

Transcriptome Assembly and Gene Expression Analysis of Leaf and Rhizome tissues of *Dioscorea opposita*

Dianyun Hou

Henan University of science and Technology

Ruiyang Cheng

China Academy of Chinese Medical Sciences

Zhihong Li

Henan university of science and technology

Jingya Xu

Henan university of Science and Technology

Zhanqiang Ma

Henan university of science and technology

Xingli Zhao

Henan university of Science and Technology

Huawei Xu

Henan University of Science and Technology

Jiang Xu (✉ jxu@icmm.ac.cn)

China Academy of Chinese Medical Sciences

Xiaogai Hou

Henan University of Science and Technology

Research

Keywords: *Dioscorea opposita*, leaf and rhizome, transcriptome assembly, expression, terpene biosynthesis

Posted Date: May 19th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-28829/v1>

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

[Read Full License](#)

Abstract

Background: *Dioscorea opposita* is a monocotyledonous plant in the Dioscoreaceae family. The rhizome of *D. opposita*, known as “dioscoreae rhiaoma” or “Chinese yam”, is not only a well-known traditional Chinese medicine, but also a popular health food.

Methods:

In this study, the leaf and rhizome tissues of *D. opposita* were collected. Their transcriptome sequence and were obtained using an Illumina Hiseq 4000 platform. The data was analyzed by bioinformatics software. Likewise, the consistent expression pattern of DEGs was validated with qRT-PCR analysis.

Results:

A total of 46,982,031 and 53,689,620 clean reads from leaf and rhizome tissues were acquired, respectively (SRP: PRJNA628588). 50,765 unigenes were obtained with an average length of 1050 base pairs (bp). A total of 13,259 significant differentially expressed genes (DEGs) were obtained from leaf versus rhizome tissues, including 5536 up-regulated and 7723 down-regulated unigenes. In addition, 20,607 simple sequence repeats (SSRs) were identified in this study.

Conclusions:

This study not only enriched the transcriptomic data of *D. opposita*, but also laid a foundation for the analysis of terpenoid synthesis pathway.

Background

Dioscorea opposita Thunb (Family Dioscoreaceae), is an important staple in tropical regions and edible crop mainly distributed in parts of Africa, South America, and Asia [1–3]. The rhizome of *D. opposita*, known as “dioscoreae rhiaoma” or “Chinese yam”, has been listed as both a nutritious food source and an invigorant in traditional Chinese medicine for thousands of years [4–6]. Various types of natural constituents, including dopamine, starch, dioscin, polysaccharides, mucin, allantoin, choline, phytosterols, and essential amino acids, have been reported to be in *D. opposita* [7–9].

There were reports that 2-C-Methyl-D-Erythritol-4-Phosphate (MEP) and mevalonate acid (MVA) pathway are the early biosynthetic pathways of dioscin in many species [10–13]. It is well known that 1-deoxy-D-xylulose-5-phosphate synthase (DXS), 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR), geranylgeranyl pyrophosphate synthase (GGPS) and hydroxymethylglutaryl-CoA reductase (HMGR) are the key enzyme genes, related to the biosynthesis of dioscin[13]. However, there are no reports on key enzyme genes, molecular biology studies or comprehensive genomic and transcriptomic data of *D. opposita*.

RNA-Seq is an efficient method for transcriptome analysis, which can also provide a highly sensitive and accurate platform at a low cost and is efficient for gene discovery and biosynthesis pathway analysis in non-model plants without knowing their genome sequences [14–17]. RNA-Seq is currently a very popular method that is applied in the exploration of plants at transcriptomic levels. The ultimate aim is often to reveal genes that are involved in the biosynthetic pathways in secondary metabolism of many species [18–21]. Until now, the genomic resources of *D. alata* [22], *D. rotundata* [23], and *D. dumetorum* [24] have been reported. Nevertheless, as a non-model species, there is lack of genomic resources, in particular, information on active substance metabolism of *D. opposita*.

In the study, the sample of leaf and rhizome tissues of *D. opposita* were collected from Wenxian County, called “Tie Gui Shan Yao”. It is a local cultivar of *D. opposita*, belonging to one of the “four huaiqing Chinese medicines”. There have reported that the content of terpenoids in “Tie Gun Shan Yao” was higher than other *D. opposita* [25].

In this study, transcriptome analysis was performed from leaf and rhizome tissues of *D. opposita*. The results will enrich the known genetic information for further gene function research and analyses, such as differential transcript abundance analysis. We also try to understand the transcript abundance difference and explore the molecular mechanisms of biosynthetic pathways in secondary metabolites in *D. opposita*.

Materials And Methods

Plant materials and RNA extraction

Leaf and rhizome tissues of *D. opposita* were collected from Wenxian County, Jiaozuo City, Henan Province of China, on October 2017[Wenxian County Weather Station] N34°56'49.54", E113°10'18.54"; T:15°C]. Three biological replicates from three leaf tissues (HSYY_1, HSYY_2, and HSYY_3) and three rhizome tissues (HSYK_1, HSYK_2, and HSYK_3) were used for the analysis. All samples were flash frozen in liquid nitrogen after cleaning and stored at –80 °C until RNA extraction.

Total RNAs were extracted from the leaf and rhizome tissues using the TRIzol® Reagent (Invitrogen, USA) following the product description with slight improvement [26, 27]. The extracted RNAs were checked through NanoDrop™ 2000 spectrophotometer (NanoDrop Technologies, USA) and RNA integrity was assessed using electrophoresis in 1.0% agarose gel. Furthermore, the RNA concentration was determined by the RNA Agilent Bioanalyzer 2100 system (Agilent, CA, USA) to guarantee the RNA Integrity Number (RIN) values were not less than 7.0.

Transcriptome sequencing and de novo assembly

The cDNA libraries were constructed using the TruSeq-TM RNA sample preparation Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. The cDNA libraries were enriched through a polymerase chain reaction (PCR) with 15 cycles. Then, the target band was extracted and recycled by a

2% low range ultra-agarose gel, and quantified by TBS380 Mini-Fluorometer. The suitable cDNA libraries were sent to Shanghai Majorbio Biopharm Biotechnology Co., Ltd. (Shanghai, China) where the RNA-Seq was performed using the Illumina HiSeq4000 SBS Kit (300 cycles, Illumina, San Diego, CA, USA) on an Illumina HiSeq4000 platform.

The raw reads were cleaned and assessed following the steps. Firstly, the adaptor sequence of reads was removed, and discarded the reads that have no inserted fragments due to the reason of adaptor self-connection; Secondly, low quality bases (quality value less than 20) were trimmed off at the end of the sequence (3'). If there are still bases with quality value less than 10 in the remaining reads, the whole reads are eliminated, otherwise it is retained. Thirdly, reads with "N" bases was also removed. Lastly, the reads less than 100bp in length were discarded after removing adapter and quality pruning. All the above mentioned processes were performed carried out using the SeqPrep (<https://github.com/jstjohn/SeqPrep>) and Sickle (Version 1.33) software [28]. The high quality clean reads of leaves and rhizomes of *D. opposita* were de novo assembled by Trinity (Version v2.8.5) software with operation parameters of k-mer size of 25, 27, 30, 31 and 32(max normalize read 50×Min contig length 200). The assembly results were assessed based on the assembly size, number of contigs, N50 and complete BUSCOs. Then, 2x150bp reads were generated and filtered, and more than 200 bp contigs were produced. At last, assembly completeness and quality also were assessed with Benchmarking Universal Single-Copy Orthologs (BUSCO, version 3.0.2) and TransRate package (version 1.0.3) [29, 30].

Functional annotation and classification of unigenes

The unigenes were compared with the NCBI BlastX program (E-value <1 E-5) against the non-redundant (NR), KEGG (Kyoto Encyclopedia of Genes and Genomes), KOG (clusters of orthologous groups for eukaryotic complete genomes), GO (gene ontology) and Pfam [31-33].

The functional classification of these unigenes was performed using Blast2GO and KEGG. The GO analysis was categorized into biological processes, molecular functions, and cellular component [34]. The pathway distributions were determined based on the KEGG database [35]. Unigenes were also aligned to the Evolutionary Genealogy of Genes: Non-Supervised Orthologous Groups (eggNOG) database for further functional prediction and classification [36].

Differential transcript abundance analysis

The clean reads were used to determine alignment counts and to quantify transcript abundance by the RSEM (RNA-seq by Expectation Maximization, <http://deweylab.github.io/RSEM/>) package [37]. The unigene and transcript abundance levels were analyzed by TPM (transcripts per million reads) [38].

The software DESeq2 (Version 1.24.0, <http://bioconductor.org/packages/ stats/bioc/DESeq2/>) was used to analyze the differentially expressed genes (DEGs) between leaf and rhizome tissues of *D. opposita*. The false discovery rate (FDR) can be used to assess p-value thresholds in multiple testing [39, 40]. The significance of DEGs was defined with a threshold of FDR < 0.05 and an absolute value of log2-fold

change of ≥ 2.0 ($|log2fc| \geq 1.0$) [41]. The DEGs enrichment analyses were carried out for GO and KEGG terms by the GOatools and KOBAS software [42, 43].

