-PCR are shown in the histograms within the corresponding circles. Transcription level is indicated as the mean ($2^{\Delta Ct}$) \pm SD. **b** Core reactions of the theanine biosynthesis pathway. **(c)** Core reactions of the caffeine biosynthesis pathway

Functional GO analysis revealed that most genes under PS were grouped into GO terms cell, cell part, binding and metabolic process (Fig. 5). Of the 625 PS genes, 68 exhibited differential expression among tissues (Additional file 7). Notably, some PS orthologs encode CHI and DFR in the flavonoid pathway. CHI is a rate-limiting enzyme, and DFR is key enzyme, in the catechin-producing branch of the flavonoid biosynthesis pathway [41, 42]. Since the Ka/Ks ratio is widely used to detect selective pressure acting on protein-coding sequences [43, 44], rapid evolution of the CHI and DFR genes might be associated with adaptive selection in plants. No PS ortholog was assigned to the theanine or caffeine pathway. Environmental factors might play an important role in the evolution of the flavonoid pathway. Indeed, the highest quality green tea from Japan (a fine powder made from tencha) was grown in the shade and contains high levels of amino acids but low levels of catechins [45].

Conclusions

In this study, we examined the levels of characteristic metabolites in tea compared to oil tea, revealing (for the first time) trace amounts of theanine in oil tea. The contents of major metabolites were higher in tea than in oil tea. The genes involved in most of these pathways were more highly expressed in tea than in oil tea, especially key enzymes that function at branch points in these pathways, which might

explain the differential biosynthesis of metabolites (resulting in different components) in tea versus oil tea. Comparative transcriptome analyses demonstrated the connection between gene expression and the biosynthesis of catechins, theanine and caffeine. Comparative transcriptome analyses comparing the levels of metabolites between tea and oil tea not only enabled us to provide a preliminary description of the gene expression profiles, but it also helped elucidate the molecular mechanisms underlying the biosynthesis of characteristic biochemicals in tea. The transcriptome data obtained in this study will serve as an invaluable platform for further studies of the molecular biology and genomes of tea and oil tea.

