

Comparative Transcriptomics of Metabolites in Wild and Cultivated *Pinellia ternata*

Jue Wang

Affiliated Hospital of Nanjing University of Chinese Medicine

Yufeng Zhu

Affiliated Hospital of Nanjing University of Chinese Medicine

Xiangyu Lv

China Pharmaceutical University

Yue Xiong

Jiangsu Province Animal Product Quality Inspection and Testing Center

Jiawei Lu

China Pharmaceutical University

Chao Jin

Affiliated Hospital of Nanjing University of Chinese Medicine

Jian Zhang

China Pharmaceutical University

Shijia Liu (✉ liushijia_2019@126.com)

Affiliated Hospital of Nanjing University of Chinese Medicine <https://orcid.org/0000-0002-3521-3618>

Research article

Keywords: RNA-Seq, *Pinellia ternata*, Differentially expressed genes, Transcriptome

Posted Date: May 26th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-29019/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: As a traditional Chinese medicine(TCM) used to treat cough, *Pinellia ternata* (PT) has been utilized for thousands of years. Studies focused on the genetic makeup underlying the differences between wild and cultivated PT are rare, while wild herb is quite effective and indispensable in the clinical use of TCM. There is still rare researches related to the cause of differences between wild and cultivated PT. With the application of transcriptome technology in the field of traditional Chinese medicine, High-throughput RNA sequencing (RNA-seq) could be employed to find out how environment affect the accumulation of active constituent in medicinal herbs.

Result: Overall, 133 503 unigenes were assembled, and 1329 genes were differentially expressed in the cultivated PT compared to the wild PT. Of those, 70 DEGs (differentially expressed genes) were significantl[1]y downregulated and 9 were upregulated in the cultivated PT compared to the wild PT after filtering out genes involved in the biosynthesis and metabolism of amino acids, glutathione, phenylpropane and other constituents. These genes were assumed to be of special interest with respect to the better efficacy of the wild PT. Meanwhile, RT-qPCR was utilized to analysis the expression patterns of 9 selected genes and HPLC were applied to quantify the content of adenine and guanosine in PT which verified the RNA-Seq result.

Conclusion: 79 DEGs were found in wild and cultivated PT, which played an important role in the synthesis and metabolism of the active ingredients of PT. And the expression of DEGs is higer in wild PT, indicated the importance of wild medicinal herb resources. The experiment produced robust resources for future investigations on the regulation of active ingredients in PT and Araceae and the promotion of the cultivation and clinical use of PT.

Background

Pinellia ternata (PT), is the dried tuber of *Pinellia ternata* (Thunb.) Breit, which belongs to the Araceae family. PT has been used for the treatment of cough and inflammation since ancient times (Xie et al. 2016). Currently, PT is widely used clinically for its antitumor, antioxidant, antitussive, expectorant, antiemetic, antibacterial and anti-inflammatory effects in China, Korea and other Asian countries (Li et al. 2016; Zhang et al. 2013; Wu et al. 2011). Plant secondary metabolites (SMs) can be a useful array of natural products, helping plants defend against pathogens or herbivores (Yang et al. 2018; Kliebenstein and Osbourn 2013). Previous phytochemical investigations have shown that there are different types of SMs ion PT, including alkaloids, essential oils, amino acids, organic acids, polysaccharides, and proteins (Wu et al. 1999; Gonda et al. 1994; Yang et al. 2014).

As a plant adapted to humid environments, genuine *P. ternata* is distributed in the southern margin of the Huanghuai Plain in China. However, the heavy use of herbicides has had a serious impact on this species. Additionally, the rapidly developing economy has caused rising labor costs, and wild resources are scarce. Therefore, *P. ternata* planting has already shifted to the western and northern regions, such as Guizhou and Gansu (Mu et al. 2013). The accumulation of SMs in PT varies in different growth environments, and its active ingredients are highly variable (Lee et al. 2009). The authenticity and quality of TCMs are directly related to their safety and efficiency in clinical applications (Lv et al. 2016). Thus, the inconsistency of the quality of wild and cultivated PT may directly affect its efficacy. Since the use of PT is still increasing, to ensure the uniformity of its clinical effects, it is necessary to conduct a comprehensive study of PT in different regions and provide support for solving the quality control problem.

The application of omics in the field of medical research is expanding, and as an emerging high-throughput sequencing technology, RNA-Seq is gradually being applied to transcriptome research due to its low cost, high speed and high precision (Xiao et al. 2016; Marioni et al. 2008), providing a particularly effective technology for the discovery of genes from many organisms and tissue types (Birol et al. 2009; Trapnell et al. 2010), especially in nonmodel species. These technologies also promote research progress in the field of traditional Chinese medicine. Since 2010, RNA-seq technology has been used in Chinese herbal medicines (Chen et al. 2011; Liu et al. 2015; Fukushima et al. 2016).

To date, he investigation of the transcriptome of PT has been performed on the heat stress response of PT leaves and the biosynthesis of alkaloids in PT (Ma et al. 2018; Zhu et al. 2013). However, there has been little research focused on the

differences between PT from planted and wild origins. Focusing on this distinction will help to reveal the quality formation of PT.

In this study, RNA-Seq technology based on de novo assembly was applied to conduct transcriptomics on PT from seven different origins (cultivated: S1-S6; wild: S7, CON). The transcript levels of 9 target genes were determined by real-time PCR, which provided a basis for revealing the molecular mechanism of the quality differences in PT from wild and cultivated fields.

Here, we report the first high coverage comparative transcriptome analysis of wild and cultivated PT from different origins. We analyzed the transcriptomes of PT from six cultivated areas and two wild areas using three biological replicates. We identified and compared DEGs in cultivated and wild PT. Our comprehensive transcriptomic analyses provides a basis for revealing the molecular mechanism of the quality difference between wild and cultivated PT as well as robust resources for future investigations on the regulation of active ingredients in PT and Araceae.

Methods

Preparation of plant materials

PT tubers were collected from 8 locations in China during 2018 (Table 1). Collected tuber samples were rinsed clean frozen immediately in liquid nitrogen and stored at -80°C before RNA extraction.

Library construction for transcriptome analysis

Total RNA was extracted using the mirVana miRNA Isolation Kit (Ambion) following the manufacturer's protocol. RNA integrity was evaluated using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The samples with an RNA integrity number (RIN) ≥ 7 were subjected to subsequent analyses. The libraries were constructed using the TruSeq Stranded mRNA LT Sample Prep Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. Then, these libraries were sequenced on an Illumina sequencing platform (HiSeqTM 2500 or Illumina HiSeq X Ten), and 125 bp /150 bp paired-end reads were generated. The raw data were submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under accession No. PRJNA623739.

Quality control and de novo assembly

Transcriptome sequencing and analyses were conducted by OE Biotech Co., Ltd. (Shanghai, China). Raw data (raw reads) were processed using Trimmomatic. The reads containing poly-N and the low-quality reads were removed to obtain the clean reads. After removing adapter and low-quality sequences, the clean reads were assembled into expressed sequence tag clusters (contigs) and de novo assembled into transcripts by using the paired-end method in Trinity (version: 2. 4).

Annotation of gene functions

The function of the unigenes was annotated by alignment of the unigenes with the NCBI non-redundant (NR), SwissProt, and Clusters of orthologous groups for eukaryotic complete genomes (KOG) database using Blastx with a threshold E-value of $1e^{-10}$. The proteins with the highest hits to the unigenes were used to assign functional annotations. Based on the SwissProt annotation, Gene Ontology (GO) classification was performed by mapping the relationships between SwissProt and the GO terms. The unigenes were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to annotate their potential metabolic pathways.

Unigene quantification, analysis of differentially expressed unigenes (DEGs), cluster analysis, GO and KEGG enrichment

The FPKM and read count values of each unigene were calculated using bowtie2 and eXpress, respectively. DEGs were identified using the DESeq (2012) functions estimate Size Factors and nbinom Test. A P value < 0.05 and fold change > 2 or < 0.5 were used as the thresholds for differential gene expression. Five was set as the threshold for significantly differential gene

expression. Hierarchical cluster analysis of DEGs was performed to explore transcript expression patterns. GO enrichment and KEGG pathway enrichment analyses of DEGs were respectively performed using R based on the hypergeometric distribution.

Real-time quantitative PCR (RT-qPCR) for expression validation

The expression patterns of 9 genes associated with metabolite and biosynthesis pathways related to the active ingredients in PT were analyzed using RT-qPCR (gene-specific primers are provided in Additional file 1: Table S1). Total RNA was isolated from the roots using TRIzol (Tiangen Bio Co., Ltd., Beijing, China), and cDNA was synthesized using the PrimeScript™ RT Reagent Kit with a gDNA Eraser. RT-qPCR was conducted using SYBR® Premix Ex Taq™ II according to the manufacturer's instructions (Takara, Tokyo, Japan). The GAPD genes (5-TGCTGGGAATGATGTTGAATG-3 and 5-TTGGCATTGTTGAGGGTTTG-3) were used as internal standards. The PCR program was as follows: 30 s of predenaturation at 95°C, followed by 40 cycles of 5 s at 95°C and 40 s at 60°C and the steps for dissociation curve generation (10 s at 95°C, 60 s at 60°C, and 15 s at 95°C). The relative gene expression of each target gene was determined using the $2^{-\Delta\Delta C_t}$ method.