SSR Mining and analysis

The unigenes were searched for the prediction and detection of simple sequence repeats (SSR) through the MISA 1.0 software with default parameters [44]. The motifs of mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide repeats were identified. A minimum of five repetitions were considered as the search criteria in MISA. The parameters of definition of microsatellites (unit size/minimum number of repeats) were used to predict SSRs. That is Mono- (1/10), Di- (2/6), Tri-(3/5), Tetra- (4/5), Penta- (5/5) and Hexa-nucleotide (6/5). And the maximal number of bases interrupting 2 SSRs in a compound microsatellite was 100.

Expression analysis of terpene biosynthetic related DEGs

We selected 2 DEGs, DN39197_c0_g1 and DN37057_c0_g1, which were the key enzyme genes in the terpene synthesis from *D. opposita*, to validate their expression using quantitative real-time PCR (qRT-PCR). Total RNAs from leaf and rhizome tissues of *D. opposita* were extracted through the same procedures as used for RNA-seq in the study. The RNA from each sample was reverse transcribed using a reverse transcription kit (PC17-TRUE script, Aidlab Biotechnologies Co., Ltd, China). The *DoActin* gene was used as an internal control. Primers for the 2 DEGs and *DoActin* were designed in the Primer Premier 5.0 software. All the primers sequences are listed in Table 1. The reaction system (10 μ L) included 5 μ L of SYBR Premix Ex TaqTM, 4 μ L of cDNA template, and 0.5 μ L of each primer (10 μ M). The qRT-PCR reactions were conducted on a Roche Light Cycle 96 real-time PCR system (Roche, USA) with SYBR green fluorescent dye (TaKaRa, China). PCR amplification was initiated at 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s, 60 °C for 30 s, and 72 °C for 30 s. Three replicates were set for the experiment. Relative transcript abundance levels were calculated by the 2- $\Delta\Delta Ct$ method [45].

Table 1
Primers for qRT-PCR of 2 DEGs and DoActin

Unigene ID	Annotation	Primer sequence(5' – 3')
DN39197_c0_g1	DXR	DXR-F: GGTGCCGATTACGACCATATC
		DXR-R: GGATTGGCAGACGCATATCA
DN37057_c0_g1	GGPS	GGPS-F: ATCGGTGGTGGTTCTGATTG
		GGPS-R: CTGCAGTCTTCCCAAGTTCT
	Actin	Actin-F: CTAAGGCCAACAGAGAGAAGATG
		Actin-R: GGTACGGCCACTTGCATAAA

Results

Transcriptome sequencing and de novo assembly

We obtained averages of 48,786,977 and 55,202,249 raw reads for the rhizome (HSYK_1, HSYK_2, and HSYK_3) tissues and leaf (HSYY_1, HSYY_2, and HSYY_3) tissues, respectively, which consisted of 7,366,833,477 bp and 8,335,539,549 bp. We obtained 46,982,031 and 53,689,620 clean reads, including 7,053,884,682 bp and 8,067,994,222 bp clean bases, respectively. The sequences can be found in NCBI SRA with the accession number PRJNA628588. Moreover, the average GC content of rhizome and leaf tissues was 46.8% and 46.7%, respectively. The k-mer size of 25, 27, 30, 31 and 32 were selected to evaluate the assembly results based on the assembly size, number of contigs, N50 and complete BUSCOs. The results showed that the k-mer size of 30 for the final assembly results in the best final assembly. Trinity assembly resulted in 50,765 unigenes with fragment size ranging from 201 bp to 16,991 bp, and the average length of each unigene was 1049.6 bp. In addition, all of the unigenes consisted of a total of 53,283,125 bp. The unigenes were retained with an N50 length of 1825 bp and N90 length of 2797 bp. The TransRate assembly and Benchmarking UniversalSingle-Copy Orthologs (BUSCO) completeness assessment score were 0.26405 and 68.9%, respectively. The detailed sequencing and assembly results for the leaf and rhizome tissues of *D. opposita* are shown in Table 2.

Table 2
Optimized assembly results of unigenes from *D. opposita* rhizome and leaf tissues

Type	Unigene
Total sequence num	50,765
Total sequence base (bp)	53,283,125
Largest (bp)	16,991
Smallest (bp)	201
Average length(bp)	1,050
N50 (bp)	1,825
N90 (bp)	2,797
TransRate score	0.26405
Complete single copy BUSCOs	68.9%

Functional annotation of *D. opposita* unigenes

The functional annotation results indicated that a total of 24,840 unigenes were annotated (Table 3). Among them, different unigenes were matched in different databases, for example, 23,848 (46.98%) unigenes were matched in the NR database. In addition, 18,531 (36.50%) unigenes were matched in the Swiss-Prot database.

Table 3
The annotation statistics of *D. opposita*

Type	Unigene number (percent%)
NR	23,848(46.98)
Swiss-Prot	18,531(36.50)
Pfam	18,805(37.04)
COG	21,973(43.28)
GO	20,014(39.42)
KEGG	10,617(20.91)
Total_annotation	24,840(48.93)
Total	50,765(1)

The length distributions of annotated and unannotated unigenes results indicated that most annotated and unannotated unigenes were distributed in the 200–499 bp range (annotated unigenes: 5970/24,840; unannotated unigenes: 16,040/25,925), followed by the 500–799 bp ranges (annotated unigenes: 3446/24,840; unannotated unigenes: 5566/25,925), and 800–1099 bp range (annotated unigenes: 2394/24,840; unannotated unigenes: 2194/25,925). Long sequences of unigenes were clearly more likely to be annotated than those with short sequences. A total of 7151 unigenes measuring over 2000 bp in length were annotated, and only 371 unigenes were unannotated.

Functional classification results of unigenes

To further evaluate the annotated unigene data, the unigene sequences were searched against the cluster of orthologous groups of proteins (COG) databases. In total, 23,139 unigenes assigned to 23 COG categories under the following types: information storage and processing, cellular processes and signaling, metabolism, and poorly characterized (Fig. 1). Among them, alphabetized “Posttranslational modification, protein turnover, chaperones” (O) (1373; 5.93 %) was the categories with the most members in the COG databases.

Using the Blast2GO database, the 20,014 unigenes in the gene ontology (GO) analysis could be summarized into three main GO categories biological process (BP), molecular function (MF), and cellular component (CC) on two levels. In the BP category, the cellular process (7345; 36.70%), metabolic process (6480; 32.389%), and biological regulation (2125; 10.62%) were represented the most. Likewise, cell part (7756; 38.75%), membrane part (7018; 35.07%), organelle (4292; 21.44%) and organelle part (3154; 15.76%) were the most abundant in the CC category. Furthermore, the MF category mainly consisted of binding (10,949; 54.71%), catalytic activity (10,794; 53.93%), transporter activity (1074; 5.37%), and structural molecule activity (730; 3.65%) (Fig. S1).

In order to perform an in-depth analysis of unigene function based on the GO database, all of the three categories (BP, MF, and CC) were subdivided into level 3 and level 4 terms in order to determine their function. For example, cellular process is the main level 2 term from the BP category. In the level 3 terms, cellular process was assigned to the following specific functions: cellular metabolic process (GO: 0044237; 5615 unigenes; 28.06%), cellular component organization (GO:0016043; 1147 unigenes; 5.73%), cellular response to stimulus (GO:0051716; 555 unigenes; 2.78%), and signal transduction (GO:0007165; 526 unigenes; 2.63%), etc. (Fig. S2).

Through the KEGG database BLAST, 10,617 unigene sequences were assigned to 145 pathways (Table S1). Among them, the three pathways with the greatest number of unigene sequences were ribosome (map03010; 605 unigenes; 5.70%), plant-pathogen interaction (map04626; 281 unigenes; 2.65%), protein processing in endoplasmic reticulum (map04141; 276 unigenes; 2.60%), and thermogenesis (map04714; 248 unigenes; 2.34%) (Table S1). In addition, all the pathways were divided into five main types of KEGG metabolic pathways (Fig. 2).

Expression analysis of unigenes from leaf and rhizome tissues

In order to explore the differential expression in leaf and rhizome tissues of *D. opposita*, the clean reads from the leaf and rhizome library were compared with the unigene database by using the RNA-seq by RSEM package. The results indicated that 41,692,444 reads, respectively 77.62% of the original clean reads (53,689,620), were aligned in leaf tissues (HSYY: average reads of HSYY_1, HSYY_2, and HSYY_3). Likewise, 37,978,278 reads, representing 80.83% of the original clean reads (46,982.031), were aligned in rhizome tissues (HSYK: average reads of HSYK_1, HSYK_2, and HSYK_3).