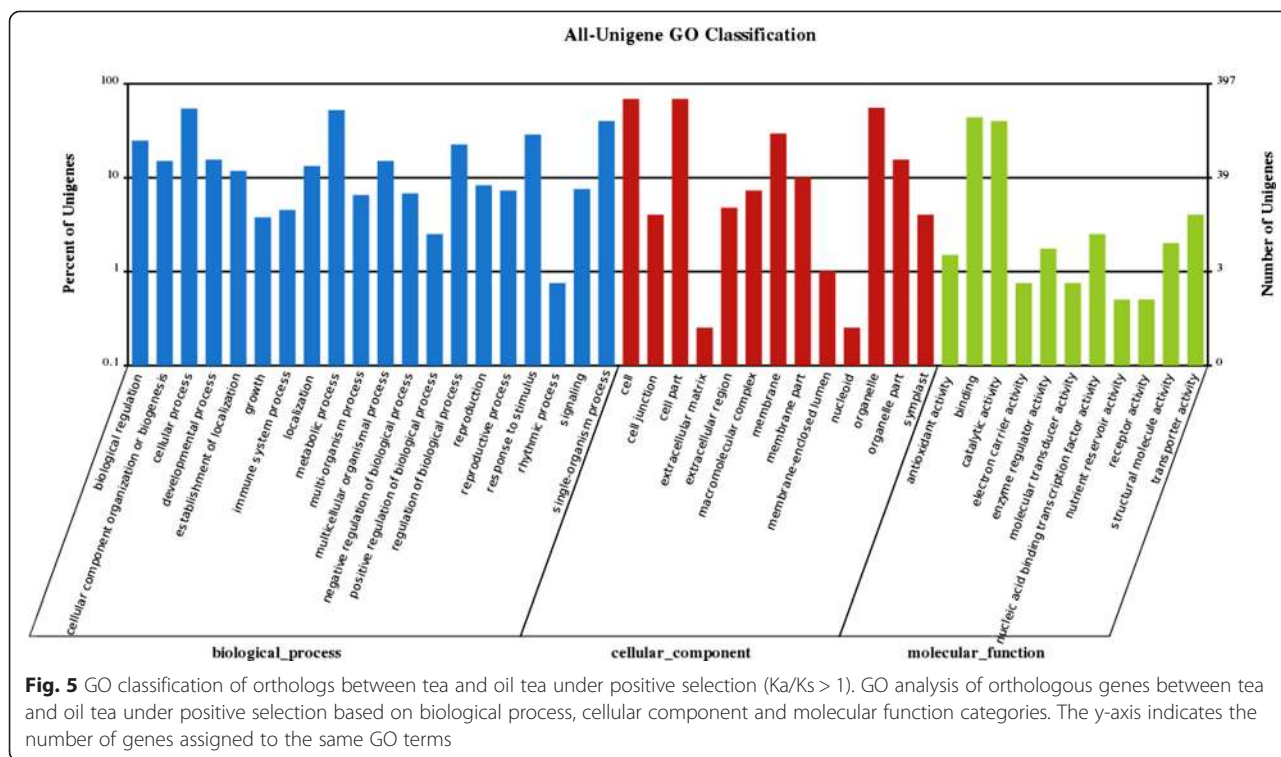
Methods

Plant materials

The six-year-old tea plants (*Camellia sinensis* [L.] O. Kuntze) and oil tea plants (*Camellia oleifera* Abel.) used in this study were grown in De Chang fabrication base in Anhui, China. One bud and five leaves were collected from each plant in the summer of 2013 (Fig. 1).

Extraction and HPLC analysis of catechins, theanine and caffeine

Catechins and caffeine were extracted from the samples according to the method described by Shan et al. [46]



with minor modifications. Briefly, 0.1 g of freeze-dried tea leaf tissue was ground in liquid nitrogen with a mortar and pestle and extracted with 3 mL 80 % methanol in an ultrasonic sonicator for 10 min at 4 °C. After centrifugation at 6,000 rpm for 10 min, the residues were re-extracted twice as described above. The supernatants were combined and diluted with 80 % methanol to a volume of 10 mL. The obtained supernatants were filtered through a 0.22 μm organic membrane before HPLC analysis.

The catechin and caffeine contents in the extracts were measured using a Waters 2695 HPLC system equipped with a 2489 ultraviolet (UV)-visible detector. A reverse-phase C18 column (Phenomenex 250 mm × 4.6 mm, 5 micron) was used at a flow rate of 1.0 mL/min. The detection wavelength was set to 278 nm, and the column temperature was 25 °C. The mobile phase consisted of 0.17 % (v/v) acetic acid (A) in water, 100 % acetonitrile (B), and the gradient elution was as follows: B 6 % from 0 to 4 min, to 14 % at 16 min, to 15 % at 22 min, to 18 % at 32 min, to 29 % at 37 min, to 45 % at 45 min, to 45 % at 50 min, to 6 % at 51 min and to 6 % at 60 min. Then, 10 μL of the filtrate was injected into the HPLC system for analysis. The filtered sample (10 μL) was injected into the HPLC system for analysis. Samples from each stage of leaf development were analyzed in triplicate.

Amino acids were extracted with hot water [47, 48]. Specifically, 0.15 g of freeze-dried tea leaves was ground in liquid nitrogen with a mortar pestle and extracted with 5 mL deionized water for 20 min in a water bath at

100 °C. After centrifugation at 6,000 rpm for 10 min, the residues were re-extracted once as described above. The supernatants were combined and diluted with water to a volume of 10 mL. The supernatants were also filtered through a 0.22 μm membrane before HPLC analysis. Theanine in tea was detected using a Waters 600E series HPLC system equipped with a quaternary pump and a 2489 ultraviolet (UV)-visible detector. A reverse-phase C18 column (Phenomenex 250 mm × 4.6 mm, 5 micron) was used at a flow rate of 1.0 mL/min. The column oven temperature was set to 25 °C. The detection wavelength was set to 199 nm for analysis [49]. The mobile phase consisted of 0.05 % (v/v) trichloroacetic acid (A) in water, 50 % acetonitrile (B), and the gradient elution was as follows: B 0 % (v/v) to 100 % at 40 min, to 100 % at 45 min and to 0 % at 60 min [31]. Then, 5 μL of the filtrate was injected into the HPLC system for analysis.

