Compatative Transcriptome Analysis Reveals Sesquiterpene Biosynthesis among 1-, 2- and 3-year Old Atractylodes Chinensis

Jianhua Zhao
Hebei Normal University of Science and Technology
Chengzhen Sun
Hebei Normal University of Science and Technology
Shanshan Ma
Hebei Normal University of Science and Technology
Jinshuang Zheng (jinshuangk@163.com)
Hebei Normal University of Science and Technology
Xin Du
Hebei Normal University of Science and Technology
Liping Zhang
Hebei Normal University of Science and Technology

Research Article

Keywords: Atractylodes chinensis (DC.) Koids., transcriptome sequencing, differentially expressed genes, bioactive compounds, sesquiterpene biosynthesis, qRT- PCR

Posted Date: March 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-272708/v1

Version of Record: A version of this preprint was published at BMC Plant Biology on July 27th, 2021. See the published version at https://doi.org/10.1186/s12870-021-03131-1.

Compatative transcriptome analysis reveals sesquiterpene biosynthesis among 1-, 2- and 3-year old *Atractylodes chinensis*

Jianhua Zhao, Chengzhen Sun, Fengyu Shi, Shanshan Ma, Jinshuang Zheng*, Xin Du, Liping Zhang

Hebei Key Laboratory of Crop Stress Biology (in preparation), Hebei Normal University of Science & Technology, Qinhuangdao, Hebei 066004, China

*** Correspondence:**

Corresponding Author: Jinshuang Zheng

email: jinshuangk@163.com

Abstract

Background: *Atractylodes chinensis* (DC.) Koidz is a well-known medicinal plant containing the major bioactive compound, atrctylodin, a sesquiterpenoids. High-performance liquid chromatography (HPLC) analysis demonstrated that atrctylodin was most abundant in 3-year old *A. chinensis* rhizomes, compared with those from 1-year and 2-year-old plants, however, the molecular mechanisms underlying accumulation of atrctylodin in rhizomes are poorly understood.

Results: In this study, we characterized the transcriptomes from 1-, 2, and 3-year old (Y1, Y2, and Y3, respectively) *A. chinensis*, to identify differentially expressed genes (DEGs). We identified 205 and 226 unigenes encoding the enzyme genes in the mevalonate (MVA) and methylerythritol phosphate (MEP) sesquiterpenoid biosynthesis pathways, respectively. To confirm the reliability of the RNA sequencing analysis, eleven genes key genes encoding factors involved in the sesquiterpene biosynthetic pathway, as well as in pigment, amino acid, hormone, and transcription factor functions, were selected for quantitative real time PCR (qRT-PCR) analysis. The results demonstrated similar expression patterns to those determined by RNA sequencing, with a Pearson's correlation coefficient of 0.9 between qRT-PCR and RNA-seq data. Differential gene expression analysis of samples from different ages revealed 52 genes related to sesquiterpenoids biosynthesis. Among these, seven DEGs were identified in Y1 vs Y2, Y1 vs Y3, and Y2 vs Y3, of which five encoded four key enzymes, squalene/phytoene synthase, squalene-hopene cyclase, squalene epoxidase and dammarenediol II synthase. These four enzymes directly related to squalene biosynthesis and subsequent catalytic action. To validate the result of these seven DEGs, qRT-PCR was performed and indicated most of them displayed lower relative expression in 3-year old rhizome, similar to transcriptomic analysis.

Conclusion: The enzymes SS, SHC, SE and DS down-regulated expression in 3-year old rhizome. This data corresponded to the higher content of sesquiterpenes in 3-year old rhizome, and confirmed by qRT-PCR. The results of comparative transcriptome analysis and identified key enzyme genes laid a solid foundation for investigation of production sesquiterpenes in *A. chinensis*.

Key words: *Atractylodes chinensis* (DC.) Koids., transcriptome sequencing, differentially expressed genes, bioactive compounds, sesquiterpene biosynthesis, qRT-PCR

1. Introduction

Atractylodes lancea and *Atractylodes. chinensis* (typically referred to as "Mao Cang Zhu" and "Bei Cang Zhu" in Chinese), together constitute the rhizome Atractylodes, and belong to the Asteraceae family. This rhizome Atractylodes are widely used in East Asia, and have great economic and medicinal value. *A. lancea* is currently on the verge of extinction, therefore, *A. chinensis* is the main source of the rhizome Atractylodes that widely distributed in most areas of Northern China. The main bioactive compounds in of *A. chinensis* rhizome are used to treat digestive disorders, rheumatic diseases, and night blindness (Committee, 2020). Modern pharmacological studies have demonstrated that *A. chinensis* also has anti-inflammatory, anti-bacterial (Hossen et al., 2019; Lyu et al., 2019), and antitumor (Ishii et al., 2020) properties. Although the sesquiterpene components of *A. chinensis* have important pharmacological activities, the molecular mechanism underling accumulation of bioactive sesquiterpenes compounds are poorly understood. In plants, sesquiterpenes are generally synthesized via MVA and MEP biosynthetic pathways.

Natural populations of *A. chinens*is currently being rapidly depleted, due to heavy use and weak reproductive capacity of perennial herbs. Thus, artificial cultivation is urgently needed to protect the natural populations and ensure sustainable utilization. A crucial question is how to ensure, or even improve, rhizome quality, in terms of sesquiterpene content. Although the phytochemistry (Xu et al., 2016; Hossen et al., 2019), pharmacology (Shimato et al., 2018; Kim et al., 2018, Cheng et al., 2019; Lyu et al., 2019), and cultivation (Xu et al., 2018; Sun et al., 2019; Zheng et al., 2018; 2019) of *A. chinens*is have been studied, the molecular mechanisms underlying their accumulation of bioactive compounds remains unclear, largely due to a lack of genomic and transcriptomic data.

Transcriptome analysis is an effective approach for analysis of secondary

metabolite biosynthesis and has been used to determine the functions of genes in medicinal plants, including Danshen (*Salvia miltiorrhiza*) (Yang et al., 2013), Renshen (*Panax ginseng*) (Chen et al., 2011), Sanqi (*Panax notoginseng*) (Liu et al., 2015), and Yunnan chonglou (*Paris polyphylla* var. *yunnanensis*) (Gao et al. 2020) among others. Recently, understanding of the molecular processes involved in sesquiterpene biosynthesis has improved, with various genes involved in this biosynthetic pathway investigated by transcriptome analysis in the genus *Atractylodes* (Huang et al., 2016; Ahmed et al., 2016, Zhao et al., 2020). Sesquiterpenes are the main bioactive components in the rhizomes of *A. lancea* and *A. chinensis*, however, there are differences between them in the composition and content of sesquiterpenes. In addition, the content of bioactive components in perennial medicinal herbs is influenced by the year of cultivation (Kong et al., 2017; Agnieszka et al., 2018; Wang and Li, 2018). To date, one study has reported the transcriptome of 3-year old *A. chinens*is rhizome (Zhao et al., 2020), however, there are no data regarding the molecular mechanism involved in the relationship between sesquiterpene accumulation and year of cultivation. Elucidating factors involved in the biosynthesis and accumulation of bioactive components and identifying key enzyme genes in the biosynthetic pathway will be important steps toward improvements in sesquiterpenes production.

Here, rhizomes from 1-, 2- and 3-year old *A. chinensis* were subjected to high throughout transcriptome sequencing, enabling us to characterize the transcriptomes and differential expression profiles of *A. chinensis* rhizomes cultivated for different ages, to profile differentially expressed genes (DEGs) among rhizomes from different years of cultivation, and to identify DEGs related to biosynthesis and accumulation of sesquiterpenes. Discovering the key enzyme genes in the sesquiterpene biosynthetic pathway is necessary to improve atractylodin production. This study could provide insights into the relationship between changes in atrctylodin content and year of cultivation, and contribute to uncovering the underlying molecular mechanisms in *A. chinensis*.

2. Materials and Methods

2.1 Plant materials

A. chinensis seeds were collected from cultivation base of Qinhuangdao Tongsheng Pharmaceutical Co., Ltd, Qinhuangdao City, Hebei Province, China. To ensure a similar physical environment, seeds were sown separately in 2016, 2017 and 2018, in the same open field at the Hebei Normal University of Science & Technology.

A. chinensis rhizome, which is the part of the plant used in medicinal preparations, serves as a store for photosynthetic products and bioactive compounds. For use as a medicine, *A. chinensis* is optimally harvested during the 3rd to 4th withering period, therefore, 1-, 2- and 3-year old rhizomes were collected during this period (as seen in Fig. 7). After collection, rhizomes were cleaned in running water, then immediately frozen in liquid nitrogen and stored at -80 °C prior to RNA extraction.

2.2 Atractylodin extraction and HPLC analysis

Samples were dried at 60 °C and then ground into a powder, 0.2 g of which was immersed in 50 mL methanol (purity≥99.9%, Grade/Application information: ACS. Reag. Ph Eur, CSA-No. 67-56-1) and ultrasonically extracted (Power 250W, Frequency 40kHz) for 1h. Next, 1mL of the supernatant was collected and passed through a 0.22μm microporous filter membrane (JTSF0311, Tianjin Jinteng Experiment Equipment Co. Ltd.).

Determination of atractylodin content was conducted using a Thermo Fisher UltiMate 3000 UPLC system, equipped with a Uv-vis detector, on C18 Column (4.6 × 250 mm, 5μm, Thermo Fisher). The mobile phase was methanol:water (79:21), with a flow rate of 1.0 $mL \cdot min^{-1}$. The HPLC chromatogram was monitored at 340 nm, and the column temperature was set at 30°C. Atractylodin content was determined by comparison with authorized standards (gbw114.com, China).

2.3 RNA sequencing and functional annotation of unigenes

To extract total RNA, three replicates of rhizomes from 1-, 2- and 3-year old *A. chinensis* were extracted using TRIzol Reagent (Invitrogen), then treated with DNase I (TaKaRa). RNA quality was tested by 1% agarose gel electrophoresisa and the concentration determined using Nanodrop spectrophotometer (Thermo). Rhizome RNA pools were prepared by mixing equal amounts of the RNA replicates. Transcriptome data were acquired using based on the Illumina HiSeqTM 2000 150PE platform, by Novogene Co. (Beijing, China). Clean reads were assembled de novo using the Trinity program. The datasets generated and analyzed during the current study are available in the Sequence Read Archive (SRA) repository (https://www.ncbi.nlm.nih.gov/sra/PRJNA698794).