Determination of adenine and guanosine

HPLC were applied to determine the content of adenine and guanosine in wild and cultivated PT. Two gram of the dry powder of PT were weighed and transferred into 50 mL erlenmeyer flask. 10mL of purified water was used for extraction. The vials containing the samples and the solvents were ultrasonic extracted at room temperature for 60 min, and then the solutions were filtered with 0.22µm micro-porous membrane before injection. Stock solution, approximately 1 mg of each adenine and guanosine reference standard were accurately weighed into a centrifuge tube, dissolved in 1 mL methanol by sonication for 1 min. Working standard solutions were prepared by 20 fold dilution of the stock solution with methanol prior to HPLC analysis.

An Agilent 1260 series liquid chromatograph system comprising vacuum degasser, quaternary pump, autosampler, thermostatted column compartment, and diode array detector was used. The column used was a C18 reversed Agilent Eclipse SB-Aq 5 mm (4.6 mm×250 mm). Mobile phases consisted of 0.2% ammonium dihydrogen phosphate solution (v/v) (eluent A) and methanol (v/v) (eluent B). The gradient was as follows: 0–6 min, 98% A B; 6– 8 min, linear gradient from 98 to 95% A; 8–15 min, linear gradient from 95 to 80% A; 15–20 min, linear gradient from 80 to 70% A; 20–25 min, 70% A; 25–45 min, 70 to 10 % A; 45–50 min, 10 to 98 %; 50–55 min, 98 % A. Post-run time was 5 min. Elution was performed at a solvent flow rate of 1 ml/min. Detection was accomplished with a diode array detector and chromatograms were recorded at 260 nm. The column was maintained at 30 C. The sample injection volume was 20 µl.

Results

De novo assembly of transcripts

cDNAs prepared from tubers of PT were sequenced using the Illumina HiSeq platform.

As a result of sequencing, more than 48 595 258 clean reads and a total of 62 348 392 clean bases were obtained from 24 PT samples. The Q30 percentage was over 93.56% (Table 2), and the distribution of GC content frequency ranged from 49.63% % to 51.82%, showing that the throughput and sequencing quality were high enough for further analysis.

A final set of unigenes was obtained by the splicing method in Trinity software and paired-end and clustered by CD-HIT software. The de novo assembly generated 57 154 unigenes that were considered for downstream analysis (Table 3, Fig. 1). The length distributions of the unigenes are shown in Table 3. There were 36 007 unigenes with lengths above 1000 bp, and the maximum length was 13 803 bp. Based on these unigenes, 76 894 unigenes above 500 bp were constructed. The length of unigenes ranged from 301 bp to 13803 kb, with an average length of 861 25 bp.

Functional annotation and classification

Unigenes were compared to the NR, KOG, GO, Swiss-Prot, eggNOG, and KEGG databases by diamond software, and the Pfam database was used to perform unigene functional analysis using HMMER software. A total of 53 323 unigenes (39.94% of the

total) were matched in the NR database, 31 981 unigenes (23.96% of the total) were matched in the KOG database, 36 197 unigenes (27.11% of the total) were matched in the Swiss-Prot database, 18,831 unigenes (14.11% of the total) were matched in the KEGG database, 33 006 unigenes (24.72% of the total) were matched in the GO database, 123 unigenes (0.09% of the total) were matched in the Pfam database, and 48 128 unigenes (36.05% of the total) were matched in the eggNOG database with an E-value $<1e-5$ (Table 4).

GO annotations were classified into three subcategories: biological process (BP), molecular function (MF), and cellular component (CC) (Additional file 2: Table S2, Fig. 2). In the cellular component category (29 007; 21.73%), cells, cell parts, and organelles were shown to be the top 3 terms. Among the molecular function category (28 318; 21.21%), binding and catalytic activities were dominant. In the biological process category (26 484; 19.84%), cellular process, metabolic process, and response to stimulus were the top terms.

To explore the pathways in which these annotated genes were contained, KEGG analysis with E values $\leq 1e-5$ was also conducted. A total of 18 831 (14.11%) annotated unigenes were assigned to 182 pathways that corresponded to four KEGG modules: metabolism, environmental information processing, genetic information processing and cellular processes. (Fig. 3, Additional file 3: Table S3).

Among them, 9 229 unigenes were assigned to metabolic pathways. As shown in Fig. 3, the dominant pathways related to metabolism were mainly classified as carbohydrate metabolism (1 833), followed by global and overview map metabolism (1 232 unigenes), amino acid metabolism (1 108 unigenes), and energy metabolism (1 062 unigenes). Metabolic pathway analysis enables us to determine which genes in the same pathway cooperate with each other as they exercise their biological functions in PT.

Overall, the possible functions of the assembled unigenes were assessed by similarity matches with the GO and KEGG databases. The results of these database searches help us to better understand the biological features of PT.

Differentially expressed genes (DEGs) in PT from wild and cultivated fields

Unigene expression was calculated using their FPKM values (Fragments Per kb Per Million Reads), and principal component analysis (Fig. 4) was performed according to the expression levels of each sample, showing that the samples from the same origin were concentrated, and the samples from the S5 and S6 groups in Hebei Province had similarities. Based on the multiple difference and significant difference test results, 16 242 differentially expressed unigenes were screened, of which 8 598 were upregulated and 7 644 were downregulated (P value < 0.05 , fold change > 2). Unsupervised hierarchical clustering analysis was performed on the differentially expressed unigenes. As shown in the heat map in Fig. 5, there were significant differences between the cultivated and wild PT samples. Overall, these results indicate that there are significant differences between PT in different regions and between cultivated and wild PT, which help in categorizing cultivated accessions.

To better understand the biological functions of the specific transcripts of wild and cultivated plants, KEGG and GO databases were both utilized for pathway analysis of the different unigenes (combined with KEGG annotation results). Significantly differentially expressed genes related to metabolic pathways of active ingredients in PT (screening for unigenes of a single group of samples ≥ 3 and P value ≤ 0.01) were selected and are shown in Additional file 4: Table S4; 70 of these genes were significantly downregulated and 9 genes were upregulated.

Validation of RNA-seq analysis by RT-qPCR

To confirm the RNA-Seq results, the average expression levels of 9 selected genes, β -glucosidase (β Glu), β -fructofuranosidase (FFase), sucrose-phosphate synthase (SPS), glutamate decarboxylase (GDA), glutamate decarboxylase (GS), 5'-nucleotidase (5'-NT), caffeic acid 3-O-methyltransferase (COMT), glutathione S-transferase (GST) and glutathione peroxidase (GPX) were quantified via RT-qPCR (Additional file 5: Table S5), and the average expression levels (FPKM values) between wild and cultivated samples were compared. There was a high degree of correlation between the RT-qPCR results and the RNA-Seq results (Fig. 6). The expression levels of the 9 selected genes were higher in wild PT than in cultivated PT. The RT-qPCR data

indicated that genes were regulated as indicated in the transcriptomic data, which further supported the accuracy of the transcriptomic results.

Determination of adenine and guanosine

The Chromatogram of catechin standards and PT are shown in Fig. 7. 10 cultivated PT and 3 wild PT dry extracts were analysed and the results are shown in Table 5 and Fig. 8. As expected, wild PT contained a higher quantity of adenine and guanosine than cultivated PT.

Discussion

Pinellia ternata is one of most important medicinal herbs in China, which makes it vital for genomic studies. With the transformation of cultivation areas and scarcity of wild PT resources, to ensure clinical efficacy, there is a growing need for the genetic characterization of wild and cultivated PT. Using transcriptomic evidence, we provided an in-depth view of the differences between wild and cultivated PT, which may explain the better clinical efficacy of wild PT. Among the DEGs screened in this experiment, compared with cultivated PT, wild PT had significantly more upregulated DEGs, which were involved in the synthesis and metabolic pathways of the active ingredients widely known to exist in PT (including amino acids, nucleosides, etc.).

Carbohydrate metabolism

Pinellia ternata deposits carbohydrates, particularly starch and polysaccharides. 15 key target DEGs involved in carbohydrate metabolism and pathways were found to be downregulated in PT from cultivated fields. These genes were fructokinase (ScrK), GDP-L-galactose phosphorylase (GGP), malate dehydrogenase (MDH), fructose-bisphosphate aldolase (ALDO), aconitate hydratase (ACO), L-ascorbate peroxidase (APX), beta-glucosidase (bglX), SPS, trehalose 6-phosphate synthase (TPS), beta-fructofuranosidase (Inv), mannose-1-phosphate guanylyltransferase (GMPP), catalase (CAT), β Glu, FFase and succinic semialdehyde reductase (GLYR). Five of these DEGs were found to be upregulated: glyceraldehyde 3-phosphate dehydrogenase (GAPDH), pyruvate kinase (PK), glucose-1-phosphate adenyltransferase (glgC), 1,4-alpha-glucan branching enzyme (glgB) and inositol 3-alpha-galactosyltransferase (GOLS).

Ascorbic acid (AsA) plays important roles as a major antioxidant and free-radical scavenger in plants. The L-galactose pathway is the main pathway for AsA biosynthesis in higher plants (LinsterC and ClarkeS 2008). As a key enzyme, GGP catalyzes GDP-L-galactose to L-galactose-1-phosphate, which is beneficial to the formation of AsA and thus responds to abiotic stress. In addition to GPX (downregulated) in cultivated PT, APX also plays a vital role in plant antioxidation and abiotic stress, including abiotic stress responses induced by light and temperature (Yoshimura et al. 2000; Kim et al. 2007). CAT has a high affinity for hydrogen peroxide in plant cells. It regulates the plant stress response mainly by removing hydrogen peroxide generated during mitochondrial electron transfer and fatty acid oxidation (Jiang et al. 2002; Xing et al. 2007).