In this study, the sample correlation analysis was performed using RSEM software (Tables S1). A high correlation coefficient was obtained among the three biological replicates of each tissue. The correlation coefficients of rhizome_1 (HSYK_1) with rhizome_2 (HSYK_2) and rhizome_3 (HSYK_3) were 0.935 and 0.913, respectively. The correlation coefficients of leaf_1 (HSYY_1) with leaf_2 (HSYY_2) and leaf_3 (HSYY_3) were 0.919 and 0.935, respectively. The correlation coefficient between two different of leaf and rhizome tissues was significantly lower than that among the three biological replicates of each leaf and rhizome tissue. The results indicated that all three biological replicates of each tissue are all clustered together, thus suggesting that the sequencing data are suitable for further biological analysis (Fig. S3).

Venn analysis of samples with a unigene expression greater than 2 was performed using RSEM software (Fig. S4). A total of 14,627 unigenes were shared among the six samples. Moreover, a total of 1552, 2295, 1271, 1195, 1588, and 1409 unigenes were specifically expressed in leaf_1 (HSYY_1), leaf_2 (HSYY_2), leaf_3 (HSYY_3), rhizome_1 (HSYK_1), rhizome_2 (HSYK_2), and rhizome_3 (HSYK_3), respectively. Overall, more unigenes were assembled in leaf tissues than in the rhizome tissues.

Principal component analysis (PCA) of the expression levels of total transcriptome (Fig. S5), revealed that the leaf (HSYY_1, HSYY_2, HSYY_3) and rhizome (HSYK_1, HSYK_2, HSYK_3) treatment groups could be

obviously distinguished. In other words, the three biological replicates of each sampling group clustered well together. Hence, the sequencing data can be used for further analysis.

Identification and functional enrichment analysis of DEGs

To identify the differentially expressed genes (DEGs), we compared the expression unigenes in leaf and rhizome tissues based on the false discovery rate (FDR) through the DESeq2 software. The results showed that 13,259 significant DEGs ($FDR < 0.05$ and $|log2FC| \geq 1$) were identified in the leaf versus rhizome tissues, which included 5536 up-regulated unigenes and 7723 down-regulated unigenes (Fig. 3, Table S2).

A total of 13259 DEGs were enriched to 179 GO terms based on thresholds of $FDR < 0.05$, $|log2FC| \geq 1$, p -value < 0.05 and p -value_corrected < 0.05 (Table S2). Fig. 4 indicates the top 20 enriched GO enrichment terms of DEGs. Among them, the membrane part (2964/8239), intrinsic component of membrane (2887/8239), integral component of membrane (2882/8239), and oxidoreductase activity (778/8239) were the most enriched GO terms of the DEGs (Fig. 4). By contrast, oxidoreductase activity, acting on iron-sulfur proteins as donors (13/8239), ferredoxin-NADP+reductase activity (10/8239) and omega peptidase activity (10/8239) were the least-enriched GO terms (p -value_corrected < 0.05).

To thoroughly understand the functions and interactions of DEGs between leaf and rhizome tissues, all DEGs were subjected to KEGG pathway enrichment analysis by mapping against the KEGG database using KEGG orthology based annotation system (KOBAS) software. The results indicated that 13,259 DEGs were involved in 144 metabolic pathways. There 23 pathways were enriched based on the p -value _uncorrected < 0.05 (Table S3). We found that these pathways were mainly related to the biosynthesis of other secondary metabolites, genetic information processing, and signal transduction, such as tropane, piperidine and pyridine alkaloid biosynthesis (map00960; 28/4059), DNA replication (map03030; 12/4059), and phosphatidylinositol signaling system (map04070; 22/4059).

The top 20 enriched pathways with a p -value of < 0.05 are presented in Fig. 5. We found that phenylpropanoid biosynthesis (map00940; 76/4059; 1.87%), photosynthesis (map04626; 54/4059; 1.33%), glutathione metabolism (map00480; 51/4059; 1.26%), glycine, serine and threonine metabolism (map00260; 49/4059; 1.21%), and flavonoid biosynthesis (map00941; 40/4059; 0.99%) were the highest enriched pathways.

Detection and Characterization of SSR marker

Using MISA 1.0 software, a total of 20,607 potential simple sequence repeats (SSRs) were identified across 13,770 unigenes (27.12%), of which 3464 unigenes contained more than one SSR. Among the 20,607 SSRs, the mono- and di-nucleotide repeats were the most abundant types (10,031, 48.68%; 5312, 25.78%, respectively), followed by tri- (4603, 22.34%), tetra- (440, 2.14%), hexa - (128, 0.62%), penta - (93, 0.45%), and nucleotide repeats. In addition, the repetitions of different SSRs types were calculated for four levels of repeat numbers 1–5, 6–10, 11–15, and >15 and are presented in Fig. S6. The results

indicated that most mono-nucleotide repetitions were distributed in repeat number 6–10 (4385), followed by mono-nucleotide in 6–10 (4385), di-nucleotide in 6–10 (4095), tri-nucleotide 6–10 (2344), tri-nucleotide 1–5 (1991), di-nucleotide in 11–15 (874), and tetra-nucleotide 1–5 (290). Analysis of the SSR motif length distribution showed that most of the SSR motifs are less than 100 bp in length. The results indicated that 17,412 (713) and 342 SSRs had motif length in the ranges of 10-50, 51-100 and 101-150 bp, respectively.

Identification of unigenes involved in terpene biosynthesis

Comparison to the unigenes with the KEGG pathway database, indicated, a total of 174 unigenes related to terpene biosynthesis (Table S1), including terpenoid backbone biosynthesis (61), monoterpenoid biosynthesis (37), diterpenoid biosynthesis (19), sesquiterpenoid and triterpenoid biosynthesis (8), biosynthesis of terpenoids and steroids (3), and ubiquinone and other terpenoid-quinone biosynthesis (43). Many unigenes were involved in the pathway of terpenoid backbone biosynthesis, which belongs to the universal metabolic pathway map00900. Among these unigenes, DN2456_c0_g1 (DXS, K01662) is the first enzyme in the MEP pathway. DXS catalyze D-glyceraldehyde 3-phosphate (GA-3P) and pyruvate to form 1-deoxy-D-xylulose-5-phosphate (DOXP). The *DXS* gene is significantly less expressed in the rhizome than in the leaf. DN26885_c0_g1 (HMGR, K00021) is the main enzyme in the MVA pathway. It catalyzes 3-hydroxy-3-ethylglutaryl CoA (HMG-CoA) to form MVA. The expression of *HMGR* gene shows no difference in the leaf and rhizome tissues of *D. opposita*.

qRT-PCR validations of DEGs related to terpene biosynthesis

To further analyze the consistency of RNA-seq data in this study, the DN39197_c0_g1 and DN37057_c0_g12 genes involved in terpene biosynthesis were selected for qRT-PCR validation. The results showed that a similar expression pattern in qRT-PCR analysis was observed in RNA-seq data from the leaf and rhizome tissues of *D. opposita* (Fig. 6). Further analysis showed that DXR (DN4169_c0_g1, K00099) and GGPS (DN17971_c0_g1, K13789) were the rate-limiting enzymes in the MVP and MEP pathways, which are the basic processes of a terpene synthesis pathway.

Discussion

In this study, six transcriptome sequencing analyses of RNA from leaf (HSYY_1, HSYY_2, HSYY_3) and rhizome (HSYK_1, HSYK_2, HSYK_3) tissues of *D. opposita* were conducted using the Illumina platform. More than five million raw reads and four million clean reads were obtained from leaf and rhizome library, respectively, which indicated that there more reads were assembled in leaf than in rhizome. Based on the large number of clean reads, 50,765 unigenes were assembled after evaluating the assembly results, of which, 46.98% unigenes were annotated by BLASTX analysis using the publicly available protein database. Otherwise, there are 49.5% unigenes were annotated from the *D. alata* [22]. This phenomenon shows that genomic data of *Dioscorea* species is relatively scarce in publicly database. There only *D. rotundata* [23], and *D. dumetorum* [24] genomic sequences were reported. The N50 is one of the most important indices used to evaluate the assembly quality. In the study, the N50 length of unigenes was

1825 bp, indicating that the sequence assembly was high in quality and suitable for intensive research. Moreover, the N50 could be compared with those of other species [46–48]. The average length of the unigenes was larger (1050 bp) (Table 2), than those in previous studies, such as those of *Myrica rubra* (531 bp) [49] and *Sorbus pohuashanensis* (770 bp) [50].

In this study, 24,840 and 25,925 unigenes were annotated and unannotated, respectively, in the NR, KEGG, COG, GO and Pfam database. However, the proportion of annotated long sequences was significantly higher than that of annotated short sequences. Similar results have been reported for other species [17, 51].

The identified unigenes were compared among different *Dioscorea* species. A total of 5575 unigenes were matched with 315 *Dioscorea* species. Unigenes with the highest similarity were found in *D. polystachya* (16.41%), followed by *D. bulbifera* (5.70%), *D. alata* (4.77%), and *D. rotundata* (3.28%), this result may indicate the evolutionary relationship among these species. The BLAST results (e-value 0.00001) of these *Dioscorea* species confirmed the correctness of transcriptome assembly and analysis in the present study. However, 296 *Dioscorea* species were matched to less than 100 unigenes of *D. opposita*. Although a small number of genes from *Dioscorea* species are available in several public database, complete transcriptome or genomic data are scarce. Indeed, to date, only the genomic resources for *D. alata* [22], *D. rotundata* [23], and *D. dumetorum* [24] have been reported. Therefore, enhancing research efforts on the genome of *Dioscorea* species is important and the main focus of the present study.