Amino acids in tea were detected using a Waters 600E series HPLC system equipped with a quaternary pump, a 2475 fluorescence detector and a 2489 ultraviolet (UV)-visible detector. The Waters AccQ•Tag method [50] with a Waters AccQ•Tag column (Nova-Pak C18, 4 μm, 150 mm × 3.9 mm) was employed to detect various amino acids according to the protocol of the AccQ•Fluor Reagent Kit [51, 52]. To determine the linearity of the chromatographic techniques, calibration plots of standards were constructed based on peak areas (y) using solutions of various concentrations (x). All plots were linear in the examined ranges; the linear ranges for

different concentrations of standard compounds are shown in the plots ($\mu\text{g mL}^{-1}$). The R^2 value refers to the correlation coefficient of the equation for calculating the content of a compound. The standard compounds C, EC, EGC, ECG, EGCG, GC, theanine and caffeine were purchased from Shanghai Winherb Medical Technology, Ltd., China.

Anthocyanin was extracted as follows: 0.1 g freeze-dry tea leaf tissue was ground in liquid nitrogen and extracted with 5 mL extraction solution (80 % methanol: 1 % hydrochloric acid [HCl]) using an ultrasonic sonicator for 10 min at room temperature. After centrifugation at 6,000 rpm for 10 min, the residues were re-extracted twice as described above. The supernatants were combined and diluted with extraction solution to 10 mL, followed by extraction with trichloromethane. The anthocyanin content was determined by colorimetry at 525 nm [53].

RNA extraction, library construction and RNA-Seq

Total RNA from tea and oil tea was extracted separately using the modified CTAB method [54]. The RNA integrity was measured using gel electrophoresis and spectrophotometry (Nanodrop). Equal amounts of RNA from three biological replicates were pooled prior to cDNA preparation. Enrichment of mRNA, fragment interruption, addition of adapters, size selection, PCR amplification and RNA-Seq were performed by staff at Beijing Genome Institute (BGI; Shenzhen, China). First, mRNA was enriched from 20 μg total RNA using magnetic beads with Oligo (dT) 25 (Invitrogen) and cleaved into short fragments. Second, using these short fragments as templates, first-strand cDNA synthesis was carried out with random primers (Japan, Takara) to produce double-stranded cDNA. Third, the ends of double-stranded cDNA fragments were further modified with T4 DNA polymerase, Klenow DNA polymerase and T4 polynucleotide kinase (Britain, NEB), and adapters were ligated to the short fragments using T4 DNA ligase (Invitrogen, USA). After the end repair process and ligation of adapters, the products were enriched by PCR to construct the final cDNA library. The cDNA library was examined using an Agilent 2100 Bioanalyzer. Finally, the four libraries were sequenced on an Illumina HiSeq™ 2000.

De novo assembly of RNA-Seq reads

Clean reads from four samples were obtained after quality control. Of these, two were from tea and two were from oil tea, which were combined and assembled separately using the transcriptome assembler Trinity [55]. The total and average lengths of assembled contigs were important criteria for transcriptome quality. Unigenes were defined after removing redundancy and short contigs from the assembly. Unigenes from tea and oil tea were aligned to each other iteratively using BLAST to identify homologous genes

in the two species; more than 80 % of the length of each gene in a pair of homologous genes was strictly aligned.

qRT-PCR analysis of the selected genes

To validate the accuracy of unigenes obtained from the assembled transcriptome and profiling of gene expression via RNA-Seq, qRT-PCR analysis was performed. RNA samples were extracted from the samples, and single-stranded cDNAs used for real-time PCR analysis were synthesized from the RNAs using a Prime-Script™ 1st Strand cDNA Synthesis Kit (TaKaRa, Dalian, China). The expression patterns of 34 transcripts were monitored. Detailed information about the selected transcripts, including their unigene IDs and the primer pairs designed in this study, is presented in Additional file 8. An IQ5 real-time PCR detection system (Bio-Rad) was utilized as previously described. The glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene was used as an internal reference gene, and relative expression was calculated using the $2^{-\Delta\Delta C_t}$ method [56]. All qRT-PCR analyses were performed in three biological and three technical replications.