We also found that the DEGs involved in the metabolism of glutathione (GSH) and related to the plant stress response were also upregulated in wild PT, including GPX, GST, glutathione reductase (GR), and aminopeptidase N (pepN).

GSH is one of the main sulfur reducers in living organisms and can be resistant to stress, such as cold damage and drought (Grill 1978; Schupp and Rennenberg, 1988). It combines the electrophilic groups of certain harmful substances with the thiol group of reduced glutathione to form more soluble and nontoxic derivatives of these substances (Coleman 1997). Increased GST activity protects plants from stresses, for example, from organic pollutants and heavy metals (Kirchhoff and Pflugmacher et al. 2000; Pflugmacher et al. 2007). GR deoxidizes oxidized glutathione disulfide to reduced glutathione, which helps to eliminate reactive oxygen species (ROS) and participates in the plant ascorbic acid GSH cycle. Therefore, wild PT may suffer more environmental stress, making its AsA biosynthesis and metabolic activity significantly stronger than those of cultivated PT.

bglX can hydrolyze β -glucoside and cellulose to glucose together with other cellulases; SPS and SPP recombine degraded sucrose for transport, storage and metabolism. The catalytic effect of SPS is irreversible. Huber (1983) pointed out that SPS activity is negatively correlated with starch accumulation but directly proportional to sucrose formation. Inv hydrolyzes sucrose into fructose and glucose, and fructose needs to be phosphorylated by ScrK or hexokinase before further metabolism. ScrK plays a major role in the metabolism and distribution of library tissues (David 2007), and it can be used as a hexose sensor and signal molecule to regulate the metabolism and growth of plants (Rolland et al. 2006). Trehalose is the most stable sugar of natural disaccharides catalyzed by TPS and TPP. Studies have shown that trehalose could enhance the resistance of plant cells to various abiotic stresses (Elbein et al. 2003). TPP has little effect on trehalose accumulation (Blázquez et al. 1998); thus, TPS may have a decisive effect on trehalose synthesis. It can be concluded that starch and sucrose metabolism in PT from wild fields was more active than in PT from cultivated fields.

Following starch and sucrose metabolism, genes become involved in glycolysis and the TCA cycle. ALDO catalyzes the conversion of fructose 1,6-diphosphate to glyceraldehyde 3-phosphate and dihydroxyacetone phosphate during glycolysis. This reaction is reversible. GOLS and PK are upregulated genes in cultivated PT that generate 1,3-bisphosphoglycerate and pyruvate during first and second substrate level phosphorylation, respectively. PK is one of the most critical rate-limiting enzymes in glycolysis. MDH catalyzes the dehydrogenation reaction of malic acid to oxaloacetate, which is accompanied by the reduction of NAD to NADH that may improve the TCA cycle (Chang et al. 2013). ACO is an important iron-thioproteinase in plant cells. In the TCA cycle, ACO catalyzes the production of isocitrate from citric acid in the cell via the intermediate product aconitate. This indicated that the TCA cycle was stronger, while glycolysis was weaker in wild PT than in cultivated PT.

Amino acid metabolism

Amino acids are the basic units that make up proteins and the precursors for many metabolites that have multiple functions in plant growth and stress responses. Thirteen DEGs were involved in amino acid metabolism, including pyrroline-5-carboxylate reductase (P5CR), acetyl-CoA acyltransferase 1 (ACAA1), GDA, peroxidase (POX), amino-acid N-acetyltransferase (argAB), 1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase (ADI1), GS, hydroxypyruvate reductase 1 (HPR1), glycine dehydrogenase (GDC), polyamine oxidase (PAO), primary amine oxidase (AOC), 3-methylcrotonyl-CoA carboxylase alpha subunit (MCCA), and L-pipecolate oxidase (PIPOX). Among these DEGs, 12 were upregulated and 1 was downregulated in PT from wild fields.

GS and GAD are two vital enzymes in the amino acid metabolism pathway. GS reduces the toxicity of excessive ammonium ions to plants by catalyzing the synthesis of glutamine. Glutamine is the main storage and transport form of ammonia. NH_4^+ participates in the process of glutamate formation through the glutamine synthetase/glutamic acid synthase cycle. As an osmotic regulator, proline is formed through the glutamic acid pathway and ornithine pathway and could improve the salt tolerance of plants. Nancy's (Roosens et al. 1999) study proved that GS could promote the synthesis of proline. Glutamate-semialdehyde (GSA) synthesizes proline through the glutamate pathway under the catalysis of P5CR (Rocha et al. 2012). Many studies have shown that the overexpression of the P5CR gene can improve plant resistance to drought. For example, under salt stress, P5CR-overexpressing sweet potatoes have increased proline contents and enhanced salt tolerance via the regulation of osmotic pressure (Gut et al. 2009). GAD is a pyridoxal phosphate-dependent enzyme that catalyzes the decarboxylation of glutamic acid to γ -aminobutyric acid. GAD activity is mainly regulated by the combination of pH and calcium ion/calmodulin (CAM) (Gut et al. 2009). When exposed to environmental stimulation (hypoxia, low temperature, acid stimulation, high salt penetration, etc.), H^+ and Ca^{2+} in the plant will increase significantly and promote GAD activity.

MCCA and carbamoyl phosphate synthetase belong to the same enzyme family that could convert the leucine intermediate metabolite 3-methylcrotonyl-CoA to 3-methylpentadienoyl-CoA (Wurtele and Nikolau 2000). In *Arabidopsis thaliana*, the gene numbered AT1G03090 encodes MCCA, a subunit of methylcrotonyl-CoA carboxylase, which is biotinylated to participate in the degradation of leucine and further affects plant development (Ding et al. 2012). In plants containing purine alkaloids, such as cocoa and coffee, MCCA is essential in the purine alkaloid synthesis pathway by participating in the synthesis of the precursors of these substances (Zrenner et al. 2006; Ashihara et al. 2008).

During serine metabolism, hydroxypyruvate is reduced to glyceric acid by HPR and returned to the chloroplast, and glycerol is catalyzed to the final product 3-phosphoglycerate by glycerol 3 kinase (Bauwe et al. 2010).

Polyamines (PAs) in plants mainly include putrescine (Put), sub-spermidine (Spd), and spermine (Spm). PAO can oxidatively degrade polyamines and reduce excess free PAs. Its diverse degradation products are also vital for plant growth and development and stress adaptation (An et al. 2008). In summary, amino acid metabolism in PT from wild fields was more active than in PT from cultivated fields.

Lipid metabolism

In cultivated PT, there were 2 upregulated and 8 downregulated DEGs associated with lipid metabolism: allene oxide cyclase (AOC), acyl-coenzyme A thioesterase 1/2/4 (ACOT), 2-acylglycerol O-acyltransferase 2 (MGAT2), lysophospholipid acyltransferase (LPLATs), secretory phospholipase A2 (PLA2), phosphoethanolamine N-methyltransferase (PEAMT), glycerophosphoryl diester phosphodiesterase (GDPD), enoyl-[acyl-carrier protein] reductase I (FabI), very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase (HACD) and long-chain acyl-CoA synthetase (ACSLs).

PLA2 catalyzes the hydrolysis of the acyl bond at the 2-position of glycerol phosphate to generate free fatty acids and lysates, which play an important role in cell signal transduction, biosynthesis of bioactive lipids, and the release of linolenic acid from phospholipids. GDPD catalyzes the hydrolysis of glycerol phosphodiesters into an alcohol molecule and glycerol triphosphate. Its products are involved in many biochemical pathways. Alpha-linolenic acid is the substrate of a synthetic pathway of jasmonic acid. AOC synthesizes 12-oxo-plant dienoic acid (OPDA) in this pathway. OPDA then undergoes reduction and oxidation to synthesize jasmonate. Jasmonic acid can induce tuber formation during plant growth and development (McConn 1996). Triacylglycerol TAG is a storage lipid and an important carbon source to support seedling development after seed germination. DGAT catalyzes the last step of the TAG synthesis pathway and is the only rate-limiting enzyme in this pathway (Settlage et al. 1998). MGAT2 is a membrane-bound acyltransferase that belongs to the diacylglycerol acyltransferase (DGAT) gene family. ACOT hydrolyzes CoA into free fatty acids (FFA) and coenzyme A (CoASH) and participates in essential physiological processes by maintaining appropriate levels of intracellular lactoyl CoA, FFA and CoASH (HuntillM and Alexson 2002). As a methyltransferase, PEAMT is a rate-limiting enzyme that catalyzes the methylation of phosphoethanolamine to choline. Recent studies have shown that PEAMT not only plays a role in the process of plant growth and development (Fulgencio et al. 2012; Mouet al. 2002) but also participates in the synthesis of the osmolyte betaine and stress-related second messenger phosphatidic acid to aid in the plant response to salt stress (Kosová et al. 2013; Huang 2000). ACSLs participate in the first reaction step of long-chain fatty acid degradation in various organisms (Fulda et al. 2002). It can be concluded that lipid metabolism in wild PT was stronger than in cultivated PT.