Most *D. alata* tubers have white flesh, and purple-flesh tubers with high anthocyanidin content are produced by spontaneous variation. In the present study, elite purple-flesh cultivar (DP) and conventional white-flesh cultivar (DW) of *D. alata* were performed using Illumina sequencing [22]. The transcriptomes were compared with the two species of *D. alata* [22], and we compared with two tissues of *D. opposita*. A total of 49.5% (60,02/125,123) [22] and 48.93% (24,840/ 50,765) unigenes were annotated in *D. alata* and *D. opposita*, respectively, which indicates that few genome sequences of *Dioscorea* species are available in public database; thus, the ratio of annotation is fairly low. For example, the alignment lengths of gene NC_039707.1 (AYI69196.1) in *D. alata* and gene DN1708_c1_g2 in *D. opposita* are identical at 2250 bp (750 aa), and the similarity of these genes is 100%. Thus, these genes are believed to be the *psbA* gene, which encodes Photosystem I P700 chlorophyll a apoprotein A1 protein. 97 unigenes that were compared with *D. dumetorum*. Among these unigenes, the unigene DN2638_c1_g5 in *D. opposita* was identified to be gene NC_039691.1 (YP_009528139.1) in *D. dumetorum*; these genes are 465 bp (155 aa), in length and encode ribosomal protein S7 in the chloroplast genome.

A previous report matched a total of 11,115, 3542, 3304, 2943, and 2859 unigenes of *D. alata* with sequences from *Vitis vinifera*, *Oryza sativa*, *Populus trichocarpa*, *Zea mays*, and *Ricinus communis*, respectively, with high similarity [22]. However, in the present study, only 510, 170, 38, 280, and 80 unigenes were matched with sequences from *V. vinifera*, *O. sativa*, *P. trichocarpa*, *Z. mays*, and *R. communis*, respectively.

Through the Blast2GO database, a large number of unigenes GO categories were assigned (Table 3), suggesting that the assembled unigenes represented a wide diversity of transcripts in the *D. opposita* genome, similar to in many previous reports [52, 53]. In the three GO categories, reproduction and metabolic processes were the most abundant classes in BP cell and cell part activity were the most abundant classes in CC, which was consistent with some previous research [17, 51, 52].

Likewise, compared with the transcriptome of differentially expressed genes of leaf versus rhizome tissues by the DESeq2 software, there were 13,259 unigenes existing in significant DEGs, with 5536 up-regulated and 7723 down-regulated unigenes. These results showed that more unigenes abundance down-regulated in rhizome than in leaf tissue, which may be related to secondary metabolites of rhizome tissue. Moreover, most of these unigenes were related to transferase activity, membrane part, photosynthesis, plant-pathogen interaction, and plant hormone signal transduction.

Trinity can be used with different k-mer size and the stranded flag for de novo transcriptome assembly [54]. The developers describe comprehensive workflow for assembly validation on the trinity software. In this study, analyses by using five k-mer sizes of 25, 27, 30, 31 and 32 were performed with Trinity (Version v2.8.5) website. The de novo transcriptome assembly was generated using the stranded flag and 2 × 150nt paired-end reads. The assembly results were evaluated on the basis of assembly size, number of contigs, and BUSCO (Table S4). The number of complete BUSCOs is highest for a k-mer size of 30. Meanwhile, the N50 of 1825 bp is in the expected range for a transcriptome assembly. So, 30 were selected as the best k-mer size for the assembly and analysis of the transcriptome data of leaf and enzymes tissues from *D. opposita*.

Based on previous reports [55, 56], we speculate the expression levels of dioscin in different tissues of *D. opposita* may be different and are related to key enzymes of the biosynthetic pathway. Therefore, to explore the dioscin accumulation and translocation in leaf and rhizome tissues, we compared the differentially expressed genes (DEGs) and elementary identified terpene biosynthetic genes through the RNA-seq, which would lay a foundation to study the terpenoids biosynthesis pathway.

We analyzed SSR sequences with length of 1–30 bp, and detected 16,417 unigenes. Among these unigenes, 2338 and 5782 were protein coding and non-coding sequences, respectively. The results also indicated that SSRs with lengths of 3 or multiples of 3 were identified more often within coding sequences than SSRs of other lengths, likely because a complete codon would not lead to a frameshift, but to the addition/deletion of an amino acid.

Terpenoid biosynthesis requires common precursor substances, including isopentenyl diphosphate (IPP) and its allylic isomer dimethylallyl diphosphate (DMAPP), which are generated through the MVA and MEP pathways [55, 56]. In the MEP pathway, DXR is a rate-limiting enzyme, that catalyze 1-deoxy-D-xylulose-5-phosphate (DXP) to generate MEP [57]. GGPS is the main enzyme in the terpene biosynthesis pathway, and can catalyze IPP to generate geranylgeranyl diphosphate (GGPP) [57]. In this study, many DEGs involved in terpenoids synthesis, such as diterpenoid biosynthesis, monoterpenoid biosynthesis and tropane, piperidine and pyridine alkaloid biosynthesis were detected. The expression patterns of the DXR

(DN4169_c0_g1) and GGPS (DN17971_c0_g1) genes were studied through qRT-PCR, the results of which were consistent with the RNA-Seq data. Our results also indicated that the abundance of DXR and GGPS genes is higher in the leaf tissue than in the rhizome tissue of *D. opposita*. This phenomenon may be related to the synthesis and accumulation of secondary metabolites in these tissues. Other enzymes, such as HMGR and hydroxymethylglutaryl-CoA synthase (HMGS) may also participate in the biosynthesis of terpenoids. These genes will be studied in great depth in future work.

Conclusions

In this study, the RNA-seq of leaf and rhizome tissues of *D. opposita* were obtained using an Illumina Hiseq 4000 platform. The results enriched the genomic data of *D. opposita*, and explore the different expression genes between leaf and rhizome tissues of *D. opposita*.

Abbreviations

MEP: 2-C-Methyl-D-Erythritol-4-Phosphate; GGPS: Geranylgeranyl pyrophosphate synthase; HMGR: Hydroxymethylglutaryl-CoA reductase; DXR: 1-deoxy-D-xylulose 5-phosphate reductoisomerase; DXS: 1-deoxy-D-xylulose-5-phosphate synthase; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; PCR: polymerase chain reaction; MVA: Mevalonate acid; TPM: Transcripts per million reads; NR: Non-redundant; KOG: clusters of orthologous groups for eukaryotic complete genomes; RSEM: RNA-seq by Expectation Maximization; DEGs: differentially expressed genes; FDR: False discovery rate; SSRs: Simple sequence repeats.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

Not applicable

Acknowledgments

We thank Dr. Shuang Zhou for her help with the data analysis. We would also thank manager Jinjing Chen Shanghai Majorbio Bio-pharm Technology Co., LTD, Shanghai, China for her guidance and assistance in data analysis.

Authors' contributions

All authors have participated extensively in the study and had proofread the final manuscript. DYH, JX and XGH conceived and designed the research. ZHL, JYX and RYC conducted the experiments. ZQM and HWX analyzed the data. DYH, RYC and ZHL wrote the manuscript. HWX and XLZ approved and reviewed the final manuscript. All authors read and approved the final manuscript.

Funding

This research was supported by the National Natural Science Foundation of China (U1404829)·Key project at central government level: The ability establishment of sustainable use for valuable Chinese medicine resources (2060302)·Science and technology project of Henan province(202102110156)·This research was also supported by a special fund for the construction of technical systems of the traditional Chinese medicine industry in Henan Province.

Availability of data and materials

The sequences can be found in the National Center for Biotechnology Information database (NCBI) SRA with the accession number PRJNA628588. Other datasets used and/or analyzed during the current study would be available from the corresponding author on reasonable request.

Author details

¹Agricultural College, Henan University of Science and Technology, Luoyang 471023, China.

²The Luoyang Engineering Research Center of Breeding and Utilization of Dao-di Herbs, Luoyang 471023, China. ³ Institute of Chinese Materia Medical, China Academy of Chinese Medical Sciences, Beijing, China.