For functional annotation, unigenes were searched against public databases, including Nr, Swiss-Prot, Pfam, GO, COG, and KEGG.

2.4 Analysis of differentially expressed genes

Analysis of differential expression between two assigned libraries was performed using the DESeq R package (1.10.1). DESeq provides statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P values were adjusted using the Benjamini and Hochberg's approach to control the false discovery rate. Genes with an adjusted P-value <0.05 according to DESeq were considered differentially expressed. Subsequently, GO functional enrichment analysis and KEGG pathway analysis of DEGs were performed using http://www.geneontology.org/ and Ipath version3 (https://pathways.embl.de/), respectively.

Differential expression of unigenes among rhizomes from the three *A. chinensis* year old plants was determined using edgeR software. Differences in gene expression were evaluated using the chi-square test and the false discovery rate (FDR) was controlled. Genes with an FDR <0.001 and for which the Fragments Per Kilobase of transcript per Million mapped reads (FPKM) estimate was 2-fold higher than that of the one with the lowest value, were identified as DEGs. GO enrichment annotation of DEGs was conducted using the GO TermFinder software (version v0.86). Corrected *P*-value ≤ 0.05 or Q-value ≤ 0.05 were used as thresholds for "enriched" DEGs. Pathfinder Internal software was used to assess the significance of the enrichment of DEGs in KEGG pathways. Heat maps were generated to display genes with significantly altered expression at the three stages. Raw intensity data (FPKM) were log2 transformed and used for calculation of Z scores.

2.5 Quantitative real-time PCR

To confirm the reliability of the RNA sequencing analysis, qRT-PCR analyses were performed using samples from the same 1-, 2- and 3-year old rhizomes as used for RNA-seq. Eleven genes (cluster-15114.3, cluster-8388.71372, cluster-8388.203329, cluster-388.168445, cluster-8388.64828, cluster-8388.299573, cluster-8388.162261, cluster-8388.157231, cluster-8388.172353, cluster-8388.295361, and cluster-8388.295722), with key functions in sesquiterpene biosynthetic pathway, as well as in pigment, amino acid, hormone, and transcription factor functions, were randomly selected for qRT-PCR analysis. Primers for qRT-PCR were designed using Primer v5.0 and synthesized by Sangon Biotech (Shanghai, China) Co., Ltd. The experimental methods used for qRT-PCR and relative expression analysis were as reported by Huang et al. (2016). Relative expression data were normalized to those of the *UBQ2* gene,

which was used as an internal control (Zhao et al., 2020). All primers used are presented in Table S1.

Validation of the seven DEGs related to sesquiterpene biosynthesis using qRT-PCR according to method of reliability confirmation for the RNA sequencing. The primers are presented in Table S2.

3. Results

3.1 The atrctylodin content in 1-, 2-, and 3-year old *A. chinensis* rhizomes

As atrctylodin is the main and index active constituent in *A. chinens*is, its levels in rhizome from 1-, 2-, and 3-year old *A. chinens*is plants were measured by HPLC analysis, with atrctylodine contents (%) recorded as 0.2252, 0.2378 and 0.2939, respectively (Table 1). These data showed that cultivation year had marked effect on atrctylodin content of *A. chinensis* rhizome, however, the molecular mechanisms underlying the higher atrctylodin content in 3-year old rhizome is unclear.

3.2 Sequencing analysis and de novo assembly

To study the molecular mechanisms involved in the relationship between increased atrctylodin content and *A. chinensis* cultivation year, transcriptome sequencing was conducted. Total RNA was extracted from 1-, 2-, and 3-year old *A. chinens*is rhizome, and the mRNA was isolated. Each sample was sequenced using the Illumina HiSeq™ 2000 platform. After filtering out adapter sequences and reads ≤ 50 bp, 29,197,007, 25,791,735, and 29,553,752 high-quality (HQ) reads were obtained from 1-, 2-, and 3-year old *A. chinensis* rhizomes, respectively. Reads from the three samples were also pooled and the above steps repeated, resulting in identification of 416,846 unigenes (mean length 825 bp, N50 length 1121 bp). The GC content of the reads and unigenes was in the range 45.21%–45.58% (Table 2). Analysis of length distribution demonstrated that 19% unigenes were > 1kb. The transcriptome data have been uploaded to SRA (BioProject ID PRJNA698794, https://www.ncbi.nlm.nih.gov/sra/PRJNA698794).

3.3 Functional annotation and classification

Unigene sequences were searched against public databases (Nr, Swiss-Prot, Pfam, COG, GO, and KEGG) using the BLAST program, with an E-value cut-off of 1.0 e^{-5}. A total of 56,759 unigenes (39.52% of the total assembled unigenes) had matches in the Nr database, with 37,475 (26.09%), 31,272 (21.77%), 6,540 (4.55%), 39,372 (27.41%) and 31,424 (14.92%) unigenes showing significant similarity to sequences in the Swiss-Prot, Pfam, COG, GO and KEGG databases, respectively (Table 3).

GO functional categorization of 39,372 unigenes mapped to at least one GO term, including 57,008 “biological process”, 55,906 “cellular component”, and 47,730 “molecular function” (Figure 1A). In total, 6,540 unigenes were annotated and grouped into 24 COG classifications (Figure 1B), among which, the cluster for “translation, ribosomal structure and biogenesis” (n = 464, 7.10%) accounted for the largest proportion, followed by “posttranslational modification, protein turnover, chaperones” (n = 377, 5.80%) and “general function prediction only” (n = 373, 5.70%).

KEGG pathway analysis was performed to functionally classify biochemical pathways active in 1-, 2-, and 3-year old *A. chinensis* rhizome. A total of 31,424 unigenes were assigned to 130 KEGG categories: “cellular processes”, “environmental information processing”, “genetic information processing”, “metabolism”, and “organismal systems pathways” was presented in Figure 2.

Of KEGG secondary metabolic pathways, most unigenes were assigned to “phenylpropanoid biosynthesis” (n = 723, ko00940), “terpenoid backbone biosynthesis” (n = 517, ko00900), “carotenoid biosynthesis” (n = 394, ko00906), and “sesquiterpenoid and triterpenoid biosynthesis” (n = 204, ko00909) (Table 4). A total of 205 unigenes were identified as key enzyme genes in the MVA pathway, with 226 unigenes in the MEP pathway (Table 5). The discovery of these genes related to sesquiterpene biosynthetic pathways may help us to elucidate the molecular mechanisms underlying the higher atrctylodin content in 3-year old rhizomes.

3.4 Differential expression of transcripts in *A. chinensis* rhizomes from different cultivation year

To compare the unigenes from different age *A. chinens*is rhizomes, a Venn diagram was constructed (Figure 3). The results showed that 31,895 (50.51%) unigenes were shared by all three samples. A total of 7,027, 8,879 and 6,109 unigenes were specific to 1-, 2-, and 3-year old *A. chinens*is rhizomes, respectively, with the 2-year-old *A. chinens*is rhizome having the highest number of unique unigenes.

To identify DEGs among the three samples, the tag frequencies of 1- vs 2-year-old (Y1 vs Y2) rhizome, 2- vs 3-year old (Y2 vs Y3) rhizome and 1- vs 3-year old (Y1 vs Y3) rhizomes were assessed, with 6,424, 3,464, and 2,869 DEGs detected between the three pairs, respectively (Figure 4). Y1 vs Y2, Y2 vs Y3, and Y1 vs Y3, revealed 3,880, 2,214, and 2,292 up-regulated genes. There were more up-regulated than down-regulated genes in Y1 vs Y2, with the opposite detected in the Y2 vs Y3, and Y1 vs Y3 comparisons.

KEGG pathway enrichment analysis of all DEGs was performed to characterize the complex biological behaviors in three samples. The 19 enriched pathways are presented in Fig 5 and reflected the preferential biological functions of samples from different age plants. Hierarchical clustering of all DEGs indicated that overall unigene enrichment characteristics were similar between the Y1 vs Y2, Y1 vs Y3 rhizomes (Fig 5A and 5B), with genes involved in "Carbohydrate metabolism", "Signal transduction", "Amino acid metabolism", "Lipid metabolism" and "Biosynthesis of secondary metabolites" over-expressed. In Y2 vs Y3, genes involved in "Lipid metabolism", "Amino acid metabolism", "Biosynthesis of secondary metabolites" and "Replication and repair" were overexpressed (Fig 5C).

Pathways involved in bioactive metabolism are of particular interest in medicinal plants. DEGs involved in "Biosynthesis of secondary metabolites" were overexpressed in all three samples; 52 genes related to sesquiterpene biosynthesis were detected, of which seven were differentially expressed in Y1 vs Y2, Y1 vs Y3, Y2 vs Y3 (Fig 6).

Heatmap trees were constructed based on gene expression levels, to further investigate the seven differentially expressed sesquiterpene biosynthesis and triterpenoids genes, including NAD-dependent epimerase/dehydratase (NDE), squalene/phytoene synthase (SS), squalene-hopene cyclases (SHC), squalene epoxidase (SE), dammarenediol-II synthase (DS), and serine/threonine-protein kinase SRK2E (SPK). All of these seven DEGs down-regulated expression in 3-year old rhizome, comparing with 1- and 2-year old samples (Fig. 7). Notably, five of the differentially expressed genes encoded 4 key enzymes: SS, SHC, SE and DS. Isoprenenyl-PP is synthesized from isoprenenyl through MVA or MEP pathway, then was catalyzed toward two biosynthesis branch pathway, sesquiterpenes biosynthesis and triterpenes biosynthesis. These four enzymes, SS, SHC, SE and DS, directly related to squalene biosynthesis and subsequent catalytic action. According to putative pathway, squalene is the first precursor in triterpenoid biosynthesis pathway (Fig. 8).

3.5 Validation of RNA-seq analysis by qRT-PCR

To confirm the reliability of the RNA sequencing analysis, eleven genes representing key genes in sesquiterpene biosynthesis pathways, as well as in pigment, amino acid, hormone, and transcription factor functions, were selected for qRT-PCR analysis. The result demonstrated similar expression patterns to those determined by RNA sequencing, with a Pearson's correlation co-efficient between qRT-PCR and RNA-seq data of 0.9 (Fig. 9).