Biosynthesis of other secondary metabolites

As an abundant and widely studied effective substance in PT, nucleoside alkaloids are considered to be more suitable for quality control than succinic acid (Lua et al. 2011) which makes an interesting study topic in future research on the differences between wild and cultivated PT an interesting target of study in future PT research. We found increased expression of 2 DEGs related to nucleoside alkaloids in wild accessions: 5'-NT and adenosine deaminase (ADA). Various nucleotides in plants are catalyzed to nucleosides through 5'-NT. ADA, a thiolase, catalyzes the conversion of adenine nucleoside to hypoxanthine nucleoside and then generates hypoxanthine through the action of nucleoside phosphorylase, which plays a vital role in purine nucleoside metabolism. This may indicate that nucleoside alkaloid biosynthesis was stronger in wild PT than in cultivated PT.

Vitamin E is a kind of fat-soluble vitamin necessary for the human body. It is an important effector substance on the cell membrane of plants that can remove singlet oxygen and inactivate reactive oxygen radicals under oxidative stress (Valentin et al. 2005). The tocopherol O-methyltransferase (γ -TMT) and tocopherol cyclase (TC) genes were upregulated in this study, and are responsible for the synthesis of vitamin E. γ -TMT is the last step in the biosynthetic pathway of vitamin E that catalyzes the methylation of γ and δ -tocopherol to α and β -tocopherol. TC catalyzes different intermediates in tocopherol synthesis to form aromatic rings. Kumar R's (2005) study found that overexpression of the TC gene in rapeseed could increase the total

tocopherol levels and change the proportion of each component of tocopherol. It can be concluded that the biosynthesis of other SMs in wild PT was stronger than that in cultivated PT.

Conclusion

At present, studies of *Pinellia ternata* (PT) are mainly focused on the mechanism of its toxicity. However, since PT is a widely distributed species with an ancient history of human interaction, there is a growing need for the genetic characterization of PT accessions to understand the genetic basis controlling its various nutraceutical traits. However, the genetic study of the pharmacodynamic components of PT is limited.

In this work, we sequenced tuber transcriptomes of 8 different wild and cultivated plant accessions of PT with the RNA-Seq method, generating an important genomic resource for understanding the native biodiversity of the herb. In particular, the identification of sequence and expression polymorphisms of transcripts related to secondary metabolite metabolism resulted in a total of 114 979 080 paired end reads, which were assembled into unigenes with an average length of 861 25 bp. Further analysis of 79 genes related to metabolic pathways that were significantly differentially expressed allowed the exploration of the molecular mechanisms of PT quality formation under different growing conditions and the quality control of PT. RT-qPCR and HPLC determination results show that RNA-Seq is a reliable method for transcriptome research of PT. The data reported here will help further genomic projects to expand our knowledge of this valuable TCM.

Abbreviations

TCM: Traditional Chinese Medicine; PT: *Pinellia ternata*; RNA-seq: RNA sequencing; SMs: secondary metabolites; DEGs: differentially expressed genes; RIN: RNA integrity number; NR: NCBI non-redundant; KOG: eukaryotic complete genomes; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; BP: biological process; MF: molecular function; CC: cellular component; β Glu: β -glucosidase; FFase: β -fructofuranosidase; SPS: sucrose-phosphate synthase; GDA: glutamate decarboxylase; GS: glutamate decarboxylase; 5'-NT: 5'-nucleotidase; COMT: caffeic acid 3-O-methyltransferase; GST: glutathione S-transferase; ScrK: fructokinase; GGP: GDP-L-galactose phosphorylase; MDH: malate dehydrogenase; ALDO: fructose-bisphosphate aldolase; ACO: aconitate hydratase; APX: L-ascorbate peroxidase; β glX: beta-glucosidase; TPS: trehalose 6-phosphate synthase; Inv: beta-fructofuranosidase; GMPP: mannose-1-phosphate guanylyltransferase; CAT: catalase; GLYR: succinic semialdehyde reductase; GAPDH: glyceraldehyde 3-phosphate dehydrogenase; PK: pyruvate kinase; β glC: glucose-1-phosphate adenylyltransferase; β glB: 4- α -glucan branching enzyme; GOLS: inositol 3- α -galactosyltransferase; AsA: Ascorbic acid; GSH: glutathione; GR: glutathione reductase; pepN: aminopeptidase N; ROS: reactive oxygen species; P5CR: pyrroline-5-carboxylate reductase; ACAA1: acetyl-CoA acyltransferase 1; POX: peroxidase; argAB: amino-acid N-acetyltransferase; ADI1: 1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase; HPR1: hydroxypyruvate reductase 1; GDC: glycine dehydrogenase; PAO: polyamine oxidase; AOC: primary amine oxidase; MCCA: 3-methylcrotonyl-CoA carboxylase α subunit; PIPOX: L-pipecolate oxidase; GSA: Glutamate-semialdehyde; CAM: calcium ion/calmodulin; PAs: Polyamines; Put: putrescine; Spd: sub-spermidine; Spm: spermine; AOC: allene oxide cyclase; ACOT: acyl-coenzyme A thioesterase 1/2/4; MGAT2: 2-acylglycerol O-acyltransferase 2; LPLATs: lysophospholipid acyltransferase; PLA2: secretory phospholipase A2; PEAMT: phosphoethanolamine N-methyltransferase; GDDP: glycerophosphoryl diester phosphodiesterase; FaBI: enoyl-[acyl-carrier protein] reductase I; HACD: very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase; ACSLS: long-chain acyl-CoA synthetase; OPDA: 12-oxo-plant dienoic acid; DGAT: diacylglycerol acyltransferase; FFA: free fatty acids; CoASH: coenzyme A; ADA: adenosine deaminase; γ -TMT: tocopherol O-methyltransferase; TC: tocopherol cyclase

Declarations

Acknowledgements

Not applicable.

Funding

This work was supported by grants 81774096 from the National Natural Science Foundation of China, ZYX03KF031 from the Open Projects of the Discipline of Chinese Medicine of Nanjing University of Chinese Medicine supported by the Subject of Academic Priority Discipline of Jiangsu Higher Education Institutions, QNRC2016642 and LGY2018061 from the Jiangsu Provincial Medical Youth Talent, QNRC2-B04 from the Young Elite Scientists Sponsorship Program by CAST and WSN-051 from the Top Six Talent Project of Jiangsu Province 2016.

Availability of data and materials

The dataset supporting the conclusions of this article is available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under accession No. PRJNA623739.

Authors' contributions

SJL and JW conceived and designed research. JW and CJ conducted experiments. YX, YFZ and JZ contributed new reagents or analytical tools. JW, XYL and JWL analyzed data. JW wrote the manuscript. All authors read and approved the manuscript.

Author's information

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Xie HH, et al. Effects of pinellia ternata (thunb.) berit. on the metabolomic profiles of placenta and amniotic fluid in pregnant rats. *Journal of Ethnopharmacology* 2016;183(13):38-45.
2. Li Y, et al. A polysaccharide from pinellia ternata inhibits cell proliferation and metastasis in human cholangiocarcinoma cells by targeting of cdc42 and 67kda laminin receptor (lr). *International Journal of Biological Macromolecules* 2016;93(A):520-525.
3. Zhang ZH, et al. General toxicity of pinellia ternata (thunb.) berit. in rat:a metabonomic method for profiling of serum metabolic changes. *Journal of Ethnopharmacology* 2013;149(1):303-310.
4. Wu XY, et al. Sedative, hypnotic and anticonvulsant activities of the ethanol fraction from rhizoma pinelliae praeparatum. *Journal of Ethnopharmacology* 2011;135(2):0-329.
5. Yang L, et al. Response of plant secondary metabolites to environmental factors. *Molecules* 2018;23(4):762-
6. Kliebenstein DJ, Osbourn A. Making new molecules—evolution of pathways for novel metabolites in plants. *Current opinion in plant biology* 2012;15(4):415-423.
7. Wu H, et al. Studies on stimulating components of raw pinellia ternata (thunb.)(banxia). *China journal of chinese materia medica* 1999;24(12):725-730.
8. Gonda R, et al. Characterization of an acidic polysaccharide with immunological activities from the tuber of pinellia ternata. *Biological & Pharmaceutical Bulletin* 1994;17(12):1549-1553.