References

1. Xie YF, Chang PR, Wang SJ, Yu JG, Ma XF. Preparation and properties of halloysite nanotubes/plasticized *Dioscorea opposita* Thunb. starch composites. Carbohydr Polym. 2011;83(1):186–91.
2. Zhang LM, Liu P, Wang YG, Gao WY. Study on physico-chemical properties of dialdehyde yam starch with different aldehyde group contents. Thermochim Acta. 2011;512(1–2):196–201.
3. Zhao GH, Li ZX, Chen ZD. Structural analysis and antitumor activity of RDPS-I polysaccharide from Chinese yam. Acta Pharm Sin. 2003;38(1):37–41.
4. Wang SJ, Yu JL, Gao WY, Liu HY, Xiao PG. New starches from traditional Chinese medicine (TCM)-Chinese yam (*Dioscorea opposita* Thunb.) cultivars. Carbohydr Res. 2006;341(2):289–93.
5. Chang SJ, Lee YC, Liu SY, Chang TW. Chinese yam (*Dioscorea alata* cv. Tainung No. 2) feeding exhibited antioxidative effects in hyperhomocysteinemia rats. J Agric Food Chem. 2004;52(6):1720–

5.

6. Ma FY, Zhang Y, Wen YR, Yao YN, Zhu JH, Liu XH, et al. Emulsification properties of polysaccharides from *Dioscorea opposita* Thunb. *Food Chem.* 2017;221(15):919–25.
7. Chen YF, Zhu Q, Wu SJ. Preparation of oligosaccharides from Chinese yam and their antioxidant activity. *Food Chem.* 2015;173:1107–10.
8. Herlina H. Deproteinase effect of hydrocolloid flour made of “Gembili Tuber” (*Dioscorea esculenta* L.) on chemical and technical functional properties. *International Journal on Advanced Science Engineering Information Technology.* 2015;5(4):298–302.
9. Chen JT, Huang WJ, Mu Q. Active components and mineral elements analysis of three kinds of Chinese yam. *Sci Tech food Industry.* 2017;38(7):352–5.
10. Guitton Y, Nicolè F, Moja S, Valot N, Legrand S, Jullien F, et al. Differential accumulation of volatile terpene and terpene synthase mRNAs during lavender (*Lavandula angustifolia* and *L. x intermedia*) inflorescence development. *Physiol plant.* 2010;138(2):150–63.
11. Sun C, Li Y, Wu Q, Luo HM, Sun YZ, Song JY, et al. De novo sequencing and analysis of American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. *BMC Genom.* 2010;11:262.
12. Zhang HL, Yang B, Liu J, Guo DL, Hou J. Analysis of structural genes and key transcription factors related to anthocyanin biosynthesis in potato tubers. *Sci Hortic.* 2017;225(1):310–6.
13. Li J, Liang Q, Li CF, Liu MD, Zhang YS. Comparative Transcriptome Analysis Identifies Putative Genes Involved in Dioscin Biosynthesis in *Dioscorea zingiberensis*. *Molecules.* 2018;23(2):454.
14. Piskol R, Ramaswami G, Li JB. Reliable Identification of Genomic Variants from RNA-Seq Data. *Am J Hum Genet.* 2013;93(4):641–51.
15. Li CF, Zhu YJ, Guo X, Sun C, Luo HM, Song JY, et al. Transcriptome analysis reveals ginsenosides biosynthetic genes, microRNAs and simple sequence repeats in Panax ginseng. *BMC Genom.* 2013;14(1):245.
16. Xiao M, Zhang Y, Chen X, Lee EJ, Barber CJ, Chakrabarty R, et al. Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J Biotechnol.* 2013;166(3):122–34.
17. Wang DJ, Yang CL, Dong L, Zhu JC, Wang JP, Zhang SF. Comparative transcriptome analyses of drought-resistant and susceptible *Brassica napus* L. and development of EST-SSR markers by RNA-SEq. *J Plant Biol.* 2015;58:259–69.
18. Pandey A, Swarnkar V, Pandey T, Srivastava P, Kanojiya S, Mishra DK, et al. Transcriptome and metabolite analysis reveal candidate genes of the cardiac Glycoside biosynthetic pathway from *Calotropis procera*. *Sci Rep.* 2016;6:34464.
19. Akhtar MQ, Qamar N, Yadav P, Kulkarni P, Kumar A, Shasany AK. Comparative glandular trichome transcriptome based gene characterization reveals reasons for differential (−)-menthol biosynthesis in *Mentha* species. *Physiol Plant.* 2017;160(2):128–41.

20. Wang XH, Hui F, Yang YC, Yang SH. Deep sequencing and transcriptome analysis to identify genes related to biosynthesis of aristolochic acid in *Asarum heterotropoides*. *Sci Rep.* 2018;8:17850.
21. Li WG, Xu RR, Yan XG, Liang DM, Zhang L, Qin XY, et al. De novo leaf and root transcriptome analysis to explore biosynthetic pathway of Celangulin V in *Celastrus angulatus* maxim. *BMC Genom.* 2019;20:7.
22. Wu ZG, Jiang W, Mantri N, Bao XQ, Chen SL, Tao ZM. Transcriptome analysis reveals flavonoid biosynthesis regulation and simple sequence repeats in yam (*Dioscorea alata* L.) tubers. *BMC Genom.* 2015;16(1):1–12.
23. Tamiru M, Natsume S, Takagi H, White B, Yaegashi H, Shimizu MY, et al. Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. *BMC Biol.* 2017;15(1):86.
24. Siadjeu C, Pucker B, Viehöver P, Albach DC, Weisshaar B. High Contiguity De Novo Genome Sequence Assembly of Trifoliate Yam (*Dioscorea dumetorum*) Using Long Read Sequencing. *Genes.* 2020;11(3):274.
25. Chen YZ, Chen JZ. Comparative study on the pharmacological active components of four kinds of Chinese yam. *Lishizhen Medicine Materia Medica Research.* 2014;25(10):2389–91.
26. Xu HW, Mo YW, Wang W, Wang H. OsPIN1a gene participates in regulating negative phototropism of rice roots. *Rice Sci.* 2014;21(2):83–9.
27. Jia XP, Dai LF, Quan JZ, Li ZY, Fan BY. Clong and mutation sites analysis of a putative HD3A_like gene in eleven foxtail millet cultivars. *Pak J Agr sci.* 2018;55(1):29–32.
28. 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;29:644–52.
29. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol.* 2018;35(3):543–8.
30. Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S. TransRate:reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* 2016;26(8):1134–44.
31. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* 2019;47(D1):D427–32.
32. Pruitt KD, Tatusova T, Maglott DR. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 2007;35(suppl_1):D61–5.
33. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421.
34. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005;21(18):3674–6.

35. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2008;36(suppl_1):D480-4.
36. Du F, Nu Er zi ya Ya Li mai ti, Hu QX, Ye ZYJ, Zhang D HJ. A Comparative Transcriptome Analysis Reveals Physiological Maturation Properties of Mycelia in *Pleurotus tuoliensis*. *Genes.* 2019; 10(9):703.
37. Li B, Dewey CN. RSEM:accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
38. Damasco G, Shivakumar VS, Misciewicz TM, Daly DC, Fine PVA. Leaf Transcriptome Assembly of Protium copal (Burseraceae) and Annotation of Terpene Biosynthetic Genes. *Genes.* 2019;10(5):392.
39. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res.* 2001;125(1–2):279–84.
40. Zhu YY, Li YQ, Xin DD, Chen WR, Shao X, Wang Y, et al. RNA-Seq-based transcriptome analysis of dormant flower buds of Chinese cherry (*Prunus pseudocerasus*). *Gene.* 2015;555(2):362–76.
41. Niu J, Hou XY, Fang CL, An JY, Ha DL, Qiu L, et al. Transcriptome analysis of distinct *Lindera glauca* tissues revealed the differences in the unigenes related to terpenoid biosynthesis. *Gene.* 2015;559(1):22–30.
42. Tang HB, Wang XY, Bowers JE, Ming R, Alam M, Paterson AH. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* 2008;18(12):1944–54.
43. Xie C, Mao XZ, Huang JJ, Ding Y, Wu JM, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 2011;39(suppl_2):W316-22.
44. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet.* 2003;106:411–22.
45. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2-DDC_t method. *Methods.* 2001;25(4):402–8.
46. Hu LS, Hao CY, Fan R, Wu BD, Tan LH, Wu HS. De Novo Assembly and Characterization of Fruit Transcriptome in Black Pepper (*Piper nigrum*). *PLoS One.* 2015;10(6):e0129822.
47. Niño MC, Kang KK, Cho YG. Genome wide transcriptional response of papain like cysteine protease mediated resistance against *Xanthomonas oryzae* pv. *oryzae* in rice. *Plant Cell Rep.* 2020;39:457–72.
48. Liu SH, Yue TT, Ahmad MJ, Hu XW, Zhang XX, Deng TX, et al. Transcriptome Analysis Reveals Potential Regulatory Genes Related to Heat Tolerance in Holstein Dairy Cattle. *Genes.* 2020;11(1):68.
49. Feng C, Chen M, Xu CJ, Bai L, Yin XR, Li X, et al. Transcriptomic analysis of Chinese bayberry (*Myrica rubra*) fruit development and ripening using RNA-SEq. *BMC Genom.* 2012;13:19.
50. Liu CC, Dou Y, Guan XL, Fu Q, Zhang Z, Hu ZH, et al. De novo transcriptomic analysis and development of EST-SSRs for *Sorbus pohuashanensis* (Hance) Hedl. *PLoS One.* 2017;12(6):e0179219.