Further validation of seven DEGs, NDE, SS, SHC, SE, DS, and SPK, related to sesquiterpene biosynthesis pathways was performed by qRT-PCR. The relative expression levels of these seven DEGs noted in 3-year old rhizome were significantly lower than those in 1- and 2-year old rhizome (Fig. 10). These results are consistent with the data of transcriptomic sequencing analysis.

4. Discussion

As genome data for the *Atractylodes* genus are not yet available, Illumina-based RNA sequencing was performed to characterize the *A. chinensis* transcriptome. We obtained 41,6846 unigene sequences, of which 40.71% could be functionally annotated based on public databases. In addition, the qRT-PCR results demonstrated similar expression patterns of eleven randomly selected genes to those determined by RNA-seq analysis, demonstrating the reliability of our *A. chinensis* transcriptome data.

Transcriptomic analysis to investigate sesquiterpene accumulation patterns in different tissues of *A. lancea* discovered 69 unigenes in the MVA pathway, including nine key enzymes, and 28 unigenes in the MEP pathway, involving seven key enzymes (Chen et al., 2017). In this study, we investigated the sesquiterpene accumulation patterns in *A. chinensis* after different cultivation year and discovered 205 unigenes in the MVA pathway, involving eleven key enzymes, and 226 unigenes in the MEP pathway, involving eleven crucial enzymes. These data will facilitate further study of the molecular mechanisms underlying sesquiterpene accumulation.

In this study, we found that atractylodin content in *A. chinensis* rhizome with the increase of cultivation year. Li et al. (2019) confirmed that the year of cultivation medicinal plants was important in increasing saponins production in *Panax notoginseng* rhizomes. Based on this natural phenomena, we performed differential expression analysis using transcriptome data from 1-, 2-, and 3-year old rhizomes, to identify candidate DEGs encoding key enzymes in sesquiterpene biosynthetic pathways. Differential gene expression patterns were further investigated to profile global gene expression differences between Y1 vs Y2, Y2 vs Y3, and Y1 vs Y3. Most DEGs between Y1 vs Y2 and Y1 vs Y3 were assigned to 19 metabolic pathways, including signal transduction, primary metabolic pathways (carbohydrate metabolism, amino acid metabolism and lipid metabolism), and biosynthesis of other secondary metabolites. In Y2 vs Y3, DEGs were assigned to 10 metabolic pathways, of which lipid metabolism, amino acid metabolism, replication and repair, and biosynthesis of other secondary metabolites comprised a higher percentage. These data indicate that the metabolic

characteristics of 2-year old rhizome are more similar to those of 3-year old rhizome, relative to 1-year old rhizome. Further, the metabolic characteristics of DEGs were consistent with the rhizome's physiological function as a storage organ for photosynthetic products and bioactive compounds. These data demonstrated that vitality of medicinal plants and the production of secondary metabolic became increased over the cultivation year, likely because they are crucial for defense against stress in older plants.

Further analysis of DEGs provided information crucial for investigation of the molecular mechanisms involved in sesquiterpene biosynthesis and accumulation in *A. chinensis*. Seven key genes related to sesquiterpene and triterpenoid biosynthesis were discovered by analysis DEGs between Y1 vs Y2, Y1 vs Y3, and Y2 vs Y3. Of the seven key genes, five encoding four enzymes: squalene epoxidase (SE), squalene-hopene cyclases (SHC), squalene/phytoene synthase (SS) and dammarenediol-II synthase (DS). The biological production of sesquiterpenes and triterpenes is an extremely complicated process, with synthesis occurring via MEP and MVA pathway. Many enzymes are involved in the process of isoprenenyl-PP biosynthesis catalysis, which was then catalyzed toward two biosynthesis branch, sesquiterpenes biosynthesis and squalene biosynthesis.

The identified four enzymes, SE, SHC, SS and DS, play important role in squalene biosynthesis and the subsequent catalytic reactions of this metabolic branch. The enzyme SS as a key enzyme in the terpenoid biosynthesis pathway catalyzes the synthesis of the first precursor of terpenoid compounds, squalene (Kim et al., 2005; Ye et al., 2014; Zheng et al., 2013; Dan et al., 2017; Shao et al., 2020). The SHC enzyme can catalyze the formation of hopene from its precursor squalene (Siedenburg and Jendrossek, 2011; Nakano et al., 2019), toward triterpenoid or steroid biosynthesis. The enzyme, SE, which catalyzes the oxidation of squalene to 2, 3-oxysqualene, is a rate-limiting enzyme in the sterol biosynthesis (Wentzinger 2002). DS was the first dedicated enzyme for ginsenoside biosynthesis, one of triterpenoid compounds (Tansakul et al., 2006).

In the case of suppression of enzyme SS activity was observed induction of sesquiterpene cyclase, toward the synthesis of sesquiterpenes (Zook and Kuć, 1991). The enzymes SS and SE catalyze the first two steps involved in sterol biosynthesis. Inhibition of either SS or SE was found to trigger a severalfold increase in enzyme activity of HMGR (Wentzinger et al. 2002). This study revealed that the four enzymes

SS, SHC, SE and DS down-regulated expression in 3-year old rhizome. This data corresponded to the higher content of sesquiterpene in 3-year old rhizome, and confirmed by qRT-PCR. This study reported the results of comparative transcriptome analysis and identified key enzyme genes, laid a solid foundation for investigation of production sesquiterpene in *A. chinensis*.

Author Contributions

JSZ, CZS and FYS: design this experiment. JHZ, SSM, LPZ and XD: analysis the data. SCZ do the qRT-PCR experiment. JHZ: write this manuscript. JSZ: revise this manuscript.

Funding

This work was supported financially by National Fund of Hebei Province, China (Project No. H2019407120), which provide support for the test of transcriptomic sequencing, Key Research and Development Project of Hebei Province, China (Project No.19226354D) and Science and Technology Project Hebei Education Department, China (Project No. QN2019162), which provide support for the test of atrctylodin. There is no role of the funding body in the design of the study.

Acknowledgments

Thanks for the help from laboratory of college of horticulture science and technology, Hebei Normal University of Science & Technology in atractylodin extraction and HPLC analysis.

Availability of data

The raw data was uploaded to Sequence Read Archive (PRJNA698794).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

Agnieszka, G., Mariola, D., Anna, P., Piotr, K., Natalia, W., Aneta, S., *et al*. (2018). Qualitative and quantitative analyses of bioactive compounds from ex vitro *Chamaenerion angustifolium* (L.) (*Epilobium augustifolium*) herb in different harvest times. *Ind. Crop. Prod*. 123, 208–220.

Ahmed, S., Zhan, C.S., Yang, Y.Y., Wang, X.K., Yang, T.W., Zhao, Z.Y., *et al*. (2016). The transcript profile of a traditional chinese medicine, *Atractylodes lancea*, Revealing its sesquiterpenoid biosynthesis of the major active components. *Plos One*. 11(3), e0151975.

Chen, F., Wei, Y.X., Zhang, J.M., Sang, X.M., Dai, C.C. (2017). Transcriptomics analysis investigates sesquiterpenoids accumulation pattern in different tissues of *Atractylodes lancea* (Thunb.) DC. plantlet. *Plant Cell Tiss. Organ. Cult.*, 130, 73-90.

Chen, S., Luo, H., Li, Y., Sun, Y., Wu, Q., Niu, Y., *et al*. (2011). 454 EST analysis detects genes putatively involved in ginsenside biosynthesis in *Panax ginseng*. *Plant Cell Rep*. 30(9), 1593-1601.

Cheng, Y., Chen, T.Y., Yang, X.L., Xue, J.H., Chen, J.J. (2019). Atractylon induces apoptosis and suppresses metastasis in hepatic cancer cells and inhibits growth in vivo. *Cancer Manage. Res*. 11, 5883-5894.

Committee, S.P. (2020). Pharmacopoeia of the People's Republic of China. Beijing: People's Medical Publishing House

Dan, J., Rong, Q.X., Chen, Y.J., Yuan, Q.J., Ye, S., Guo, J., *et al*. (2017). Molecular cloning and functional analysis of squalene synthase (SS) in *Panax notoginseng*. *Int. J. Biol. Macromol*. 95, 658–666.

Gao, X.Y., Zhang, X., Chen, W., Li, J., Yang, W.J., Zhang, X.W., *et al*. (2020). Transcriptome analysis of *Paris polyphylla* var. *yunnanenesis* illuminates the biosynthesis and accumulation of steroidal saponins in rhizomes and leaves. *Phytochemistry*. 178, 112460.

Hossen, M.J., Chou, J.Y., Li, S.M., Fu, X.Q., Yin, C., Guo, H., *et al*. (2019). An ethanol extract of the rhizome of *Atractylodes chinensis* exerts anti gastritis activities and inhibits Akt/NF-kappa B signaling. *J. Ethnopharmacol*. 228, 18-25.

Huang, Q.Q., Huang, X., Deng, J., Liu, H.G., Liu, Y.W., Yu, K., *et al*. (2016). Differential gene expression between leaf and rhizome in *Atractylodes lancea*: a comparative transcriptome analysis. *Front. Plant Sci*. 7, 348.

Ishii, T., Okuyama, T., Noguchi, N., Nishidono, Y., Okumura, T., Kaibori, M., *et al*. (2020). Antiinflammatory constituents of *Atractylodes chinensis* rhizome improve glomerular lesions in immunoglobulin A nephropathy model mice. *J. Nat. Med*. 74(3), 616-616.

Kim, J.K., Doh, E.J., Lee, G. (2018). Chemical differentiation of genetically identified *Atractylodes japonica*, *A. macrocephala*, and *A. chinensis* rhizomes using high-performance liquid chromatography with chemometric analysis. *J. Evidence-based complementary Altern. Med*. 2018, 4860371.

Kim, O.T., Seong, N.S., Kim, M.Y., Hwang, B. (2005). Isolation and characterization of squalene synthase cDNA from *Centella asiatica* (L) urban. *J. Plant Biol*. 48, 263-269.

Kong, D., Li, Y., Bai, M., Deng, Y., Liang, G., Wu, H. (2017). A comparative study of the dynamic accumulation of polyphenol components and the changes in their antioxidant activities in diploid and tetraploid *Lonicera japonica*. *Plant Physiol. Biochem*. 112, 87–96.

Li, J., Ma, L., Zhang, S.T., Zuo, C.L., Song, N., Zhi, S.S., Wu, J.S. (2019). Transcriptome analysis of 1-and 3-year-old *Panax notoginseng* rhizomes and functional characterization of saponin biosynthetic genes *DS* and *CYP716A47 like*. *Planta*. 249, 1229-1237.