9. Yang BY, et al. A system research on hplc characteristic fingerprint of pinelliae rhizoma and pinelliae rhizoma praeparatum cum zingibere et alumine. Chinese Traditional and Herbal Drugs 2014;45(5):652-658.
10. Mu ET, Zhou JL. Endangered Resources and Change of Origin of Pinellia ternate in Huanghuai Plain. China Journal of Chinese Materia Medica 2013;38(21):3795-3796.
11. Lee EJ, et al. Quality assessment of ginseng by (1) h nmr metabolite fingerprinting and profiling analysis. Journal of Agricultural & Food Chemistry. 2009;57(16):7513-22.
12. Lv X, et al. Integration of HPLC-based fingerprint and quantitative analyses for differentiating botanical species and geographical growing origins of Rhizoma coptidis. Pharmaceutical Biology 2016;54(12):1-8.
13. Xiao TB, Xiao X, Guo JM. Acting mechanism and research methods of long-chain non-coding RNA. Hereditas 2013;35(3):269-280
14. Marioni JC, et al. RNA-seq:an assessment of technical reproducibility and comparison with gene expression arrays. GenomeRes 2008;18(9):1509-1517
15. Birol I, et al. De novo transcriptome assembly with ABySS. Bioinformatics 2009;25(21):2872-2877
16. Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 2010;28(5):511–515
17. Chen S, et al. 454 EST analysis detects genes putativelyinvolved in ginsenoside biosynthesis in Panax ginseng. Plant Cell Rep. 2011;30(9):1593-1601.
18. Liu ZQ, et al. Transcriptome sequencing and analysis of the entomopathogenic fungusHirsutella sinensis isolated from Ophiocordyceps sinensis. BMC Genomics. 2015;16(1):106-122
19. Fukushima. A, et al. Comparative characterization of the leaf tissue of physalis alkekengi and physalis peruviana using ma-seq and metabolite profiling. Frontiers in Plant Science 2016;7(598):1883.
20. Ma S, Wang L, Ye D, Xu T. Transcriptome analysis of Pinellia ternata mutant bulbs based on high-throughput sequencing. Journal of Zhejiang Agricultural Science 2018;59(10):107-112.
21. Zhu Y, et al. A comparative proteomic analysis of pinellia ternata leaves exposed to heat stress. International Journal of Molecular Sciences 2013;14(10):20614-20634.
22. Linster CL, Clarke SG. l-Ascorbate biosynthesis in higher plants:the role of VTC2. Trends in Plant Science 2008;13(11):567-573.
23. Yoshimura K, et al. Expression of spinach ascorbate peroxidase isoenzymes in response to oxidative stresses. Plant Physiol. 2000;123(1):223-233.
24. Kim DW, et al. Gene transcription in the leaves of rice undergoing salt-induced morphological changes (Oryza sativa L.). Mol Cells. 2007;24(1):45-59.
25. Jiang M, Zhang J. Involvement of plasma-membrane NADPH oxidase in abscisic acid-and water stress-induced antioxidant defense inleaves of maize seedlings. Planta 2002;215:1022-1030.
26. Xing Y, Jia W, Zhang J. AtMEK1 mediates stress-induced gene expression of CAT1 catalase by triggering H2O2production inAra-bidopsis.J Exp Bot. 2007;58(11):2969-2981.
27. Grill ED. Seasonal Variation of Glutathione and Glutathione Reductase in Needles of Picea abies. Plant Physiology. 1978;61(1):119-121.
28. Schupp R, Rennenberg H. Diurnal changes in the glutathione content of spruce needles (picea abies l.). Plant Science. 1988;57(2):113-117.
29. Coleman JOD. Detoxification of xenobiotics by plants :chemical modification and vacuolar compartmentation. Trends in Plant Science. 1997;2(4):144-151.
30. Kirchhoff A, Pflugmacher S. Comparison of the detoxication capacity of the limnic and marine form of the green algae Enteromorpha compressa. Marine Environmental Research 2000;50(1-5):72-73.
31. Pflugmacher S, Olin M, Kankaanpää H. Nodularin induces oxidative stress in the Baltic Sea brown alga Fucus vesiculosus (Phaeophyceae). Marine Environmental Research 2007;64(2):149-159.

32. Huber SC. Role of Sucrose-Phosphate Synthase in Partitioning of Carbon in Leaves. *Plant physiology* 1983;71:818-821.
33. David G. Role of tomato hexose kinases. *Functional Plant Biology* 2007;34(6):564-570.
34. Rolland F, Baena-Gonzalez E, Sheen J. Sugar sensing and signaling in plants: conserved and novel mechanisms. *Annual review of plant biology* 2006;57(0):675-709.
35. Elbein AD. New insights on trehalose: a multifunctional molecule. *Glycobiology* 2003;13(4):17-27.
36. Blázquez MA, et al. Isolation and molecular characterization of the Arabidopsis TPS1 gene, encoding trehalose-6-phosphate synthase. *The Plant Journal* 1998;13(5):685-689.
37. Chang YY, et al. Cloning, overexpression, purification and crystallization of malate dehydrogenase from *Thermus thermophilus*. *Acta Crystallogr. Sect. F: Struct. Biol. Cryst. Commun.* 2013;69(11):1249-1251.
38. Roosens, NH, et al. Biesemans M., Jacobs M.. Proline metabolism in the wild-type and in a salt-tolerant mutant of *Nicotiana glauca* studied by ¹³C-nuclear magnetic resonance imaging. *Plant physiology* 1999;121(4):1281-1290.
39. da Rocha IM, et al. Exogenous ornithine is an effective precursor and the δ -ornithine amino transferase pathway contributes to proline accumulation under high N recycling in salt-stressed cashew leaves. *Journal of Plant Physiology* 2012;169(1):0-49.
40. Gut H, et al. A common structural basis for pH- and calmodulin-mediated regulation in plant glutamate decarboxylase. *Journal of Molecular Biology* 2009;392(2):334-351.
41. Wurtele ES, Nikolau BJ. Characterization of 3-methylcrotonyl-CoA carboxylase from plants. *Methods in Enzymology* 2000;324:280-292.
42. Ding G, et al. Genetic dissection of methylcrotonyl CoA carboxylase indicates a complex role for mitochondrial leucine catabolism during seed development and germination. *The Plant Journal: for cell and molecular biology* 2012;70(4):562-577.
43. Zrenner R., Stitt M., Sonnewald U., Boldt R.. Pyrimidine and purine biosynthesis and degradation in plants. *Annual Review of Plant Biology* 2006;57(1). 805-836
44. Ashihara H, Sano H, Crozier A. Caffeine and related purine alkaloids: biosynthesis, catabolism, function and genetic engineering. *Phytochemistry* 2008;39(22):841-856
45. Bauwe H, Hagemann M, Fernie AR. Photorespiration: players, partners and origin. *Trends Plant Sci.* 2010;15(6):330-336
46. An ZF, et al. Hydrogen peroxide generated by copper amine oxidase is involved in abscisic acid-induced stomatal closure in *Vicia faba*. *J Exp Bot.* 2008;59(4):815-825
47. McConn M. The Critical Requirement for Linolenic Acid Is Pollen Development, Not Photosynthesis, in an Arabidopsis Mutant. *The plant cell online* 1996;8(3):403-416.
48. Settlage SB, Kwanyuen P, Wilson RF. Relation between diacylglycerol acyltransferase activity and oil concentration in soybean. *Journal of the American Oil Chemists' Society* 1998;75(7):775-781.
49. Hunt LM C, Alexson SEH. The role acyl-CoA thioesterases play in mediating intracellular lipid metabolism. *Progress in Lipid Research* 2002;41(2):99-130
50. Fulgencio AC, et al. Translational regulation of Arabidopsis XIPOTL1 is modulated by phosphocholine levels via the phylogenetically conserved upstream open reading frame 30. *Journal of Experimental Botany* 2012;63(14):5203-5221.
51. Mou ZL, et al. Silencing of Phosphoethanolamine N-Methyltransferase Results in Temperature-Sensitive Male Sterility and Salt Hypersensitivity in Arabidopsis. *Plant Cell* 2002;14(9):2031-2043.
52. Kosová K, Práil IT, Vítámvás P. Protein Contribution to Plant Salinity Response and Tolerance Acquisition. *International Journal of Molecular Sciences* 2013;14(4):6757-6789.
53. Huang J. Genetic Engineering of Glycinebetaine Production toward Enhancing Stress Tolerance in Plants: Metabolic Limitations. *Plant physiology* 2000(3);122:747-756.
54. Fulda M, et al. Two long-chain acyl-CoA synthetases from Arabidopsis thaliana involved in peroxisomal fatty acid β -oxidation. *Plant J* 2002;32(1):93-103.

55. Lu D, et al. Determination of nucleosides in Rhizoma Pinelliae high performance liquid chromatography. Chinese journal of chromatography 2011;29(1):83-86.

56. Henry E, V, Qi QG. Biotechnological production and application of vitamin E:current state and prospects. Applied Microbiology and Biotechnology 2005;68(4):436-444.

57. Kumar R, et al. Characterisation of plant tocopherol cyclases and their overexpression in transgenic brassica napus seeds. Febs Letters 2005;579(6):0-1364.

Tables

Table 1 Samples of PT.

Origins	NO.	Group	Sorts
Tianshui City, Gansu Provice	A1, 2, 3	S1	cultivated
Weinan City, Gansu Provice	B1, 2, 3	S2	cultivated
Xihe City, Gansu Provice	C1, 2, 3	S3	cultivated
Dafang City, Guizhou Provice	D1, 2, 3	S4	cultivated
Anguo City, Hebei Provice	E1, 2, 3	S5	cultivated
Shijiazhuang City, Hebei Province	F1, 2, 3	S6	cultivated
Tongliang City, Chongqing Province	I1, 2, 3	S7	wild
Chengdu, Sichuan Province	H1, 2, 3	CON	wild

Table 2 Statistics of transcriptome sequencing.