51. Hou DY, Shi LC, Yang MM, Li J, Zhou S, Zhang HX, et al. De novo transcriptomic analysis of leaf and fruit tissue of *Cornus officinalis* using illumina platform. *PLoS One*. 2018;13(2):e0192610.
52. Tang LD, Liu JM, Liu LH, Yu YH, Zhao HY, Lu W. De Novo Transcriptome Identifies Olfactory Genes in *Diachasmimorpha longicaudata* (Ashmead). *Genes*. 2020;11(2):144.
53. Kim SH, Kim SW, Lim GH, Lyu JI, Choi HI, Jo YD, et al. Transcriptome analysis to identify candidate genes associated with the yellow-leaf phenotype of a *Cymbidium* mutant generated by γ -irradiation. *PLoS One*. 2020;15(1):e0228078.
54. Haak M, Vinke S, Keller W, Droste JL, Rückert C, Kalinowski J, et al. High quality de Novo transcriptome assembly of *Croton tiglum*. *Front Mol Biosci*. 2018;5:62.
55. Xu C, Wei H, Movahedi A, Sun WB, Ma XX, Li DW, et al. Evaluation, characterization, expression profiling, and functional analysis of DXS and DXR genes of *Populus trichocarpa*. *Plant Physiol Biochem*. 2019;142:94–105.
56. Tong YR, Su P, Zhao YJ, Zhang M, Wang XJ, Liu YJ, et al. Molecular Cloning and Characterization of DXS and DXR Genes in the Terpenoid Biosynthetic Pathway of *Tripterygium wilfordii*. *Int J Mol Sci*. 2015;16(10):25516–35.
57. Ma J, Xu ZS, Tan GF, Wang F, Xiong AS. Distinct transcription profile of genes involved in carotenoid biosynthesis among six different color carrot (*Daucus carota L.*) cultivars. *Acta Biochim Biophys Sin*. 2017;49(9):817–26.

Additional Files

Additional file 1: Table S1. The 145 pathways statistics

Additional file 2: Table S2. Summary of significant DEGs with leaf versus rhizometissues

Additional file 3: Table S3. Summary of KEGG pathway enrichment of significant DEGs with leaf versus rhizome tissues

Additional file4: Table S4. Results of different k-mer sizes were assessed for the de novo transcriptome of leaf and rhizome tissues.

Additional file 5: Fig. S1 Gene ontology (GO) level 2 term category classification of unigenes. The upper x-axis indicates the percentage of unigenes, and the lower x-axis indicates the number of unigenes. The y-axis indicates the specific categories of genes in the main categories.

Additional file 6: Fig. S2 Gene ontology (GO) terms level 3 categories classification of unigenes

Additional file 7: Fig. S3 The correlation coefficients heat map of the comparative transcriptome analysis of leaf (HSYY) and rhizome (HSYK) tissues. HSYY_1, HSYY_2, HSYY_3 and HSYK_1, HSYK_2, HSYK_3 indicate three biological replications of leaf and rhizome tissues, respectively,

Additional file 8: Fig. S4 The venn diagram of six samples, including three biological replicates of leaf (HSYY_1, HSYY_2, and HSYY_3) and rhizome (HSYK_1, HSYK_2, and HSYK_3) tissues,

Additional file 9: Fig. S5 The PCA analysis of leaf (HSYY_1, HSYY_2, and HSYY_3) and rhizome (HSYK_1, HSYK_2, and HSYK_3) tissues (A score plot of the first two PCs for samples grouped, the ellipses represent group membership assuming the 95% confidence). The x-axis and y-axis indicate the contribution degree of a principal component N (PCN) and principal component M (PCM) to the differentiated samples in the two-dimensional figure, respectively.

Additional file 10: Fig. S6 The distribution statistics of SSR types. The x-axis indicates the different SSR types. The y-axis indicates the number of SSRs of different types corresponding to different repetition times,

Figures

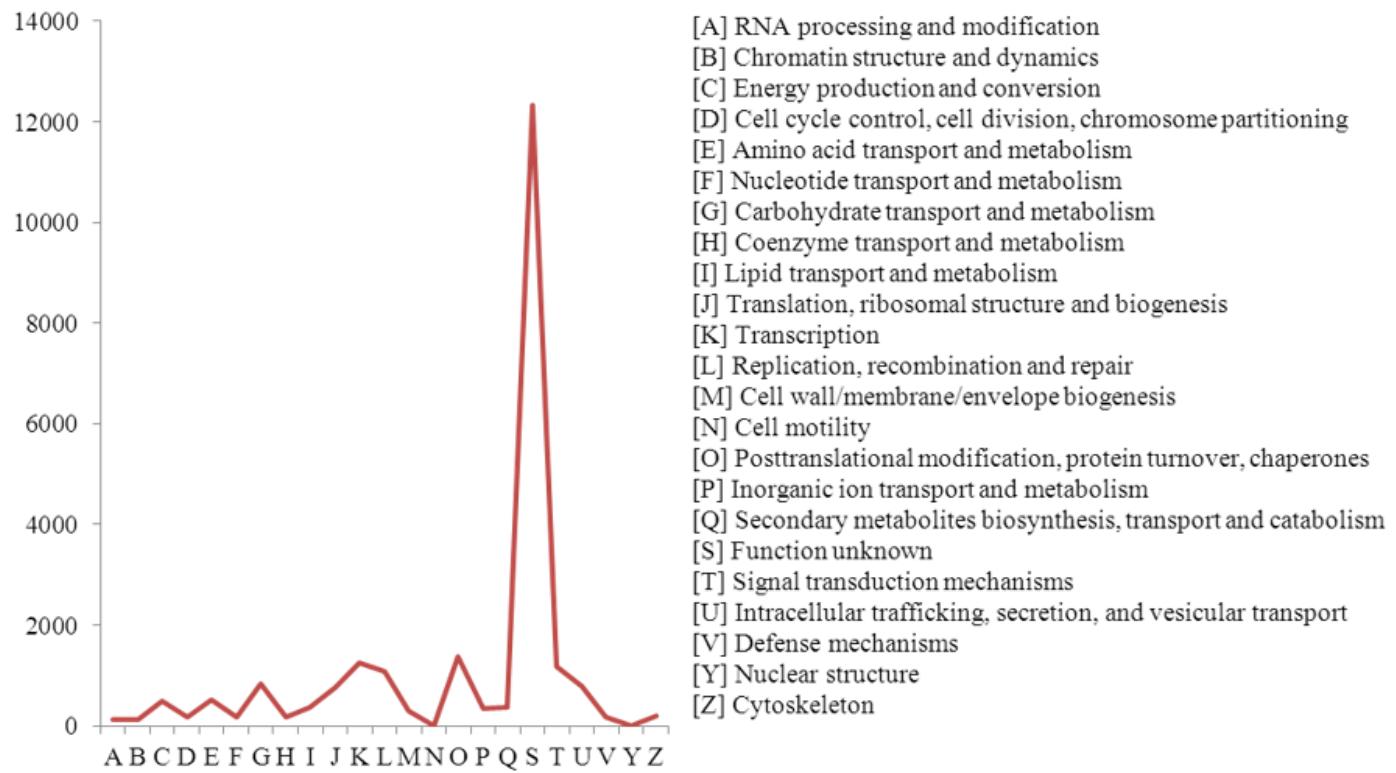


Figure 1

Clusters of COG functional classification of *D. opposita* unigenes ($E\text{-value} \leq 1.0 \times 10^{-5}$). The capital letters on the x-axis indicate the COG categories as listed to the right of the histogram. The y-axis indicates the number of unigenes.