Unigene functional annotation and classification

The unigenes were aligned to the protein sequence database NR, the Swiss-Prot protein database and COG [57] by Blastx with an E-value threshold of 1×10^{-5} . The unigenes were mapped to the KEGG metabolic pathway database [58]. Using KEGG annotation, metabolic pathway annotations of unigenes can be obtained, which helps elucidate the complex biological behaviors of genes. Using the COG database, orthologous gene products can be classified, and the possible functions of unigenes can be predicted. Based on NR annotation, GO classifications of unigenes were obtained using WEGO software [59] (<http://wego.genomics.org.cn/cgi-bin/wego/index.pl>) after annotation by the Blast 2 GO program (Version 2.3.4) [60] to elucidate the distribution of gene functions of a species at the macro level.

Comparison of nucleotide and protein sequence in tea and oil tea

Protein sequences from tea and oil tea were compared by BLAST and MUMmer (<http://mummer.sourceforge.net/>), and sequences with homology ≥ 70 % were retained.

Differentially expressed genes related to major secondary metabolism

KEGG pathway analysis was carried out to identify genes with different expression levels. Unigene expression was calculated using the FPKM method. The identification of differentially expressed genes (DEGs) was performed according to “The significance of digital gene expression profiles” [61], which was modified using a rigorous algorithm. $\text{FDR} \leq 0.001$ and the absolute value of

$\log_2\text{Ratio} \geq 1$ were chosen as the thresholds for judging the significance of differential expression of each gene. For a given unigene, four FPKM values were generated from the four transcriptomes, respectively. Hierarchical clustering was performed using Cluster 3. DEGs that may play important roles in major secondary metabolism (catechins, theanine and caffeine metabolism) were identified by ggplot2 (http://docs.ggplot2.org/current/geom_point.html) [62]. After further investigating metabolic pathways, several representative pathways were selected for more detailed analyses, including flavonoid metabolism, theanine metabolism and caffeine metabolism.

Identification of orthologous genes between tea and oil tea

To identify genes that are putatively orthologous between tea and oil tea, a RBH [63] method based on the Blastn program was used. Similar methods and criteria have been used in previous studies [64, 65]. Furthermore, GO classifications of putative orthologs under negative selection ($Ka/Ks < 1$) and under positive selection ($Ka/Ks > 1$) were compared using WEGO.

Estimation of synonymous and non-synonymous substitution rates between orthologous gene pairs

To calculate Ks and Ka substitution rate for each orthologous gene pair, the equivalent of a biological measurement of the nonsynonymous to synonymous substitution ratio Ka/Ks [66] was introduced, in which Ks and Ka were estimated by codeml (of the PAML package) using the likelihood method [67, 68].

Availability of supporting data

The Illumina RNA-seq data generated from buds and leaves of *Camellia sinensis* and *Camellia oleifera* are available in the NCBI SRA (<http://trace.ncbi.nlm.nih.gov/Traces/sra>) with accessions SRR1928149 and SRR1928150.

Additional files

Additional file 1: Annotation of assembled unigenes of tea and oil tea. (XLSX 16149 kb)

Additional file 2: Statistical comparison of DEGs between any two tissues. The genes were classified into three classes: red indicates up-regulated genes, green indicates down-regulated genes and blue indicates genes that are not differentially expressed. Tea buds, second leaves of tea, oil tea buds and second leaves of oil tea are abbreviated as TB, TL, OTB and OTL, respectively. (TIFF 2076 kb)

Additional file 3: The rich factors and number of genes involved in each pathway derived from a comparison of gene expression in TBvsTL, OTBvsOTL, TBvsOTB and TLvsOTL2. TBvsOTB stands for tea buds versus oil tea buds, TBvsTL for tea buds versus tea leaves, TLvsOTL for tea leaves versus oil tea leaves and OTBvsOTL for oil tea buds versus leaves. (XLSX 16 kb)

Additional file 4: Information about the genes involved in the flavonoid, theanine and caffeine biosynthesis pathways. Unigenes

involved in three metabolic pathways were annotated by alignment to the Swiss-Prot, COG and KEGG databases with e-values less than 1×10^{-30} . (XLSX 82 kb)

Additional file 5: List of DEGs analyzed by qRT-PCR. A total of 34 genes involved in the flavonoid, theanine and caffeine biosynthesis pathways were analyzed by qRT-PCR. (XLSX 16 kb)

Additional file 6: Estimated Ka and Ks values of 625 orthologous gene pairs ($Ka/Ks > 1$). The sequence length, nonsynonymous substitution rate (Ka), synonymous substitution rate (Ks), Ka/Ks and P-value are shown. (XLSX 51 kb)