Lyu, Z., Ji, X.F., Chen, G., An, B.Y. (2019) Atractylodin ameliorates lipopolysaccharide and D-galactosamine-induced acute liver failure via the suppression of inflammation and oxidative stress. *Inter. Immunopharmacol*. 72, 348-357.

Nakano, C., Watanabe, T., Minamino, M., Hoshino, T. (2019). Enzymatic syntheses of novel carbocyclic scaffolds with a 6,5 + 5,5 ring system by squalenehopene cyclase. *Org. Biomol. Chem*. 17, 9375–9389.

Shao, C.M., Wang, C.K., Zhang, S.X., Shi, Y.Y., MA, K.L., Yang, Q.S., *et al*. (2020). Transcriptome analysis of *Clinopodium gracile* (Benth.) Matsum and identification of genes related to Triterpenoid Saponin biosynthesis. *BMC Genomics*. 21, 49.

Shimato, Y., Ota, M., Asai, K., Atsumi, T., Tabuchi, Y., Makino, T. (2018). Comparison of byakujutsu (*Atractylodes rhizome*) and sojutsu (*Atractylodes lancea* rhizome) on anti-inflammatory and immunostimulative effects in vitro. *J. Nat. Med*. 72, 192-201.

Sun, W.M., Wen, X.L., Qi, H.X., Feng, L.N., Cao, J., Han, Z.L., *et al*. (2019). First Report of Anthracnose of *Atractylodes chinensis* (DC.) Koidz. caused by *Colletotrichum chlorophyti* in China. *Plant disease*. 103(4), 764-764.

Wang, Y.Z., Li, P. (2018). Effect of cultivation years on saponins in *Paris Polyphylla* var. *yunnanensis* using ultra-high liquid chromatography–tandem mass spectrometry and Fourier transform infrared spectroscopy. *Plant Growth Regul.* 84, 373–381.

Wentzinger, L.F., Bach, T.J., Hartmann, M.A. (2002). Inhibition of squalene synthase and squalene epoxidase in tobacco cells triggers an up-regulation of 3-Hydroxy-3-Methylglutaryl coenzyme A reductase. *Plant Physiol.*, 130(1), 334-346.

Xu, H.J., Zhou, R.J., Fu, J.F., Yuan, Y., Ge, X.X., Damm, U. (2018) Colletotrichum atractylodicola sp nov.: the anthracnose pathogen of *Atractylodes chinensis* in China. *Mycological progress*. 17(3), 393-402.

Xu, J., Chen, D., Liu, C., Wu, X.Z., Dong, C.X., Zhou, J. (2016) Structural characterization and anti-tumor effects of an inulin-type fructan from *Atractylodes chinensis*. *Int. J. Boil. Macromol*. 82, 765-771.

Yang, L., Ding, G., Lin, H., Cheng, H.N., Kong, Y., Wei, Y.K., *et al*. (2013). Transcriptome analysis of medicinal plant *Salvia miltiorrhiza* and identification of genes related to tanshinone biosynthesis. *Plos One*. 8, e80464.

Ye, Y., Wang, R.F., Jin, L., Shen, J.H., Li, X.T., Yang, T., *et al*. (2014). Molecular cloning and differential expression analysis of a squalene synthase gene from *Dioscorea zingiberensis*, an important pharmaceutical plant. *Mol. Biol. Rep*. 41, 6097–6104.

Zhao, J.H., Zhao, C.Y., Sun, C.Z., Shi, F.Y., Chen, L.N., Zheng, J.S. (2020). Transcriptomic analysis of *Atractylodes chinensis* and elucidation of genes in sesquiterpenes biosynthesis. *Plant Physiol. J*. 56(7), 1458-1466. (in Chinses with English abstract, doi: 10.13592/j.cnki.ppj.2019.0560)

Zheng, J.S., Wang, W.P., Li, Y.S. (2018). Effects of temperature and substrate water content on seed germination and seedling morphogenesis of *Atractylodes chinensis*. *J. Chinses Medi. Materi*. 41(6), 1282-1284. (in Chinses with English abstract, doi: 10.13863/j.issn1001-4454.2018.06.007)

Zheng, J.S., Wang, W.P., Wu, Y.G. (2019). Effects of different sowing depth and seedling substrate on seed emergence of *Atractylodes chinensis*. *J. Chinses Medi. Materi*. 41(12), 2501-2504. (in Chinses with English abstract, doi: 10.13863/j.issn1001-4454.2018.12.005)

Zheng, Z.J., Cao, X.Y., Li, C.G., Chen, Y.Q., Yuan, B., Xu, Y., *et al*. (2013). Molecular cloning and expression analysis of a squalene synthase gene from a medicinal plant *Euphorbia pekinensis. Rupr. Acta Physiol. Plant*. 35, 3007–3014.

Zook, M.N., Kuć, J.A. (1991). Induction of sesquiterpene cyclase and suppression of squalene synthase activity in elicitor-treated or fungal-infected potato tuber tissue. *Physiol Mol Plant Pathol.* 39, 377-390.

Siedenburg, G., Jendrossek, D. (2011). Squalene-hopene cyclases. *Appl. Environ. Microbiol.,* 77(12), 3905-3915

Tansakul, P., Shibuya, M., Kushiro, T., Ebizuka, Y. (2006). Dammarenediol-II synthase, the first

463 dedicated enzyme for ginsenoside biosynthesis, in *Panax ginseng*. *FEBS Lett.*, 580(22), 5143-
464 5149.
465
466

Additional files:

Table S1 Primers of qRT-PCR for validation of the reliability of RNA-seq analysis

Table S2 Primers of qRT-PCR for validation of the seven DEGs involved in sesquiterpenoid and triterpenoid biosynthesis pathway

473 Table 1 Atrctylodine content (%) in rhizome of 1-, 2- and 3-year old *A. chinens*is

Cultivation year	Content (%)
1	0.2252
2	0.2378
3	0.2939

474

475

476 Table 2 Summary of Illumina sequencing and assembly of *A. chinens*is

	1- year-old rhizome	2- year-old rhizome	3- year-old rhizome
Number of raw reads	59297549	52484111	60220581
Number of clean reads	58573338	51757902	59322720
Q30 (%)	92.15	9176	92.12
GC content (%)	45.55	45.21	45.58
Number of unigene[a]		416846	
Length of unigene (bp) [a]		343909490	
Average length of unigene (bp) [a]		825	
N50 of unigene (bp) [a]		1121	

477 [a] The total number of contigs and singletons.

478

479

Table 3 Summary of annotations on the unigenes in the *A. chinens*is rhizome transcriptome against public databases

Database	Unigenes	Percentage (%)
Nr	56,759	39.52
Swiss-Prot	37,475	26.09
Pfam	31,272	21.77
COG	6,540	4.55
GO	39,372	27.41
KEGG	31,424	14.92
Total annotation	58,466	40.71

485

486 Table 4 Number of unigenes related to secondary metabolites in *A. chinensis*

Secondary metabolites biosynthesis pathway	Number of unigenes	Pathway ID
Phenylpropanoid biosynthesis	723	ko00940
Terpenoid backbone biosynthesis	517	ko00900
Carotenoid biosynthesis	394	ko00906
Sesquiterpenoid and triterpenoid biosynthesis	204	ko00909
Tropane, piperidine and pyridine alkaloid biosynthesis	194	ko00960
Zeatin biosynthesis	170	ko00908
Monobactam biosynthesis	157	ko00261
Isoquinoline alkaloid biosynthesis	138	ko00950
Stilbenoid, diarylheptanoid and gingerol biosynthesis	116	ko00945
Limonene and pinene degradation	116	ko00903
Flavonoid biosynthesis	82	ko00941
Diterpenoid biosynthesis	71	ko00904
Monoterpenoid biosynthesis	69	ko00902
Caffeine metabolism	55	ko00232
Brassinosteroid biosynthesis	44	ko00905
Flavone and flavonol biosynthesis	38	ko00944
Glucosinolate biosynthesis	28	ko00966
Betalain biosynthesis	20	ko00965
Anthocyanin biosynthesis	8	ko00942
Isoflavonoid biosynthesis	4	ko00943
Indole alkaloid biosynthesis	3	ko00901