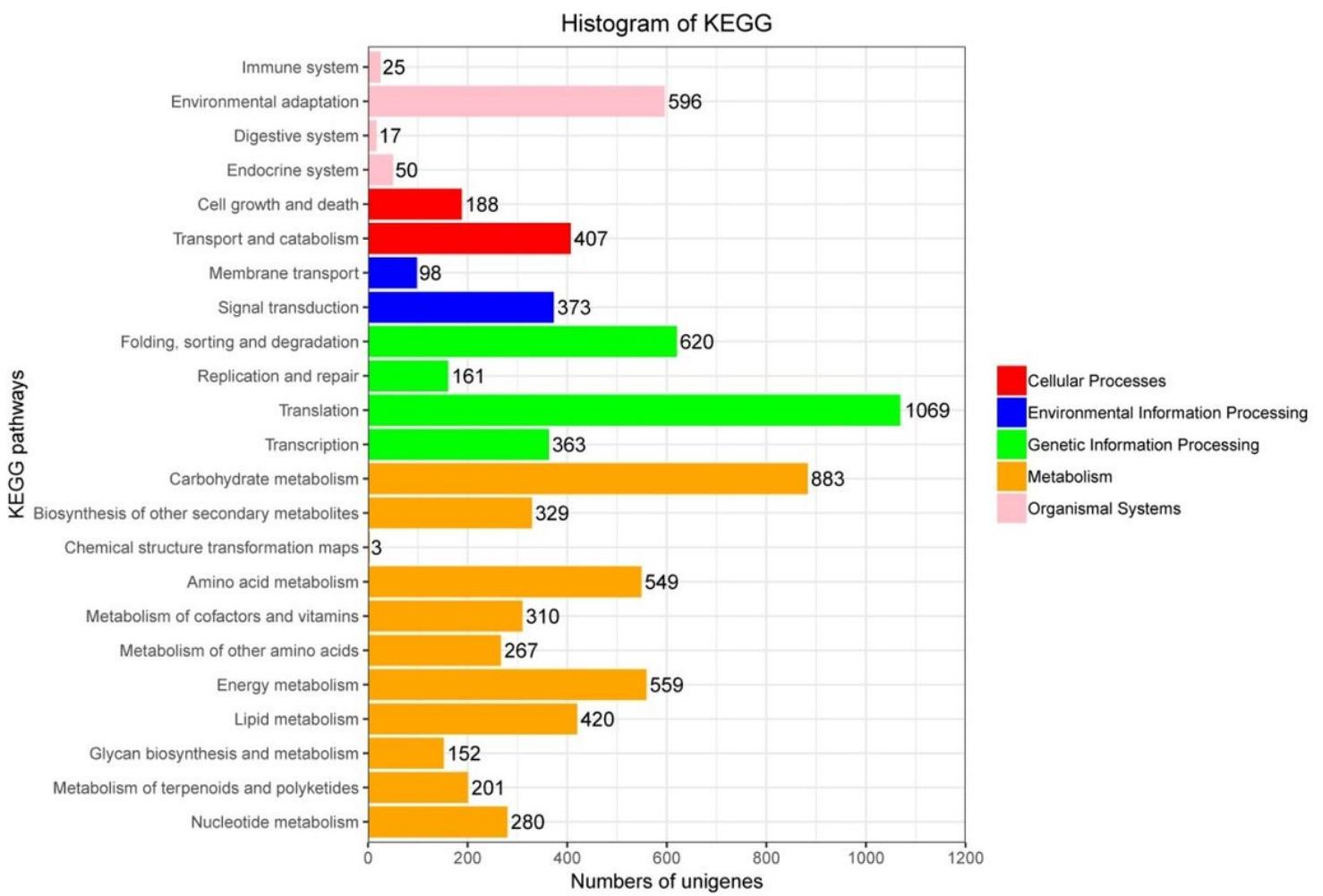


Figure 2

Classification map of KEGG metabolic pathways. The different colors indicate the different metabolic pathways.

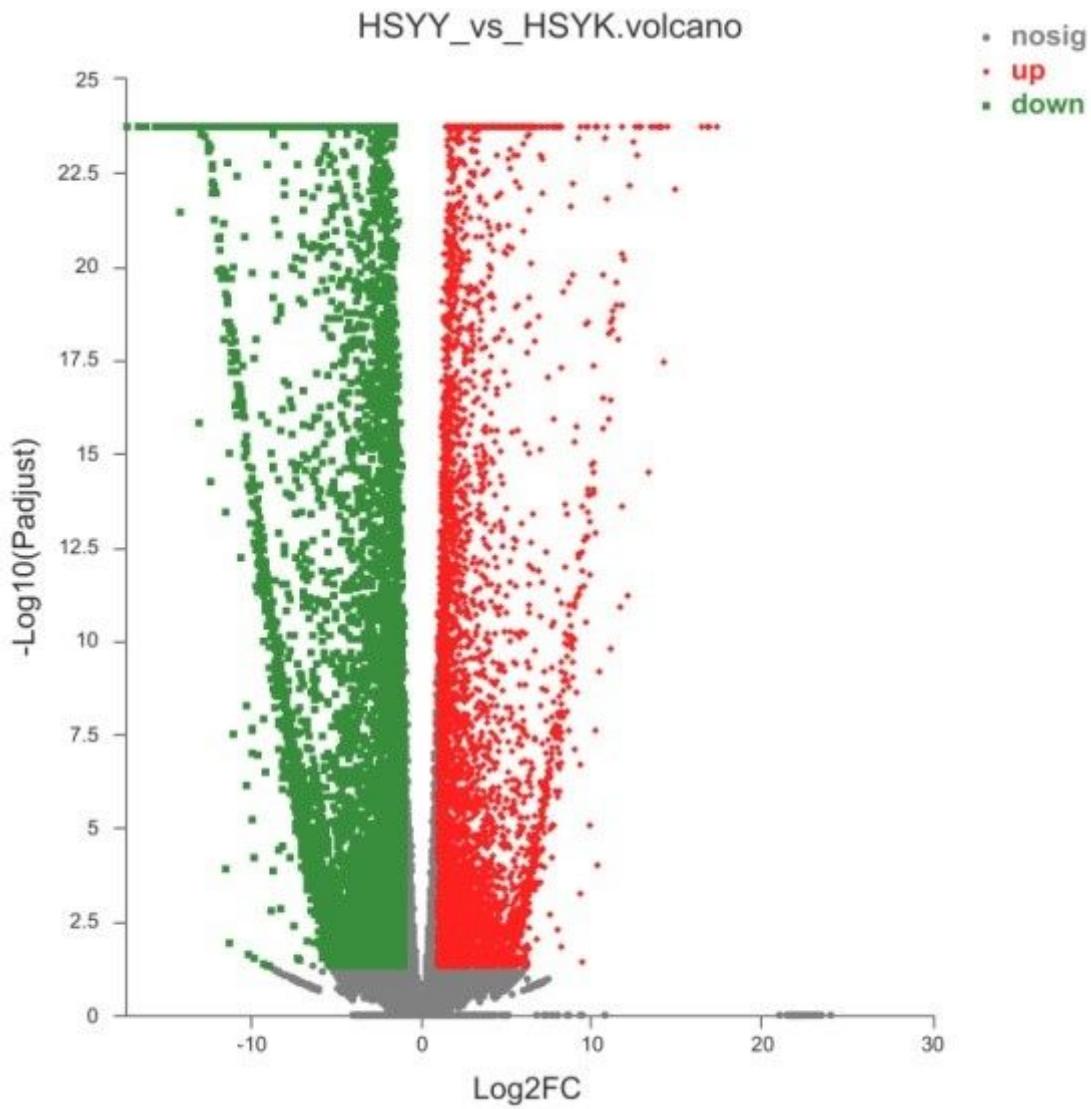


Figure 3

Volcano plots showing the expression level of each unigene of leaf (HSYY) and rhizome (HSYK) tissues ($p\text{-value} \leq 0.05$, $|\log_2\text{FC}| \geq 1$).