Additional file 7: Distribution of DEGs under positive selection in annotated pathways. (XLSX 12 kb)

Additional file 8: Primers used for qRT-PCR. (DOCX 18 kb)

Abbreviations

HPLC: High-performance liquid chromatography; qRT-PCR: Quantitative real-time PCR; EC: Epicatechin; GC: Gallocatechin; EGC: Epigallocatechin; C: Catechin; EGCG: Epigallocatechin gallate; ECG: Epicatechin gallate; FPKM: Fragment Per Kilobase of exon model per Million mapped reads; NR: Non-redundant protein database; NT: Non-redundant nucleotide database; Swiss-Prot: Annotated protein sequence database; KEGG: Kyoto encyclopedia of genes and genomes; COG: Clusters of orthologous groups of protein; GO: Gene ontology; DEGs: Differentially expressed genes; TB: Tea buds; OTB: Oil tea buds; TL: Tea leaves; OTL: Oil tea leaves; PAL: Phenylalanine ammonia-lyase; CHI: Chalcone isomerase; F3H: Flavanone 3-hydroxylase; F3'5'H: Flavonoid 3',5'-hydroxylase; DFR: Dihydroflavonol 4-reductase; ANR: Anthocyanidin reductase; GS: Glutamine synthetase; GDH: Glutamate dehydrogenase; TS: Theanine synthetase; TCS: Tea caffeine synthase; IMPDH: Inosine-5'-monophosphate dehydrogenase; SAMS: S-adenosylmethionine synthetase; RBH: Reciprocal best hit; Ka : Non-synonymous; Ks : Synonymous; PS: Positive selection; HCl: Hydrochloric acid; GAPDH: Glycerinaldehyde-3-phosphate dehydrogenase.

Competing interest

The authors declare that they have no competing interests.

Authors' contributions

YLT performed the experiments and interpreted the sequence data. XCW conceived and revised the manuscript. CLW conceived and designed the experimental plan. HY designed the experimental plan and revised the manuscript. JZ participated in sample collection and RNA preparation. LZ designed the HPLC method. CTH revised the manuscript. SW was actively involved in manuscript revision and data analysis. CBF revised the manuscript. WWD was involved in experimental design. QC provided valuable comments and suggestions for improving the quality of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We appreciate Binbin Li, Wei Chen and Mingming Shi (Beijing Genome Institute at Shenzhen, China) for their technical support and initial data analysis. We thank Xiangyu Meng for material collection. We are grateful to Shancen Zhao for comments and revision on the manuscript. This work was supported by grants from the Major Project of Chinese National Programmes for Fundamental Research and Development (No. 2012CB722903), the Science and Technology Project of Anhui Province, China (Project 13Z03012), the Tea Genome Project of Anhui Province, China, Anhui Major Demonstration Project for Leading Talents Team on Tea Chemistry and Health and the National Natural Science Foundation of China (NSFC) projects (No. 31170283, Project 31300578 and 31300576), and the Academic Backbone Cultivation Project of Anhui Agricultural University (2014XKPY-34). We are also grateful to the unknown editor at the elixigen editing service (ID141224-4541) for the English polishing.

Author details

¹State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei 230036, People's Republic of China. ²Department of Food Science, Rutgers University, Rutgers, NJ, USA.