487

488

489 Table 5 Discovery of unigenes involved in sesquiterpene biosynthesis in *A. chinensis*

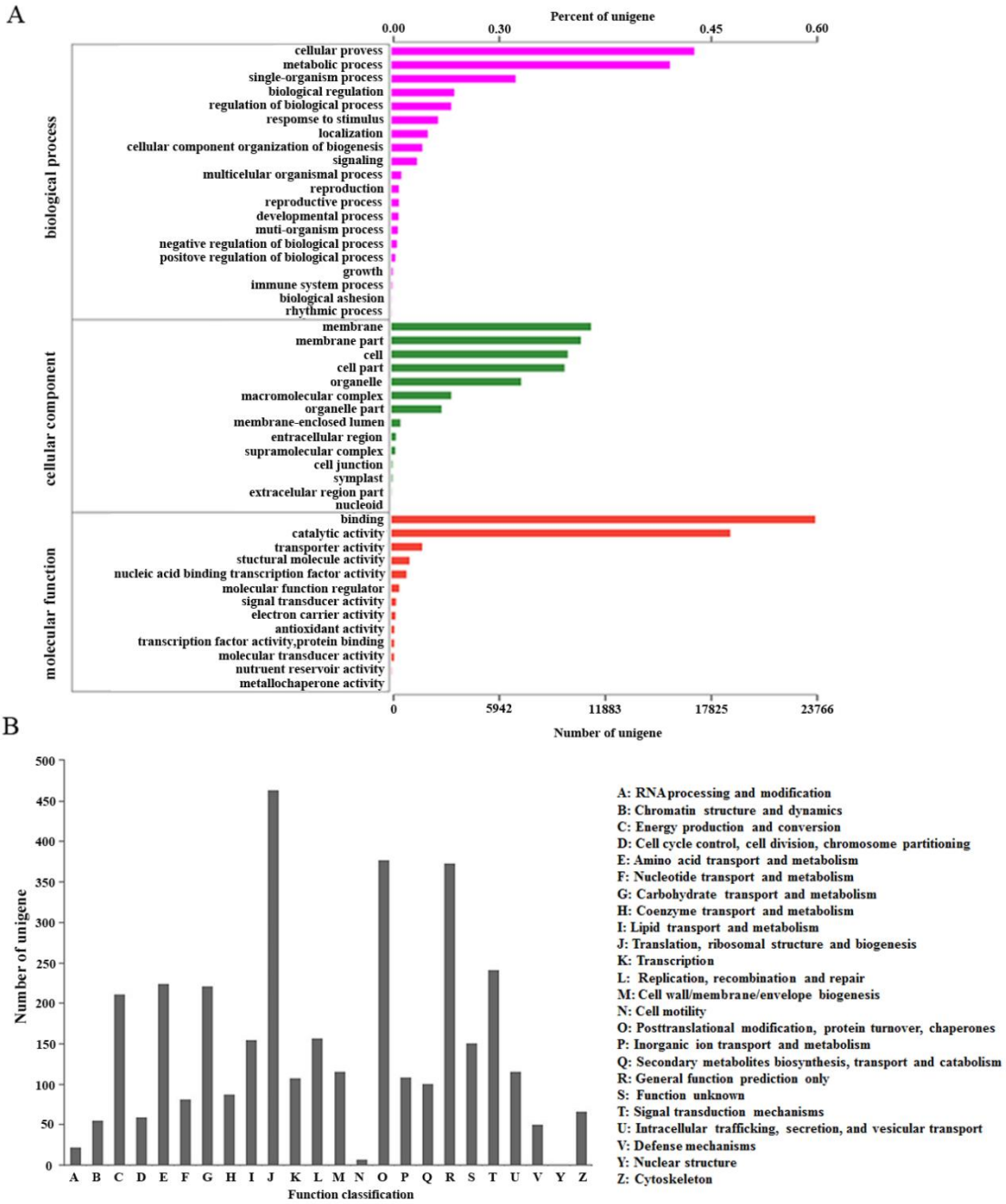
Pathway	Enzymes name	Abbreviation	Number of unigenes
MVA	Acetyl-CoA C-acetyltransferase	AACT	27
	3-hydroxy-3-methylglutaryl coenzyme A synthase	HMGS	35
	3-hydroxy-3-methylglutaryl coenzyme A reductase	HMGR	31
	mevalonate kinase	MK	2
	phosphomevalonate kinase	PMK	38
	Mevalonate-5-pyrophosphate decarboxylase	MDC	8
	Geranyl diphosphate synthase	GPPS	39
	Farnesyl diphosphate synthase	FPPS	10
	Beta-caryophyllene synthase	QHS1	5
	Germacrene D synthase	GDS	9
	Germacrene A synthase	GAS	1
MEP	1-deoxy-*D*-xylulose-5-phosphate synthase	DXPS	74
	1-deoxy-*D*-xylulose-5-phosphate reductoisomerase	DXR	22
	2-*C*-methyl-*D*-erythritol 4-phosphate cytidylyl transferase	MCT	22
	4-diphosphocytidyl-2-*C*-methyl-*D*-erythritol kinase	CMK	1
	4-hydroxy-3-methyl but-2-(*E*)-enyl-diphosphate synthase	HDS	36
	4-hydroxy-3-methyl but-2-(*E*)-enyl-diphosphate reductase	HDR	14
	Sesquiterpene synthase	TPS	7
	Isopentenyl-diphosphate delta-isomerase	IPPI	19
	Farnesyl diphosphate synthase	FDPS	10
	Mevalonate pyrophosphate decarboxylase	MVD	8
	Isopentenyl diphosphate isomerase	IDI	13

490

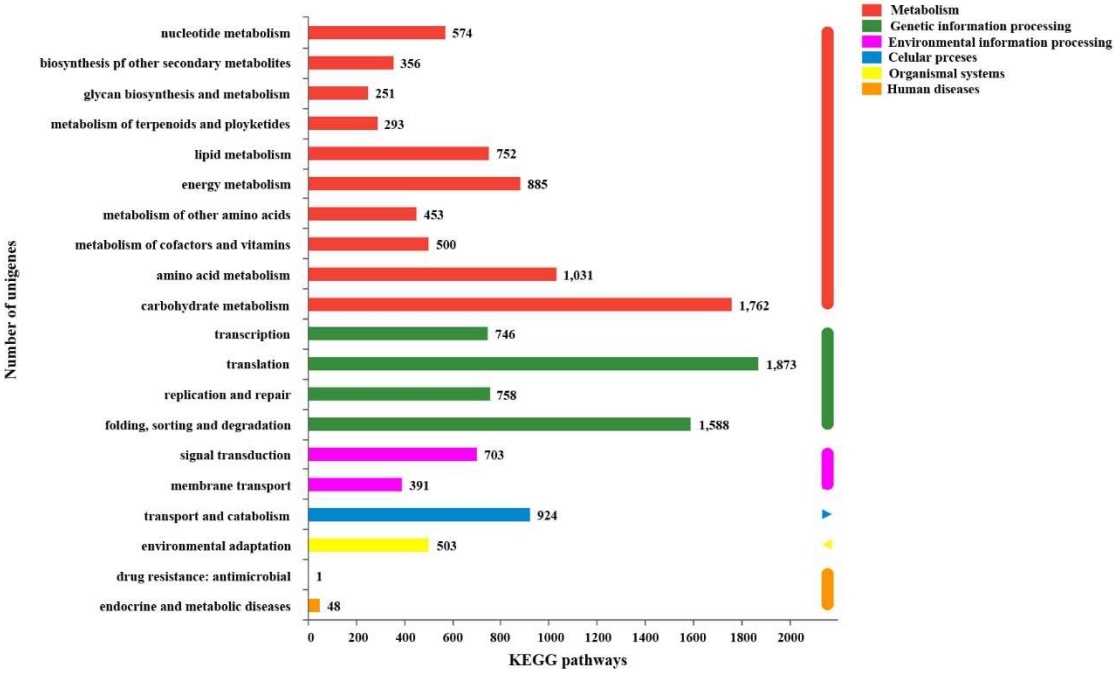
491

Figure agents

Fig. 1 GO and COG classification of assembled unigenes. (A) CO classification; (B) COG classification.

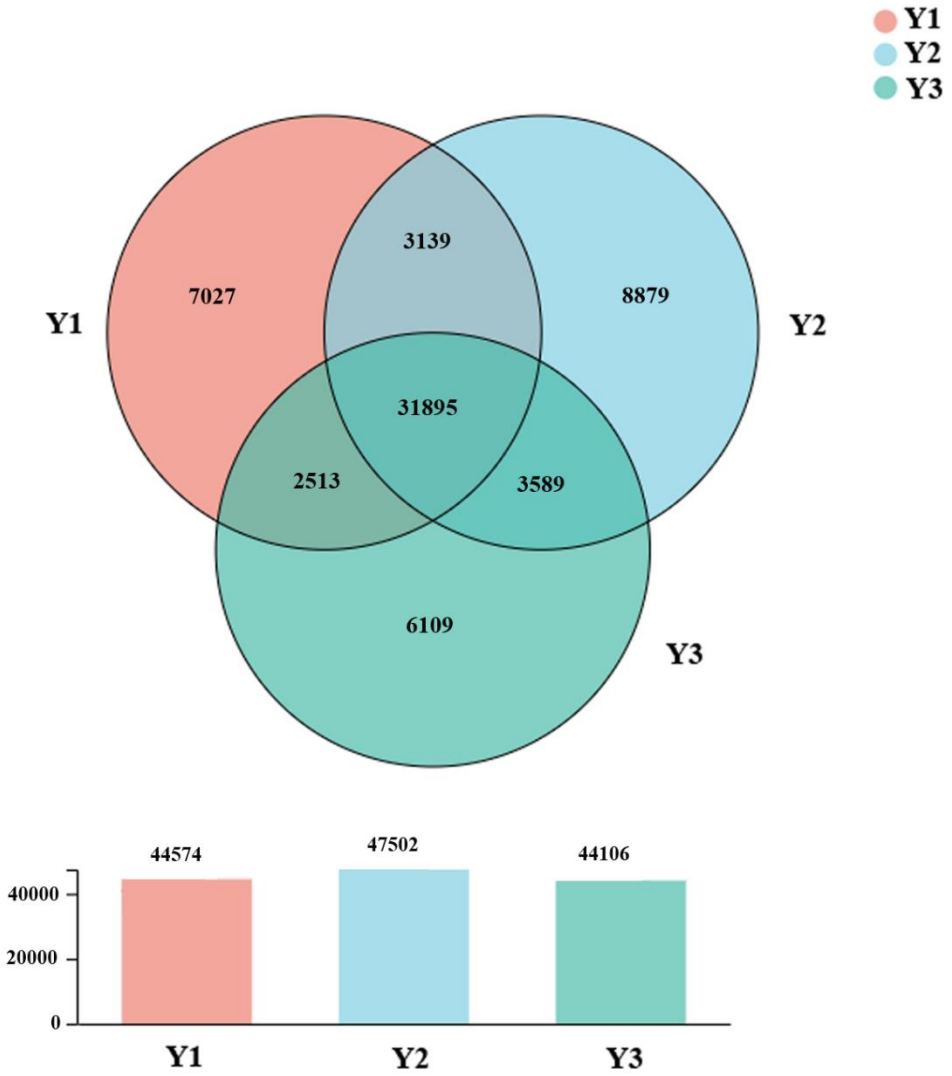


497 Fig. 2 Functional classification and pathway assignment of assembled unigenes by
498 KEGG.



499
500

501 Fig. 3 Venn diagram of unigenes from 1-, 2- and 3-year old rhizome.

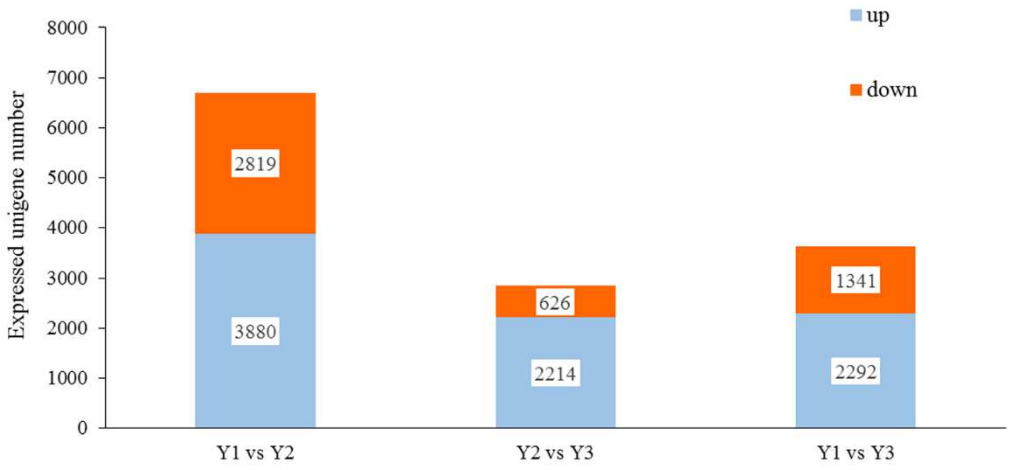


502

503

504

505 Fig. 4 The number of up-down regulated DEGs of Y1 vs Y2, Y1vs Y3 and Y2 vs Y3.



506

507

Fig. 5 Functional classification and pathway assignment of DEGs by KEGG in Y1 vs Y2, Y1 vs Y3 and Y2 vs Y3. (A) Y1 vs Y2; (B) Y1 vs Y3; (C) Y2 vs Y3.