Sample	raw_reads	raw_bases	clean_reads	clean_bases	valid_bases	Q30	GC
A1	49233876	7.38E+09	46848848	6.75E+09	91.36%	92.76%	50.72%
A2	50832238	7.62E+09	48978380	7.13E+09	93.56%	94.12%	50.15%
A3	53541196	8.03E+09	51844206	7.55E+09	93.98%	94.50%	49.98%
B1	48922418	7.34E+09	46497054	6.68E+09	90.97%	92.70%	51.13%
B2	58643440	8.8E+09	56129076	8.17E+09	92.90%	93.82%	50.40%
B3	57638038	8.65E+09	55660040	8.14E+09	94.12%	94.22%	49.73%
C1	49588850	7.44E+09	47287858	6.81E+09	91.58%	92.92%	52.12%
C2	53671546	8.05E+09	51548918	7.49E+09	93.08%	93.69%	51.42%
C3	57761362	8.66E+09	55796528	8.1E+09	93.54%	94.04%	51.56%
D1	49865966	7.48E+09	47286322	6.80E+09	90.95%	92.41%	49.32%
D2	53228552	7.98E+09	51483322	7.51E+09	94.11%	94.31%	49.63%
D3	53760272	8.06E+09	52059784	7.59E+09	94.12%	94.41%	50.79%
E1	49784312	7.47E+09	46821618	6.77E+09	90.61%	92.19%	51.69%
E2	56222680	8.43E+09	54409736	7.95E+09	94.31%	94.35%	50.52%
E3	57085156	8.56E+09	54283848	7.86E+09	91.77%	93.56%	50.43%
F1	49160062	7.37E+09	46649064	6.73E+09	91.27%	92.60%	51.23%
F2	55864836	8.38E+09	53327870	7.57E+09	90.37%	93.99%	50.16%
F3	58802194	8.82E+09	55921586	8.06E+09	91.38%	93.58%	50.39%
G1	49123410	7.37E+09	46873272	6.74E+09	91.54%	93.19%	51.06%
G2	52035850	7.81E+09	50185334	7.34E+09	94.08%	94.08%	50.65%
G3	58082460	8.71E+09	56524354	8.25E+09	94.71%	94.88%	50.51%
H1	49452664	7.42E+09	47095608	6.74E+09	90.89%	93.04%	51.54%
H2	51179962	7.68E+09	49377576	7.23E+09	94.16%	94.08%	50.85%
H3	50107400	7.52E+09	48595258	7.06E+09	93.96%	94.51%	51.82%

Table 3 Summary of De novo assembly of PT.

Term	All(>300bp)	>=500bp	>=1000bp	N50	Total_Length	Max_Length	Min_Length	Average_Length
Unigene	133503	76894	36007	1162	1.15E+08	13803	301	861.25

Table 4 Summary of annotations of unigenes in public databases.

Anno_Database	Annotated_Number	300<=length<1000	length>=1000
NR	53323(39. 94 %)	25899(19. 40 %)	27424(20. 54 %)
Swissprot	36197(27. 11 %)	15282(11. 45 %)	20915(15. 67 %)
KEGG	18831(14. 11 %)	7927(5. 94 %)	10904(8. 17 %)
KOG	31981(23. 96 %)	14283(10. 70 %)	17698(13. 26 %)
eggNOG	48128(36. 05 %)	21960(16. 45 %)	26168(19. 60 %)
GO	33006(24. 72 %)	13749(10. 30 %)	19257(14. 42 %)
Pfam	123(0. 09 %)	106(0. 08 %)	17(0. 01 %)

Table 5 The content of adenine and guanosine in wild and cultivated PT.

No.	Origins	Type	The content of adenine(%)	The content of guanosine(%)
D1	Xihe City, Gansu Provice	cultivated	0.0096	0
D2	Xihe City, Gansu Provice		0.0077	0.0168
D3	Xihe City, Gansu Provice		0.0097	0.0180
D4	Xihe City, Gansu Provice		0.0105	0.0309
D5	Xihe City, Gansu Provice		0.0146	0.0309
D6	Anguo City, Hebei Provice		0.0225	0.0467
D7	Dafang City, Guizhou Provice		0.0148	0.0291
D8	Shijiazhuang City, Hebei Province		0.0173	0.0440
D9	Weinan City, Gansu Provice		0.0226	0.0297
D10	Tianshui City, Gansu Provice		0.0014	0.0267
D11	Tongliang City, Chongqing Province	wild	0.0132	0.0441
D12	Chengdu, Sichuan Province		0.0158	0.0255
D13	Chengdu, Sichuan Province		0.0239	0.0389

Figures

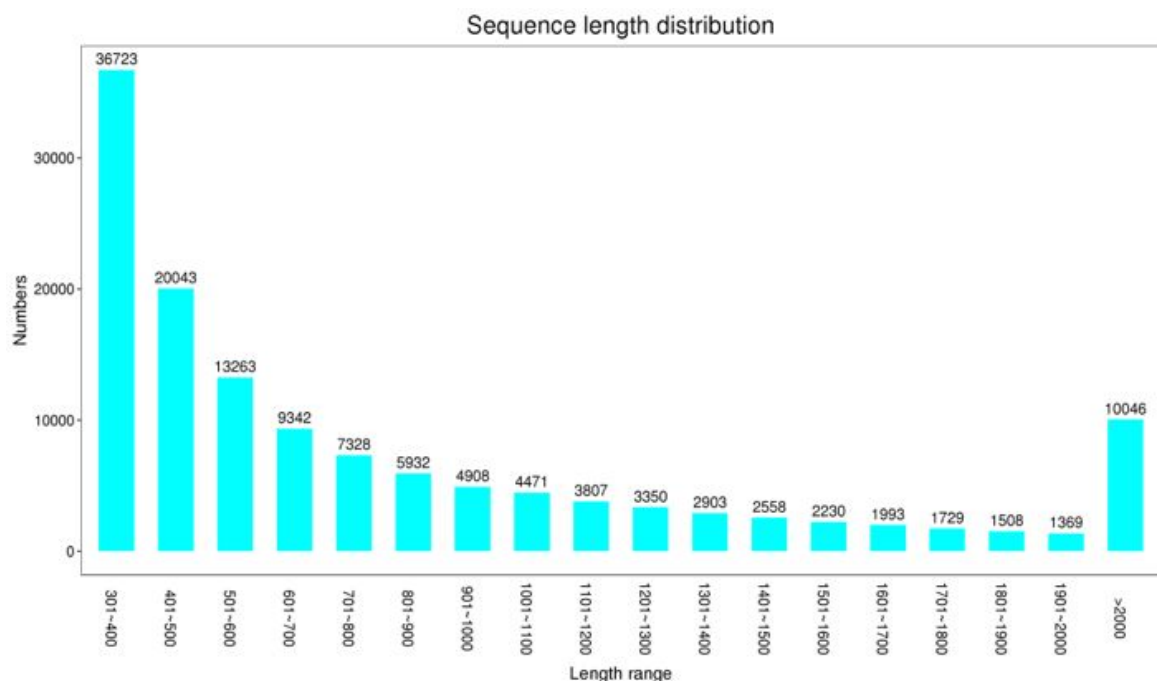


Figure 1

Distribution of unigenes.

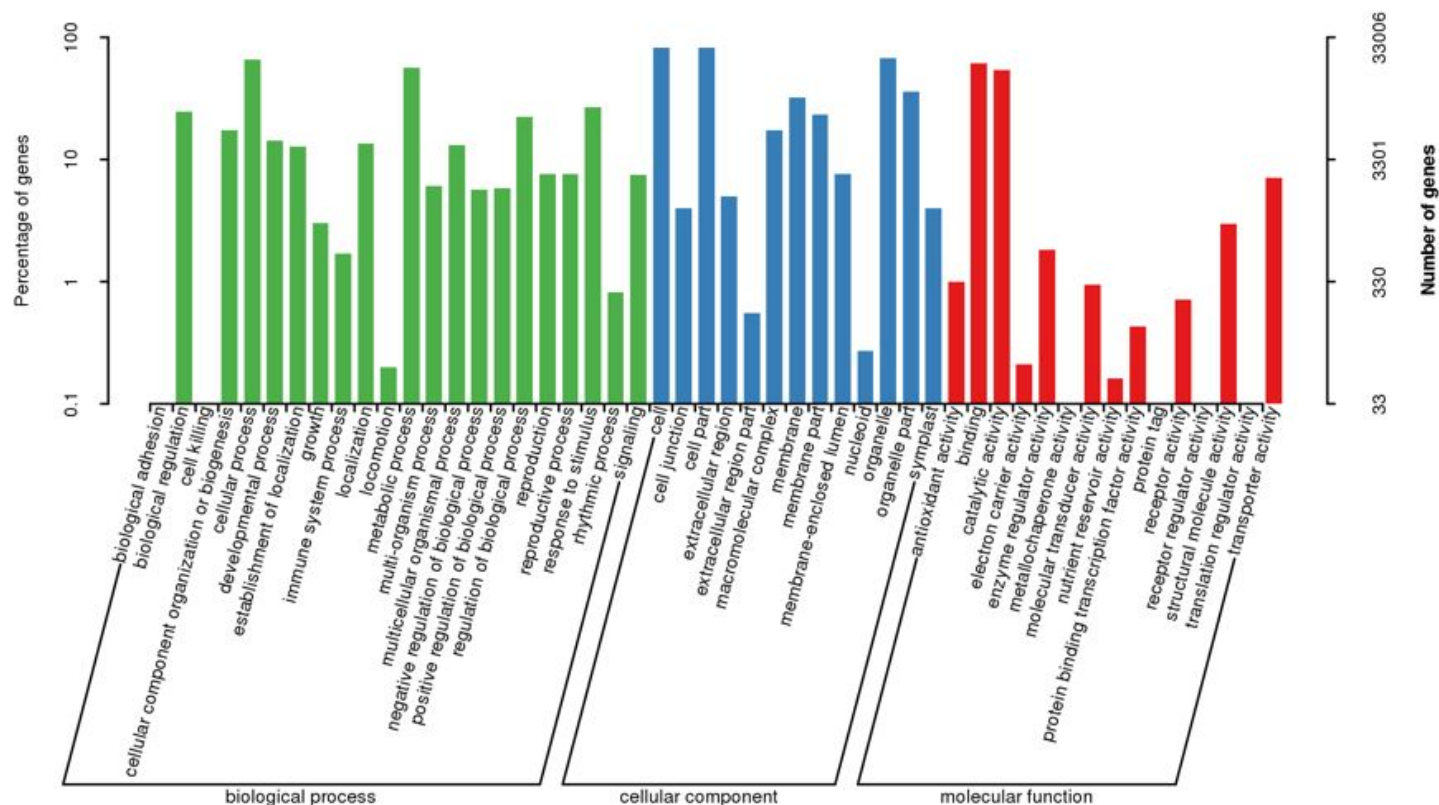


Figure 2

Frequency Distribution Ontological Classification. The results are grouped into cellular component, molecular function and biological process. The right y-axis indicates the number of genes in category. The left y-axis indicates the percentage of a

specific category of genes in the main category.