Received: 1 March 2015 Accepted: 17 July 2015

Published online: 07 August 2015

References

- Liang YR, Ma WY, Lu JL, Wu Y. Comparison of chemical compositions of *Ilex latifolia* Thunb and *Camellia sinensis* L. *Food Chem.* 2001;75(3):339–43.
- Mamati GE, Liang Y, Lu J. Expression of basic genes involved in tea polyphenol synthesis in relation to accumulation of catechins and total tea polyphenols. *J Sci Food Agric.* 2006;86(3):459–64.
- Punyasiri PAN, Abeyasinghe ISB, Kumar V, Treutter D, Duy D, Gosch C, et al. Flavonoid biosynthesis in the tea plant *Camellia sinensis*: properties of enzymes of the prominent epicatechin and catechin pathways. *Arch Biochem Biophys.* 2004;431(1):22–30.
- Obanda M, Owuor PO. Impact of shoot maturity on chlorophyll content, composition of volatile flavour compounds and plain black tea chemical quality parameters of clonal leaf. *J Sci Food Agric.* 1995;69(4):529–34.
- Owuor PO, Obanda M, Nyiranda HE, Mandala WL. Influence of region of production on clonal black tea chemical characteristics. *Food Chem.* 2008;108(1):263–71.
- Nagar PK, Sood S. Changes in endogenous auxins during winter dormancy in tea (*Camellia sinensis* L.) O. Kuntze. *Acta Physiol Plant.* 2006;28(2):165–9.
- Zhao LP. Construction of cDNA Library of Tea Plant (*Camellia sinensis*) and EST Sequences Analysis of Tea Specific Genes. Beijing: Tea Research Institute, Chinese Academy of Agricultural Sciences; 2004.
- Chen L, Zhao L, Gao Q. Generation and analysis of expressed sequence tags from the tender shoots cDNA library of tea plant (*Camellia sinensis*). *Plant Sci.* 2005;168(2):359–63.
- Das A, Das S, Mondal TK. Identification of differentially expressed gene profiles in young roots of tea [*Camellia sinensis* (L.) O. Kuntze] subjected to drought stress using suppression subtractive hybridization. *Plant Mol Biol Report.* 2012;30(5):1088–101.
- Wang XC, Zhao QY, Ma CL, Zhang ZH, Cao HL, Kong YM, et al. Global transcriptome profiles of *Camellia sinensis* during cold acclimation. *BMC Genomics.* 2013;14(1):415.
- Thirugnanasambantham K, Prabu G, Palanisamy S, Chandrabose SRS, Mandal AKA. Analysis of dormant bud (Banjhi) specific transcriptome of tea (*Camellia sinensis* (L.) O. Kuntze) from cDNA library revealed dormancy-Related genes. *Appl Biochem Biotechnol.* 2013;169(4):1405–17.
- Shi CY, Yang H, Wei CL, Zhang ZH, Cao HL, Kong YM, et al. Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genomics.* 2011;12(1):131.
- Liu YJ, Gao LP, Liu L, Yang Q, Lu ZW, Nie ZY, et al. Purification and characterization of a novel galloyltransferase involved in catechin galloylation in the tea plant (*Camellia sinensis*). *J Biol Chem.* 2012;287(53):44406–17.
- Yuan Y, Song LP, Li MH, Liu GM, Chu YN, Ma LY, et al. Genetic variation and metabolic pathway intricacy govern the active compound content and quality of the Chinese medicinal plant *Lonicera japonica thunb*. *BMC Genomics.* 2012;13(1):195.
- Feng C, Chen M, Xu C, Bai L, Yin XR, Li X, et al. Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-Seq. *BMC Genomics.* 2012;13(1):19.
- Zhang J, Liang S, Duan J, Wang J, Chen SL, Cheng ZS, et al. De novo assembly and Characterisation of the Transcriptome during seed development, and generation of genic-SSR markers in Peanut (*Arachis hypogaea* L.). *BMC Genomics.* 2012;13(1):90.
- Hao DC, Ge GB, Xiao PG, Zhang YY, Yang L. The first insight into the tissue specific taxus transcriptome via Illumina second generation sequencing. *PLoS ONE.* 2011;6(6), e21220.
- Sun W, Xu X, Zhu H, Zhu HS, Liu AH, Liu L, et al. Comparative transcriptomic profiling of a salt-tolerant wild tomato species and a salt-sensitive tomato cultivar. *Plant Cell Physiol.* 2010;51(6):997–1006.
- Koenig D, Jiménez-Gómez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, et al. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc Natl Acad Sci.* 2013;110(28):E2655–62.
- Wu ZJ, Li XH, Liu ZW, Xu ZS, Zhuang J. De novo assembly and transcriptome characterization: novel insights into catechins biosynthesis in *Camellia sinensis*. *BMC Plant Biol.* 2014;14(1):277.
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 2012;40(D1):D109–14.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008;5(7):621–8.