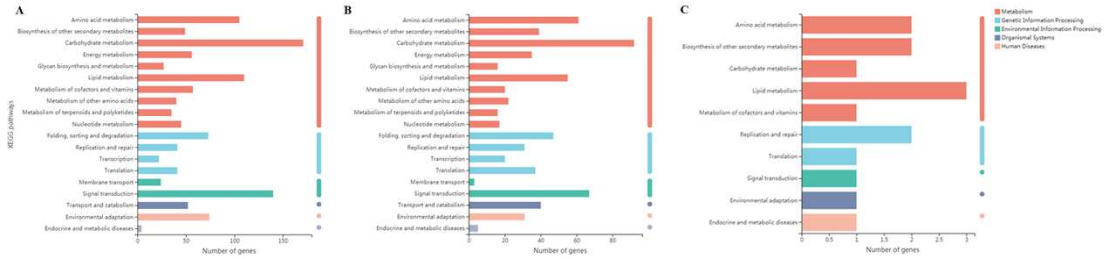


Fig. 6 A venn diagram of DEG statistics from Y1 vs Y2, Y1 vs Y3, Y2 vs Y3 and sesquiterpene genes.

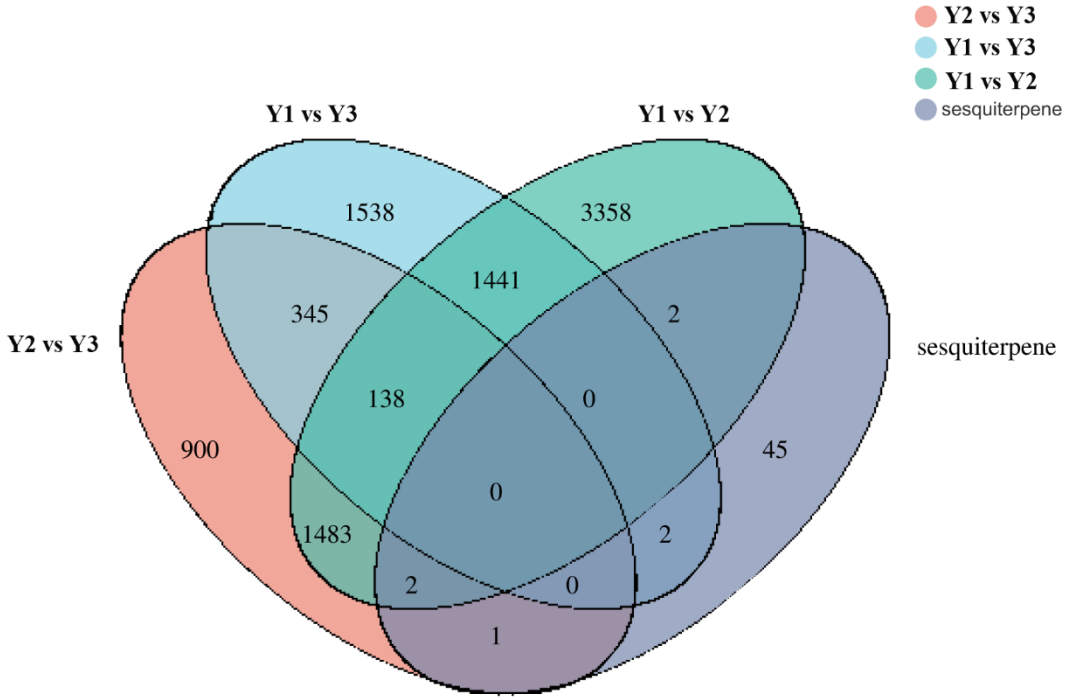


Fig. 7 Heatmap of expressions for DEGs related to sesquiterpene biosynthetic pathway.

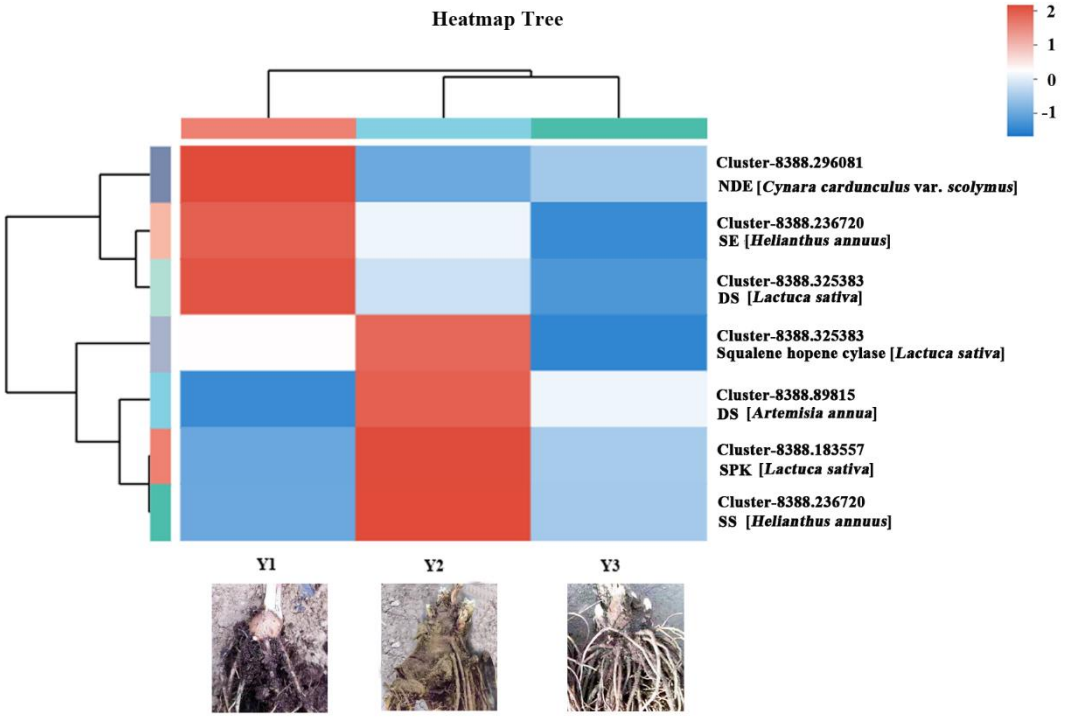


Fig. 8 QRT-PCR validation of transcriptome sequencing analysis. Heat map showed the mean value of transcript levels detected in three biological replicates. Relative transcript levels as detected by RNA-Seq (top) or by qRT-PCR (bottom) were shown by color scales. R, correlation coefficient value between RNA-seq data and qRT-PCR data.

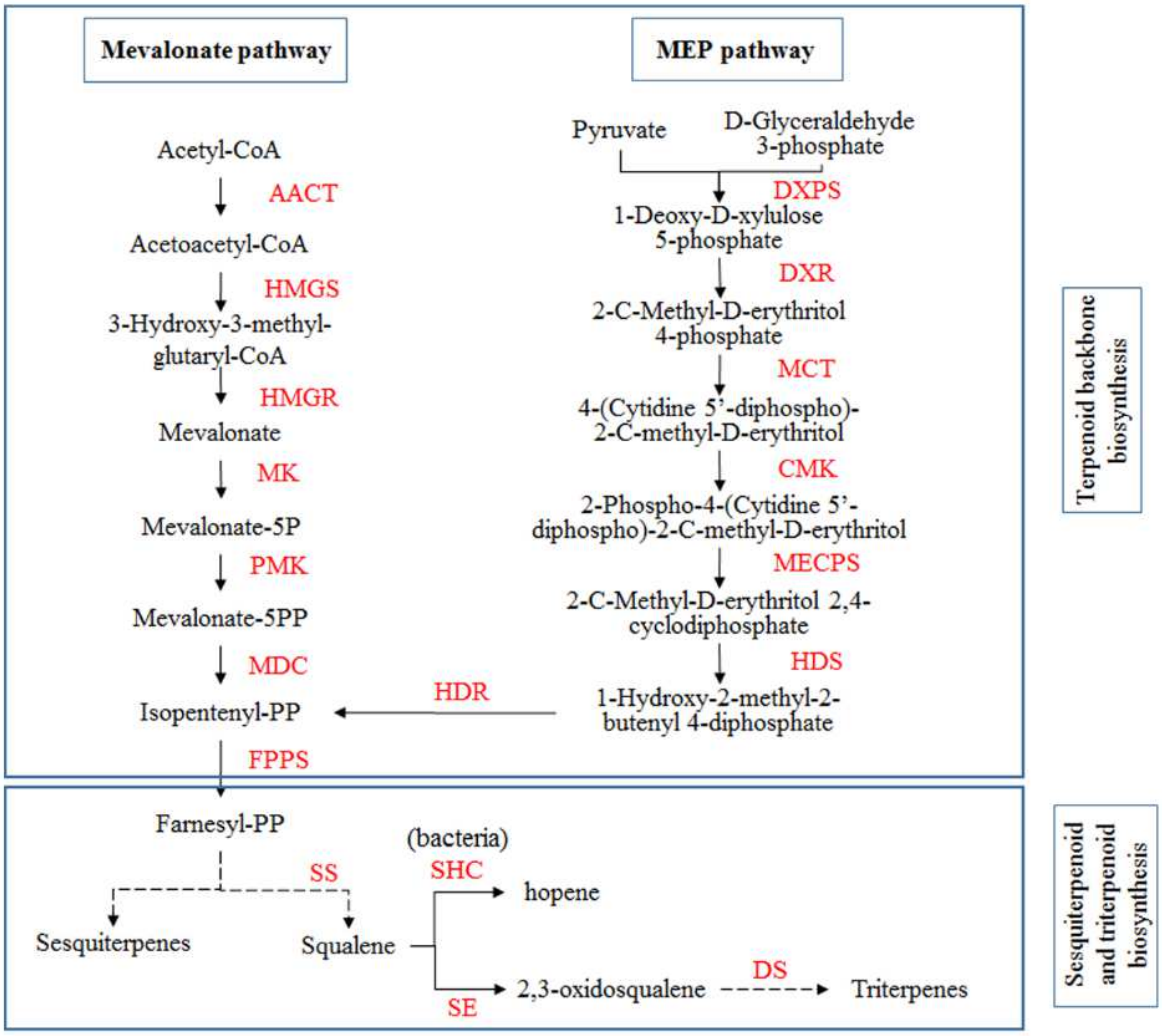


Fig. 9 Flow diagram of Putative sesquiterpenoid and triterpenoid biosynthetic pathway. The red letters represent key enzymes for the action of sesquiterpenoid and triterpenoid biosynthetic pathway. Solid line represented directly catalytic reaction, and dotted line for indirectly catalytic reaction.