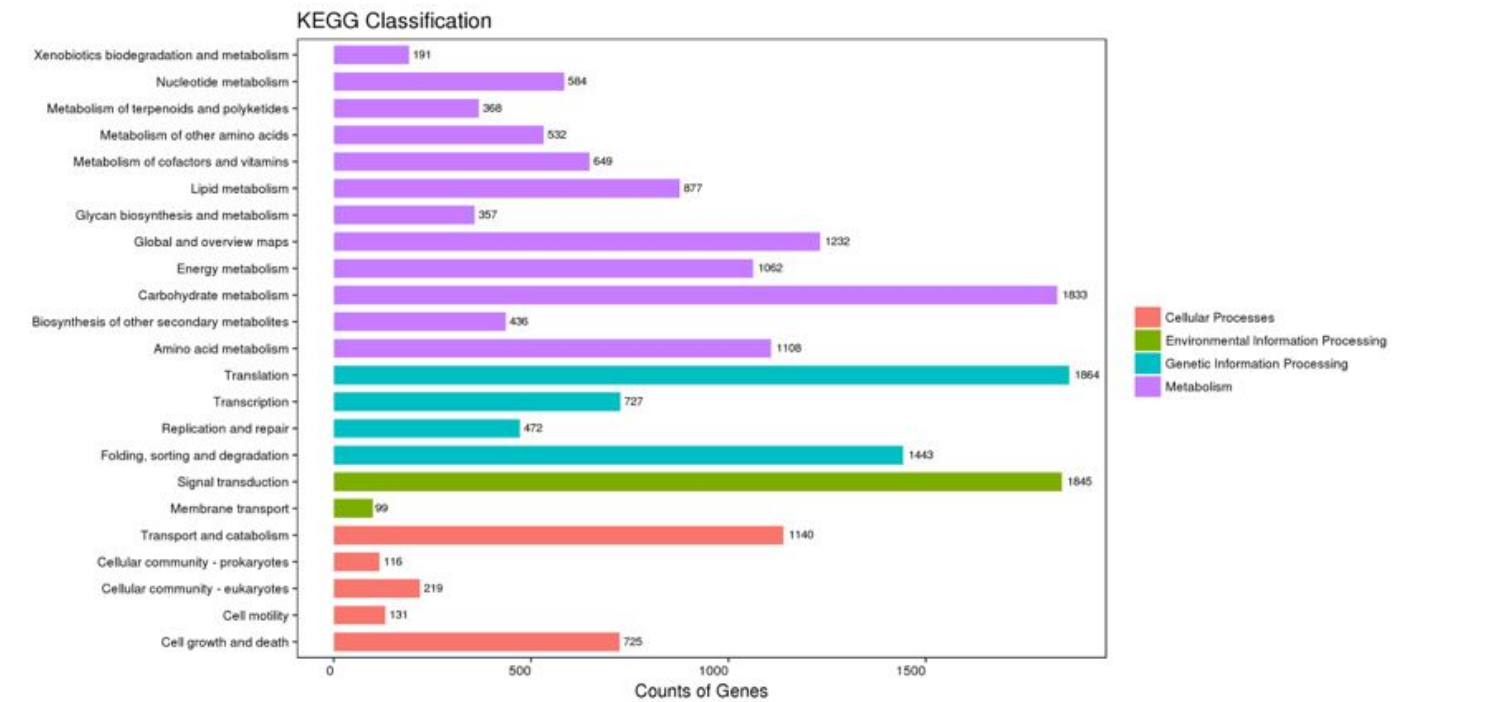


Figure 3

KEGG annotation of the non-redundant sequences of all the samples. The y-axis indicates the name of the KEGG metabolic pathway. The x-axis indicates the number of genes. The unigenes were divided into four branches according to the KEGG metabolic pathway: Cellular Processes, Environmental Information Processing, Genetic Information Processing, and Metabolism.

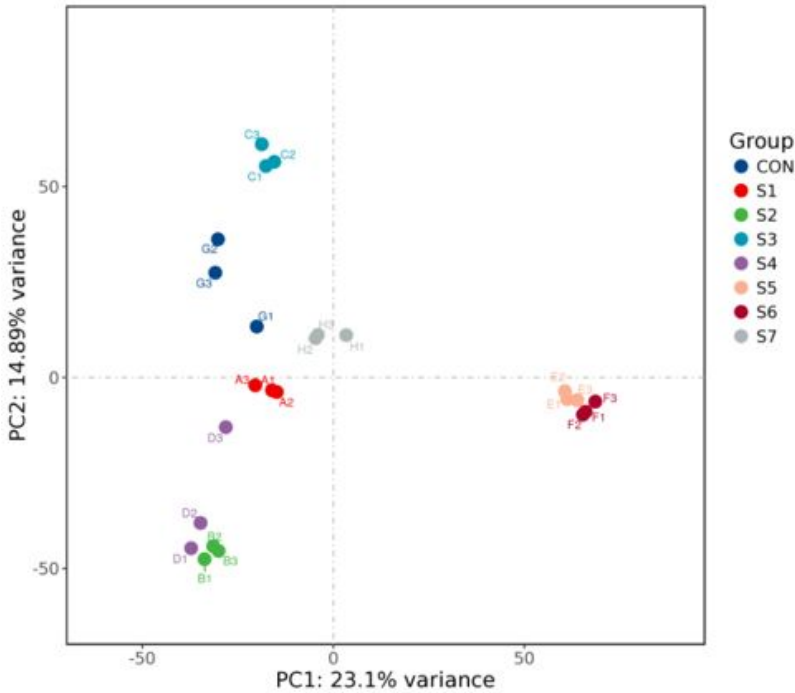


Figure 4

Principal component analysis of PT samples.

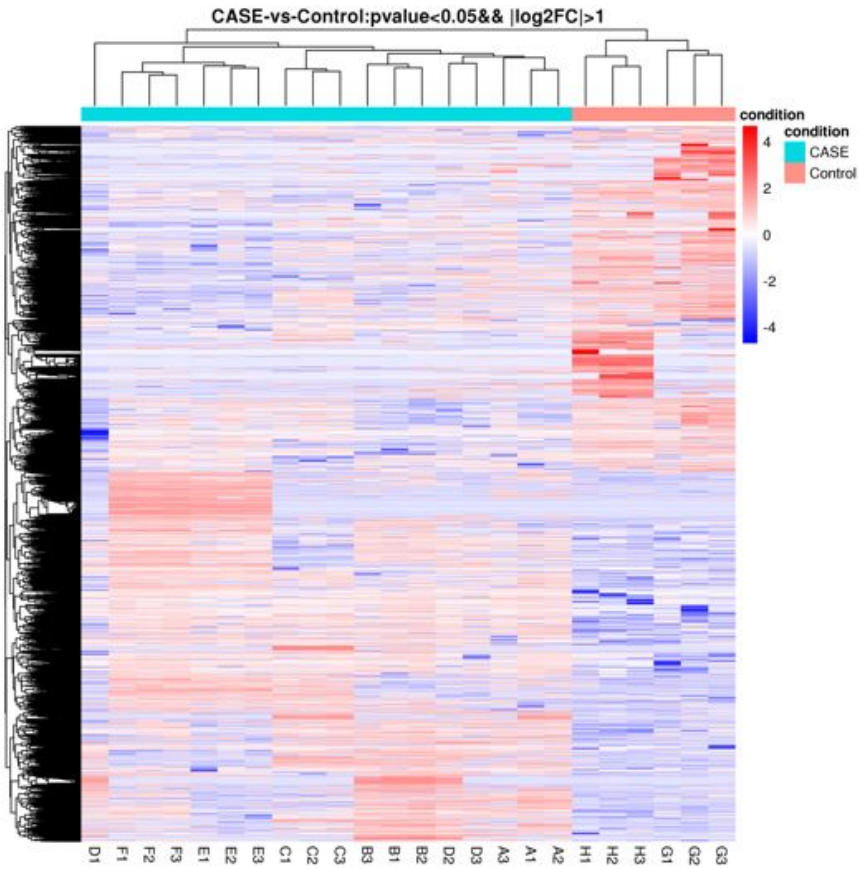


Figure 5

Heatmap showing differentially expressed transcripts among 24 Hierarchical clustering was performed with gene expression values through unsupervised hierarchical clustering analysis.