- Eungwanichayapant PD, Popleuchai S. Accumulation of catechins in tea in relation to accumulation of mRNA from genes involved in catechin biosynthesis. *Plant Physiol Biochem.* 2009;47(2):94–7.
- Ashihara H, Deng WW, Mullen W, Crozier A. Distribution and biosynthesis of flavan-3-ols in *Camellia sinensis* seedlings and expression of genes encoding biosynthetic enzymes. *Phytochemistry.* 2010;71(5):559–66.
- Wu H, Chen D, Li J, Li JX, Yu B, Qiao XY et al. De novo characterization of leaf transcriptome using 454 sequencing and development of EST-SSR markers in tea (*Camellia sinensis*). *Plant Mol Biol Report.* 2013;31(3):524–38.
- Ashihara H, Sano H, Crozier A. Caffeine and related purine alkaloids: biosynthesis, catabolism, function and genetic engineering. *Phytochemistry.* 2008;69(22):841–56.
- Li N, de Silva J. Theanine: Its Occurrence and Metabolism in Tea. Annual Plant Reviews Volume 42: Nitrogen Metabolism in Plants in the Post-Genomic Era. UK:Oxford;2010.p.171–206.
- Zhang LL, Wang YM, Wu DM, Xu M, Che JH. Microwave-assisted extraction of polyphenols from *Camellia oleifera* fruit hull. *Molecules.* 2011;16(6):4428–37.
- Khan N, Mukhtar H. Tea polyphenols for health promotion. *Life Sci.* 2007;81:519–33.
- Rusak G, Komes D, Likić S, Horžić D, Kovač M. Phenolic content and antioxidative capacity of green and white tea extracts depending on extraction conditions and the solvent used. *Food Chem.* 2008;110(4):852–8.
- Ying Y, Ho JW, Chen ZY, Wang J, et al. Analysis of theanine in tea leaves by HPLC with fluorescence detection. *J Liq Chromatogr Relat Technol.* 2005;28(5):727–37.
- Chen JH, Liao BC, Jong TT, Chang CMJ. Extraction and purification of flavanone glycosides and kaempferol glycosides from defatted (*Camellia oleifera*) seeds by salting-out using hydrophilic isopropanol. *Sep Purif Technol.* 2009;67(1):31–7.
- Xiao X S, Zhang M, Yuan XY, Li D, Yang Y. Studies on the extraction and purification of total flavones in *Camellia oleifera* leaves. *China Food Additives.* 2012;4:93–7.
- Yang D, Liu Y, Sun M, Zhao L, Wang YS, Chen XT, et al. Differential gene expression in tea (*Camellia sinensis* L.) calli with different morphologies and catechin contents. *J Plant Physiol.* 2012;169(2):163–75.
- Blankenship SM, Unrath CR. PAL and ethylene content during maturation of red and golden delicious apples. *Phytochemistry.* 1988;27(4):1001–2.
- Kataoka I, Kubo Y, Sugjura A, Tomana T. Changes in L-phenylalanine ammonia-lyase activity and anthocyanin synthesis during berry ripening of three grape cultivars. *J Jpn Soc Hortic Sci.* 1983;52:273–9.
- Xiong L, Li J, Li Y, Yuan L, Liu S, Huang J et al. Dynamic changes in catechin levels and catechin biosynthesis-related gene expression in albino tea plants (*Camellia sinensis* L.). *Plant Physiol Biochem.* 2013;71(2):132–43.
- Sakuma S, Pourkheirandish M, Matsumoto T, Takashi K, Takao K. Duplication of a well-conserved homeodomain-leucine zipper transcription factor gene in barley generates a copy with more specific functions. *Funct Integr Genomics.* 2010;10(1):123–33.
- Miyata T, Yasunaga T. Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *J Mol Evol.* 1980;16(1):23–36.
- Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol.* 2000;15(12):496–503.
- Liao JJ, An CC, Wu S, Chen ZL. The role of chalcone synthase gene in the defense response of higher plants against pathogens. *Acta Sci Nat Univ Pekin.* 2000;36(4):565–75.
- Nishihara M, Nakatsuka T, Yamamura S. Flavonoid components and flower color change in transgenic tobacco plants by suppression of chalcone isomerase gene. *FEBS Lett.* 2005;579(27):6074–6078.2.
- Du YY, Liang YR, Wang H, Wang KR, Lu JL, Zhang GH, et al. A study on the chemical composition of albino tea cultivars. *J Hortic Sci Biotech.* 2006;81(5):809–12.
- Feng L, Gao MJ, Hou RY, Hu XY, Liang Z, Wan XC, et al. Determination of quality constituents in the young leaves of albino tea cultivars. *Food Chem.* 2014;155:98–104.
- Ku KM, Choi JN, Kim J, Kim JK, Lee SJ, Lee CH. Metabolomics analysis reveals the compositional differences of shade grown tea (*Camellia sinensis* L.). *J Agric Food Chem.* 2009;58(1):418–26.
- Yu S, Li WW, Wang YS, Liu YJ, Wang HX, Wang XF, et al. Catechins synthesis and accumulation in tea seedlings at different development stages. *J Anhui Agric Univ.* 2011;38(4):600–5.