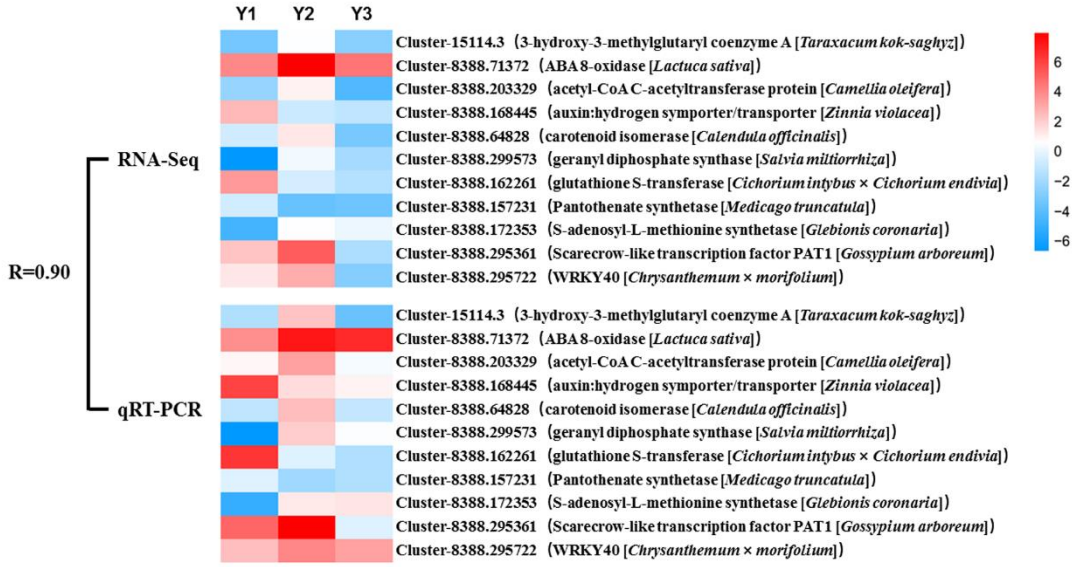
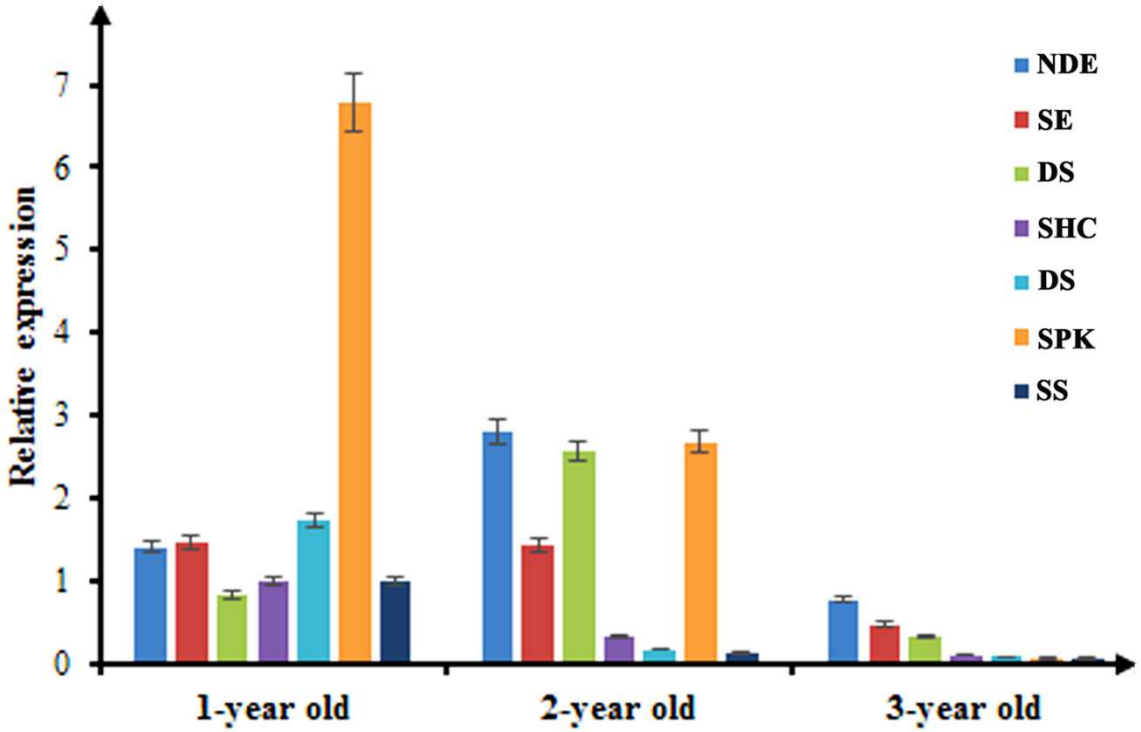


Fig. 10 QRT-PCR analysis of sven DEGs involved in sesquiterpenoid and triterpenoid biosynthesis pathway.



542 Table S1 Primers of qRT-PCR for validation of the reliability of RNA-seq analysis

clusters	primers (5'→3′)	
cluster-15114.3	F: TGTATCACCTCCTCCTCGTCTTCG	R: AATTAGCCAGTCATGCCGTCATCC
cluster-8388.71372	F: CAGTGACAGCTTGTAGGACAGTGG	R: ACAACATCATCGGCGTCATCTTCG
cluster-8388.203329	F: TGGATACTCTGTGCTGCCAACATG	R:CCAGGCAAGCGGCATTAGGTG
cluster-8388.168445	F: GACGCTGGTCTTGGAATGGCTATG	R: GTACGATTGCGACATGGAGGAGTG
cluster-8388.64828	F: GTCGGTGGTGGGTATGATGC	R: CCTGAGCTCCCACCAGGAAT
cluster-8388.299573	F: TGATTCATGTTGCCAGCCTCCTTC	R:ATGCTCCACCGCTGTTGCTATTAG
cluster-8388.162261	F: CAGAGGCTCCTTGTCAACGAAGTC	R: GAGGCATGGCAGAACAAGTCTCC
cluster-8388.157231	F: CCTCGGTTCTTGGTCTGAAGTGTG	R: TGGAGGCTCTCGTTGGAAGTCTG
cluster-8388.172353	F: TGTAGGCGGCGGCTGGTTAC	R: ACACCGTGTTAGATCATGCACCTG
cluster-8388.295361	F: GGAAGCCTACTCAGCACATGGATC	R: CGTGAGAGGTGCAAGTATGGTCAG
cluster-8388.295722	F: GAGCAAGTGGTAGTGGTAGGAAGC	R: CAACCAACCAACCAACCAACGAC

543

544

545 Table S2 Primers of qRT-PCR for validation of the seven DEGs involved in

546 sesquiterpenoid and triterpenoid biosynthesis pathway

Genes	Primers (5'→3′)	
NDE	F: ACAGTGTCAAGAGCGCGAAGAAC	R: CTGAAGATGGCGGCGAAGAAGG
SE	F: CCGCCTGGTTGTAGCAGTTCAC	R: ACCGTAGGATGGACGCAGAGTAC
DS	F: TGTCAACCTTGGACTTGAGCATCG	R: CGCAGATCATGGCAATTGGAACAC
SHC	F: GTGTTGGTGACTGGTGCTTCTGG	R: CAAGAGCAACGCCTTCGGATAGAG
DS	F: AGCGCCAAGAAGTCGAAGATGC	R: CTTCACCGCCGTCGTCACTG
SPK	F: TGCGTAAGCATCAGCAACGGTAG	R: TGTGATGCAGGAGAAGCTTGGATG
SS	F: CCGGACACAGAACTAAGGAGATCG	R: ACAAGCTACTAGCGCATCTACTGC