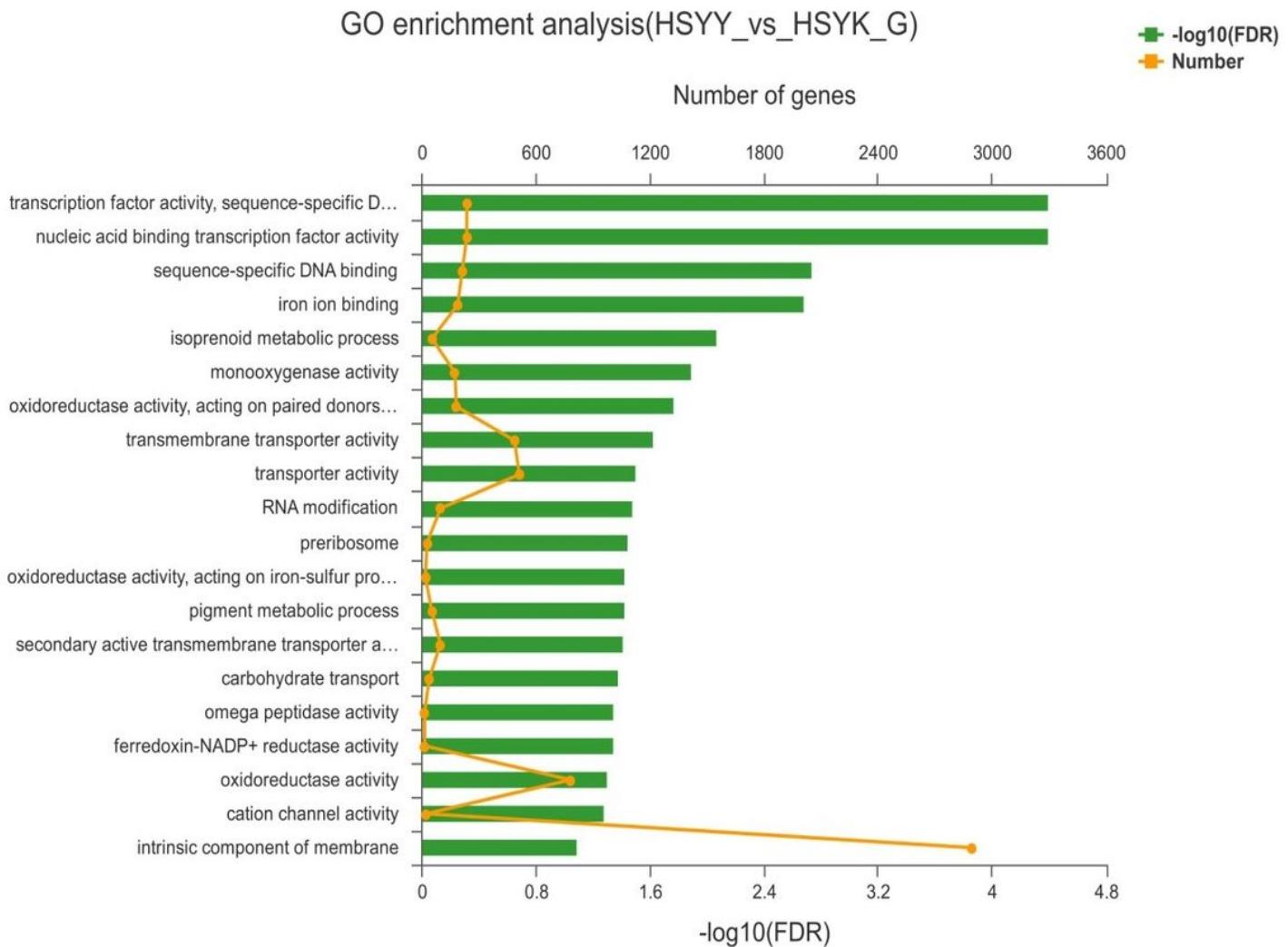


Figure 4

The top 20 enriched GO enrichment terms of differentially expressed genes (DEGs) of leaf (HSYY) and rhizome (HSYK) tissues. The upper x-axis indicates the number of unigenes, and the lower x-axis indicates the false discovery rate (FDR). The larger the FDR value, the more significant the enrichment. The y-axis indicates the different metabolic pathways. Each point on the fold represents the number of each metabolic pathway.

KEGG enrichment analysis(HSYY_vs_HSYK_G)

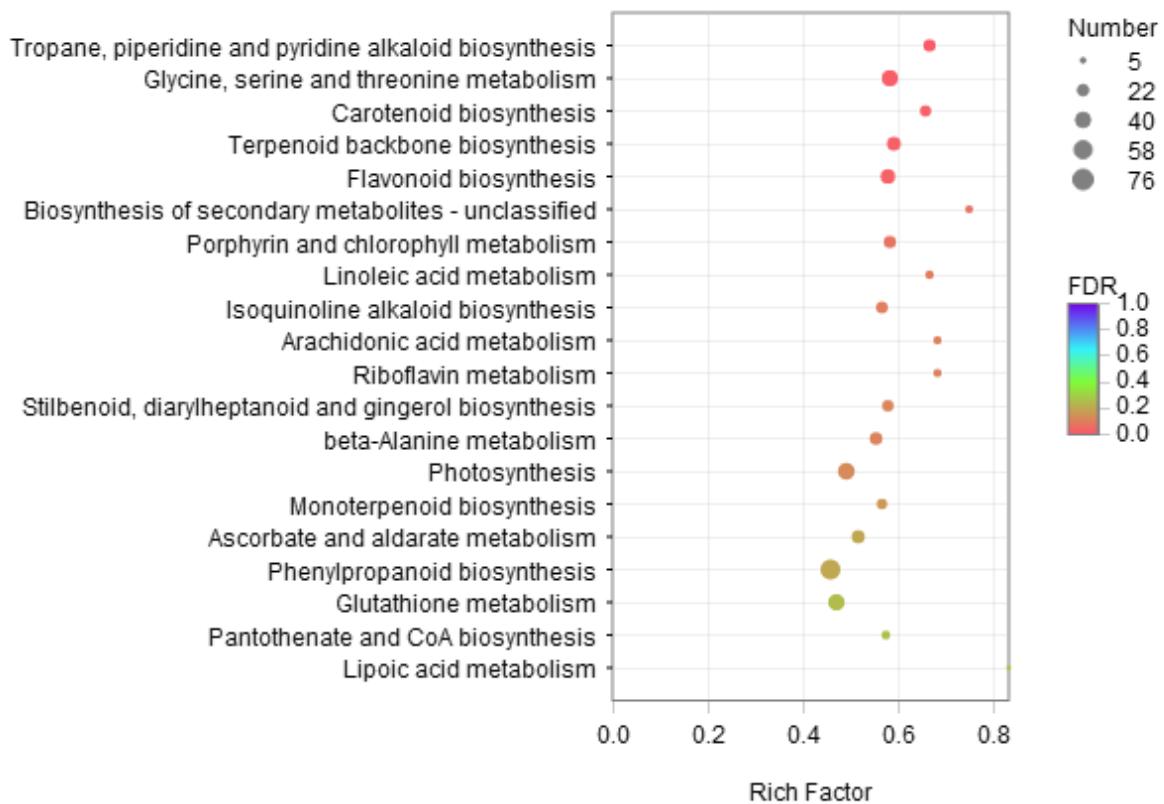


Figure 5

The top 20 enriched KEGG pathways of DEGs between leaf (HSYY) and rhizome (HSYK) tissues.

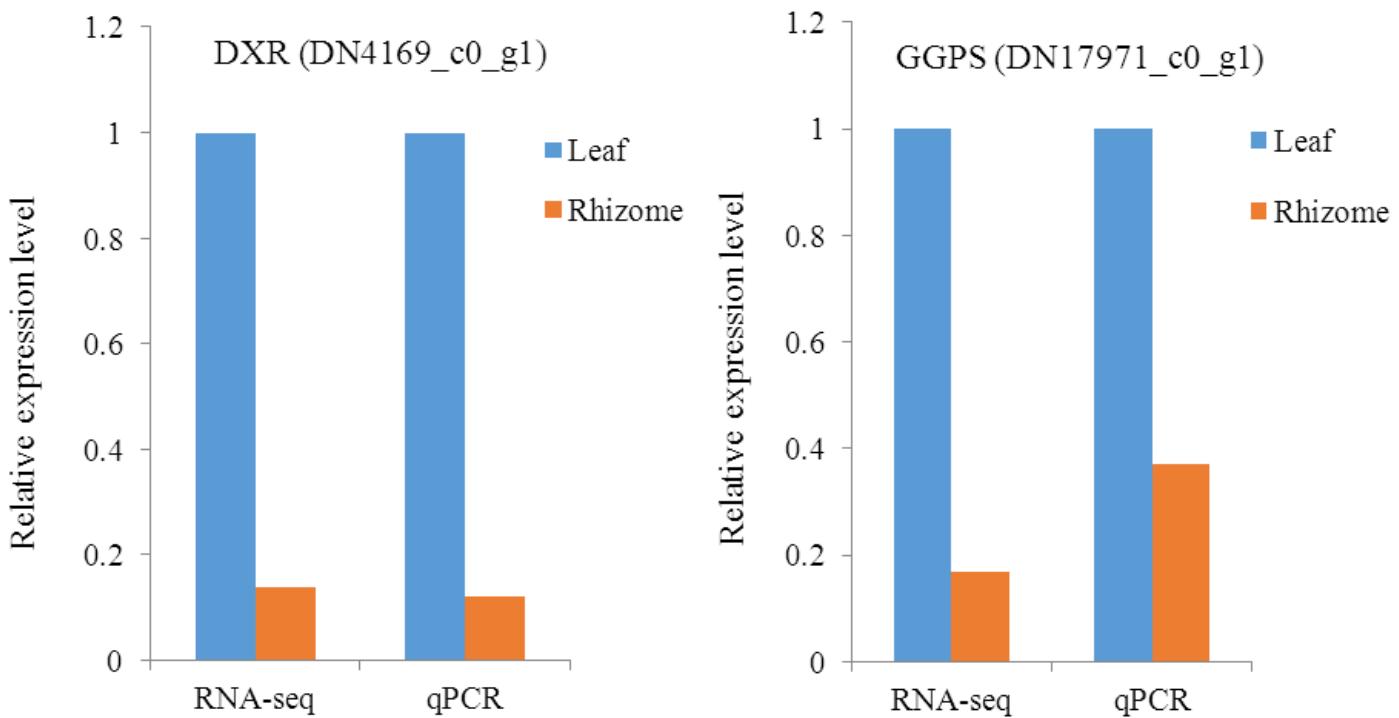


Figure 6

Expression of two selected unigenes as determined through qRT-PCR in comparison to the RNA-seq data.

Supplementary Files

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

- Fig.S1.pdf
- TableS1S4.xlsx
- Fig.S5.pdf
- Fig.S3.pdf
- Fig.S6.pdf
- Fig.S4.pdf
- Fig.S5.pdf
- Fig.S1.pdf
- Fig.S4.pdf
- Fig.S3.pdf
- Fig.S2.pdf
- TableS1S4.xlsx
- Fig.S6.pdf

- Fig.S2.pdf