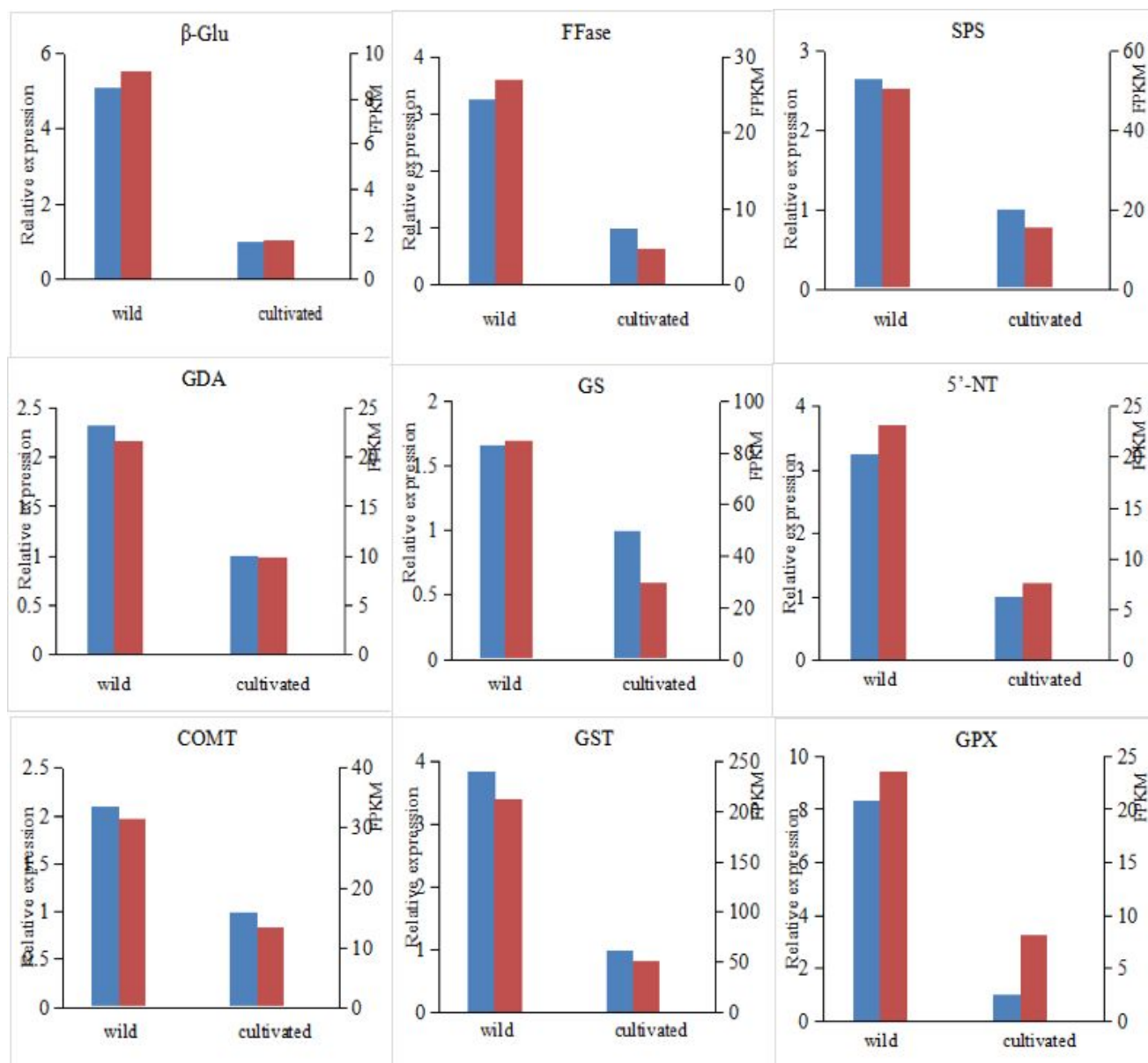


Figure 6

RT-qPCR verification of selected genes of RT. The blue bars represent the average relative expression determined with RT-qPCR, and the orange bars represent the level of expression (FPKM) of the transcripts.

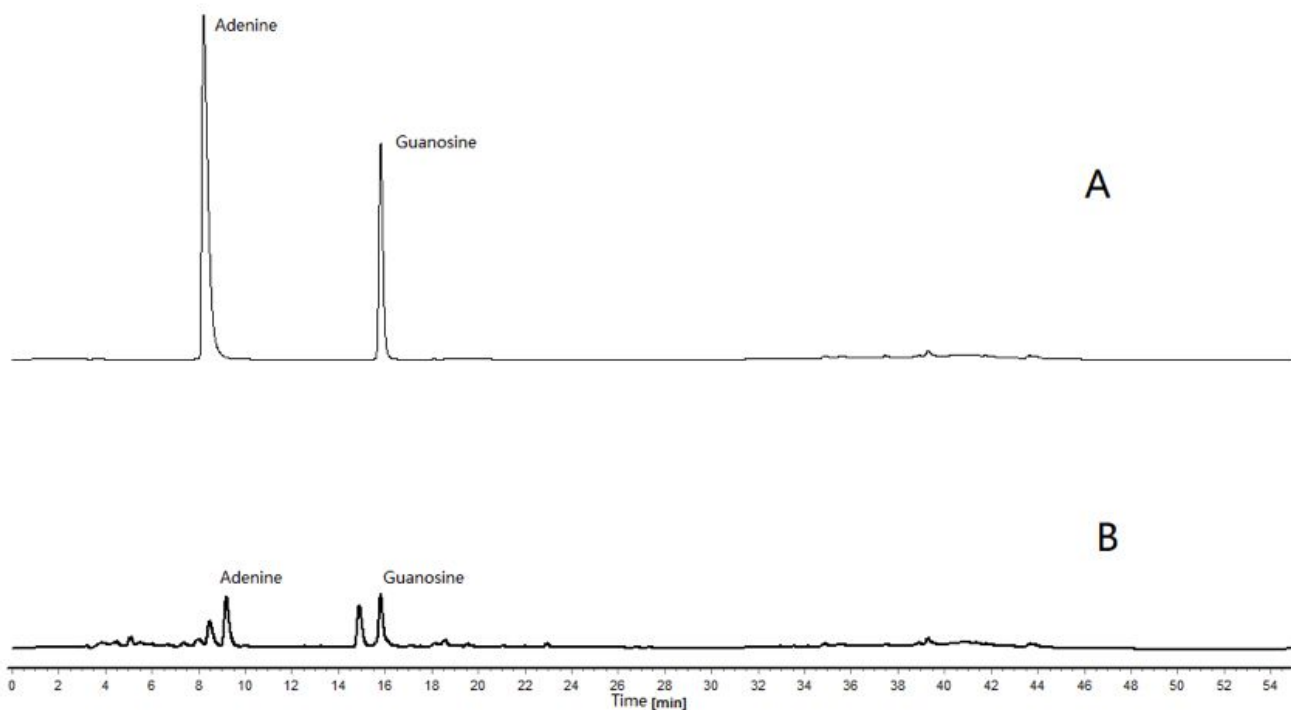


Figure 7

Chromatogram of adenine and guanosine standards (A) and a PT extract (B).

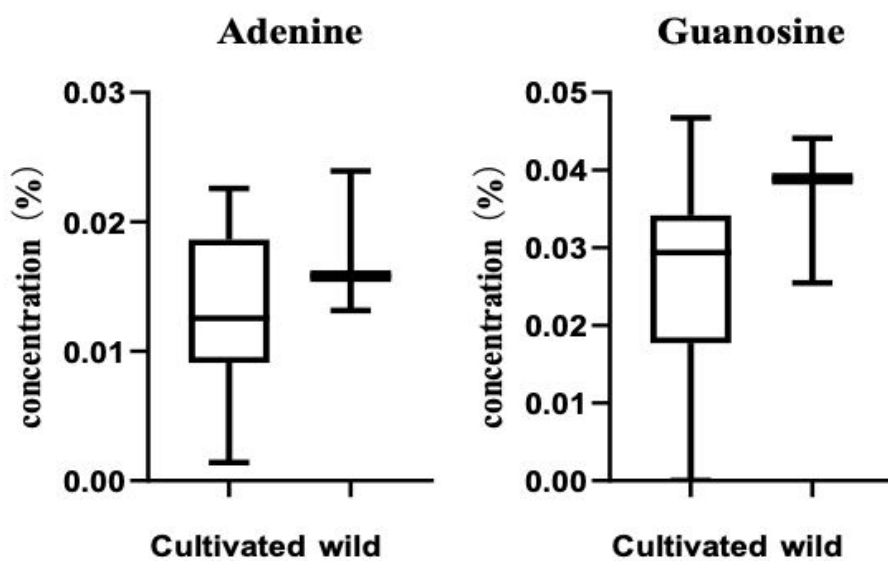


Figure 8

Box plot of the content of adenine and guanosine in wild and cultivated PT.

Supplementary Files

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

- [Additionalfile1.xlsx](#)
- [Additionalfile3.xlsx](#)

- [Additionalfile4.xlsx](#)
- [Additionalfile5.xlsx](#)
- [Additionalfile2.xlsx](#)