47. Wang L, Gong LH, Chen CJ, Han HB, Li HH. Column-chromatographic extraction and separation of polyphenols, caffeine and theanine from green tea. *Food Chem.* 2012;131(4):1539–45.
48. Yen GC, Chen HY. Antioxidant activity of various tea extracts in relation to their antimutagenicity. *J Agric Food Chem.* 1995;43(1):27–32.
49. Zhu XL, Chen B, Luo XB, Yao SZH, et al. Determination of theanine in tea by reversed-phase high performance liquid chromatography. *Chin J Chromatogr.* 2003;21(4):400–2.
50. Anderstam B, Katzarski K, Bergström J. Serum levels of NG, NG-dimethyl-L-arginine, a potential endogenous nitric oxide inhibitor in dialysis patients. *J Am Soc Nephrol.* 1997;8(9):1437–42.
51. Ni J, XU H. Research progresses in the method for determining the amino acids in tea. *J Tea.* 2007;2:001.
52. Heresztyn T, Worthley MI, Horowitz JD. Determination of L-arginine and N G, N G-and N G, N G'-dimethyl-L-arginine in plasma by liquid chromatography as AccQ-Fluor™ fluorescent derivatives. *J Chromatogr B.* 2004;805(2):325–9.
53. Lee J, Rennaker C, Wrolstad RE. Correlation of two anthocyanin quantification methods: HPLC and spectrophotometric methods. *Food Chem.* 2008;110(3):782–6.
54. Shi CY, Wan XC, Jiang CJ, Sun J. Method for high-quality total RNA isolation from tea plants [*Camellia sinensis* (L.) O. Kuntze]. *Journal of Anhui Agricultural University.* 2007;34(3):360–3.
55. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29(7):644–52.
56. Livaka KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods.* 2001;25(4):402–8.
57. Tatusov RL, Natale DA, Garkavtsev IV, Tatischeva T, Ueda S, Bachoti SR. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 2001;29(1):22–8.
58. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 2004;32 suppl 1:D277–80.
59. Ye J, Fang L, Zheng H, Chen J, Zhang Z, Wang J, et al. WEGO: a web tool for plotting GO annotations. *Nucleic Acids Res.* 2006;34 suppl 2:W293–7.
60. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005;21(18):3674–6.
61. Audic S, Claverie JM. The significance of digital gene expression profiles. *Genome Res.* 1997;7(10):986–95.
62. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn J, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol.* 2013;31(1):46–53.
63. Xia EH, Jiang JJ, Huang H, Zhang LP, Zhang HB, Gao LZH. Transcriptome analysis of the oil-rich tea plant, *Camellia oleifera*, reveals candidate genes related to lipid metabolism. *PLoS ONE.* 2014;9(8), e104150.
64. Wang JT, Li JT, Zhang XF, Zhang XF, Sun XW. Transcriptome analysis reveals the time of the fourth round of genome duplication in common carp (*Cyprinus carpio*). *BMC Genomics.* 2012;13(1):96.
65. Zhang L, Yan HF, Wu W, Yu H, Ge XJ. Comparative transcriptome analysis and marker development of two closely related Primrose species (*Primula poissonii* and *Primula wilsonii*). *BMC Genomics.* 2013;14(1):329.
66. Hu T, Banzhaf W. Nonsynonymous to Synonymous Substitution Ratio k_a/k_s : Measurement for Rate of Evolution in Evolutionary Computation Parallel Problem Solving from Nature—PPSN X. Berlin Heidelberg: Springer; 2008. p. 448–57.
67. Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol.* 1994;11(5):725–36.
68. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24(8):1586–159.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