547

Figures

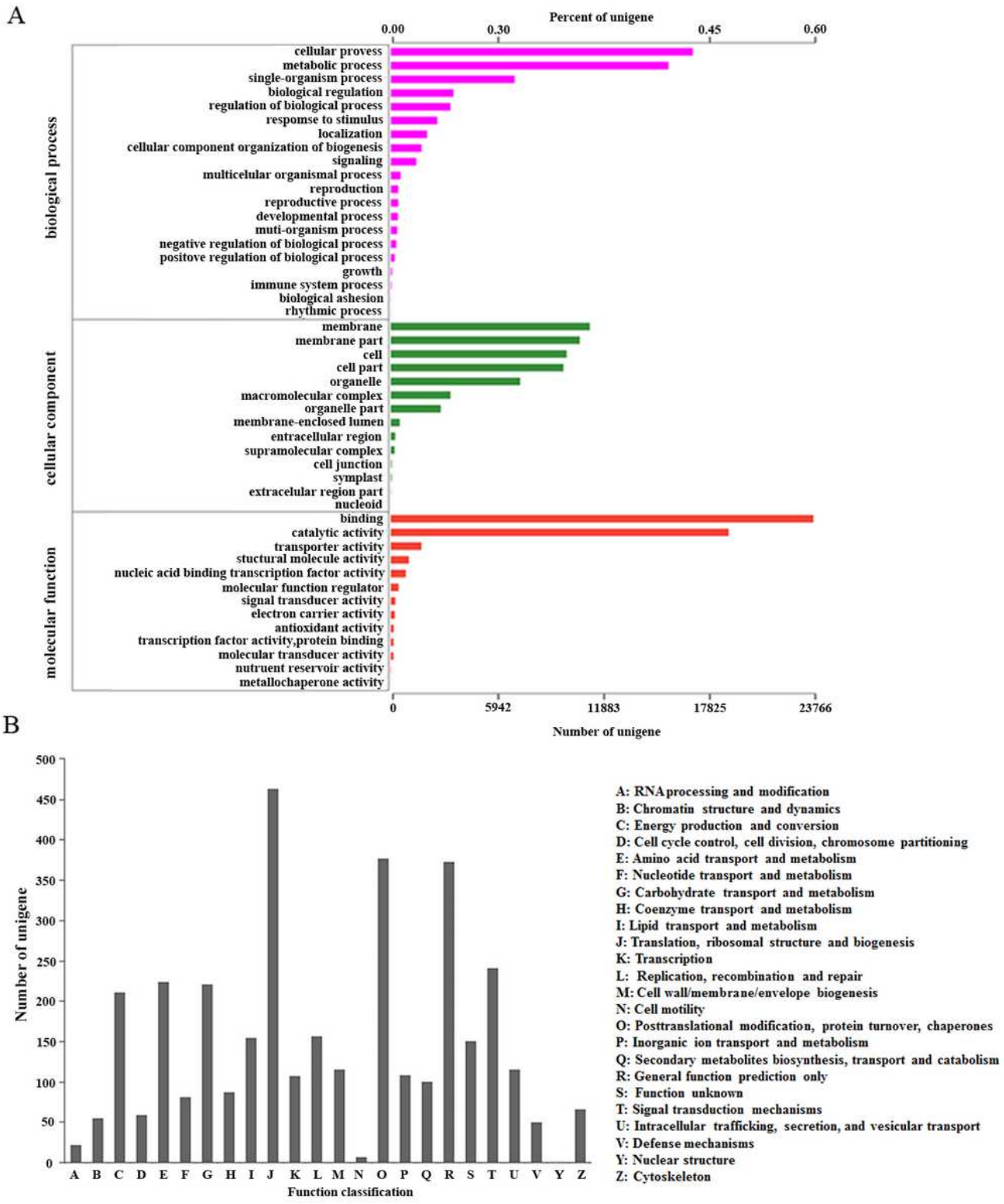


Figure 1

GO and COG classification of assembled unigenes. (A) CO classification; (B) COG classification.

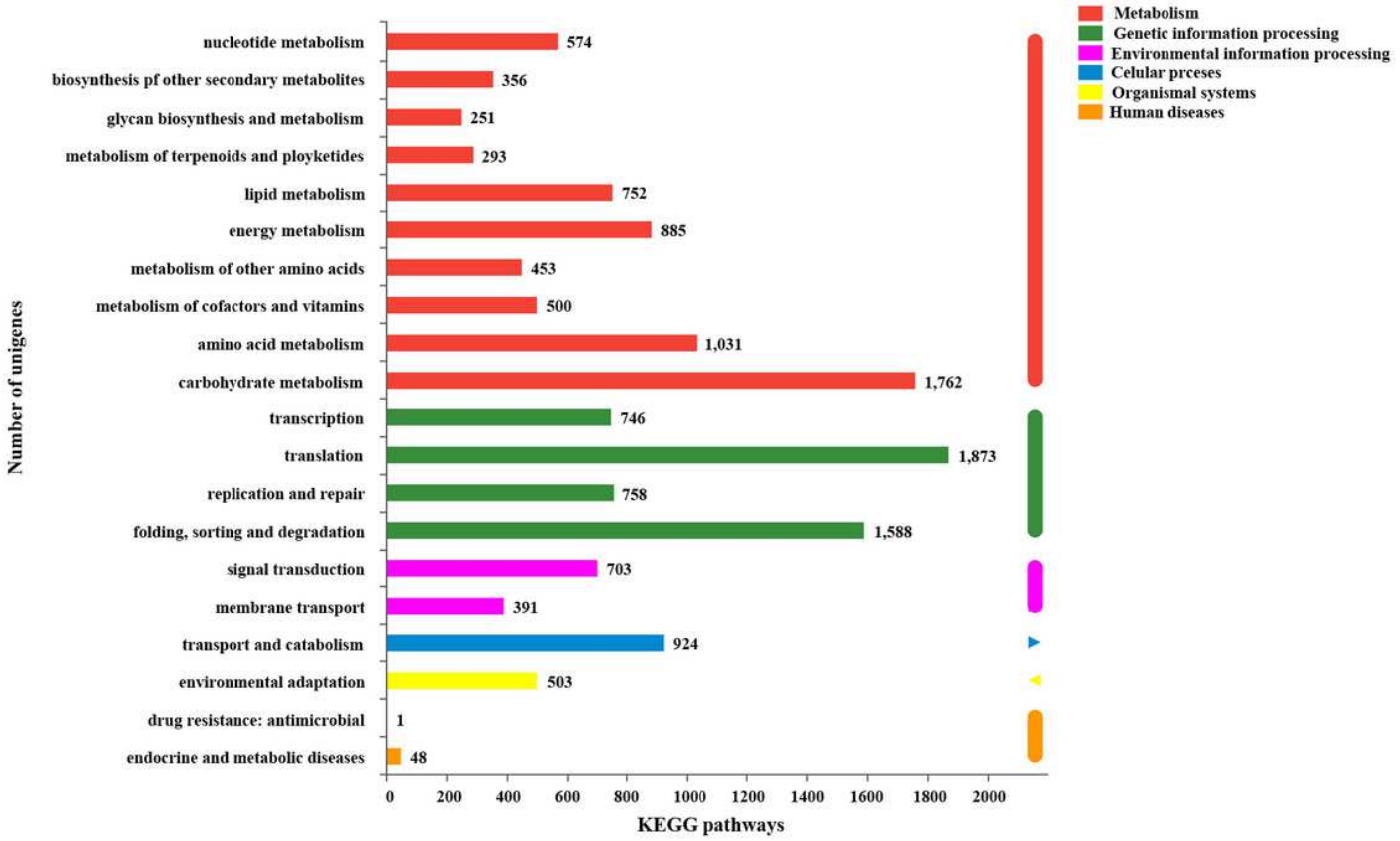


Figure 2

Functional classification and pathway assignment of assembled unigenes by KEGG.

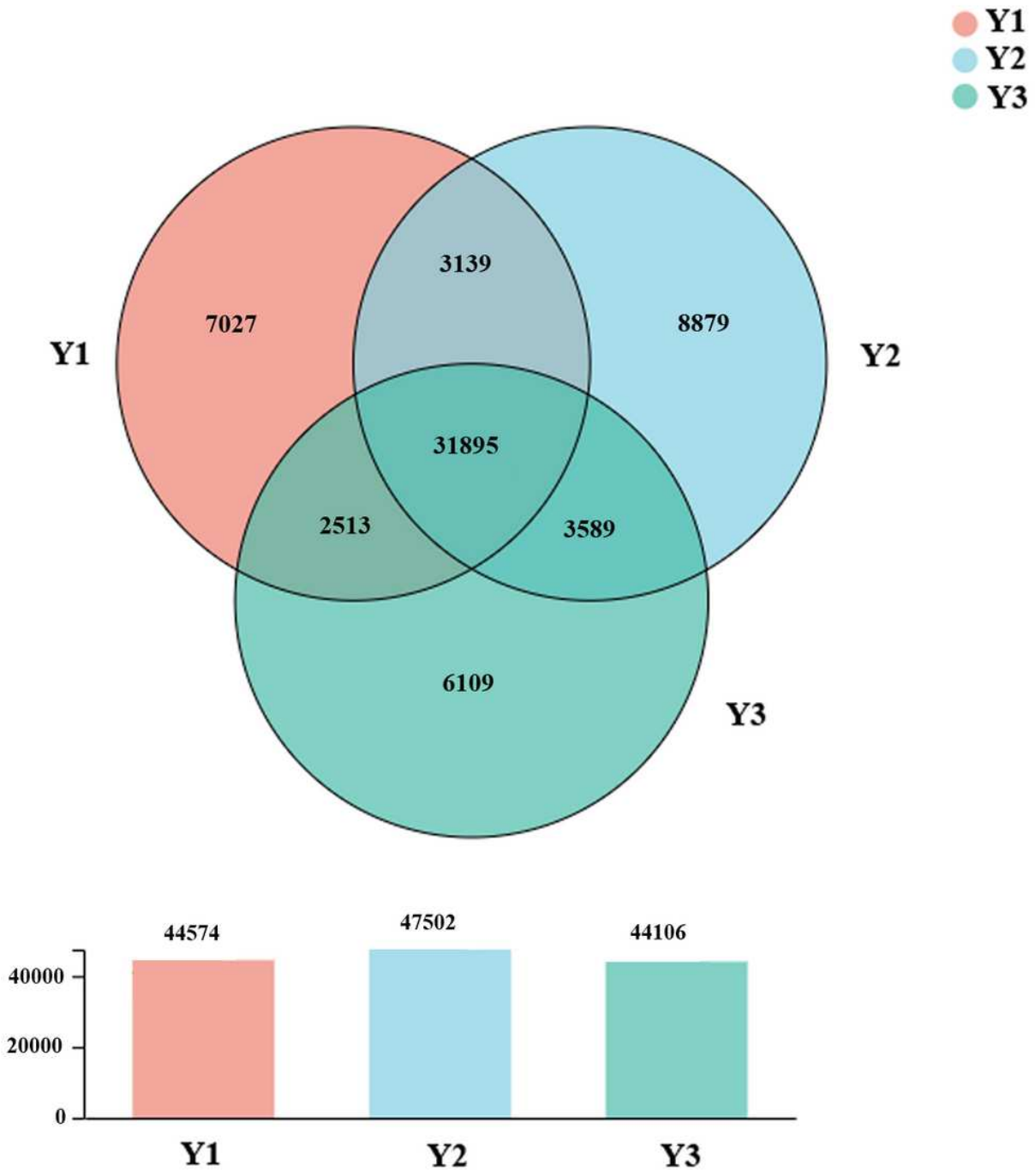


Figure 3

Venn diagram of unigenes from 1-, 2- and 3-year old rhizome.

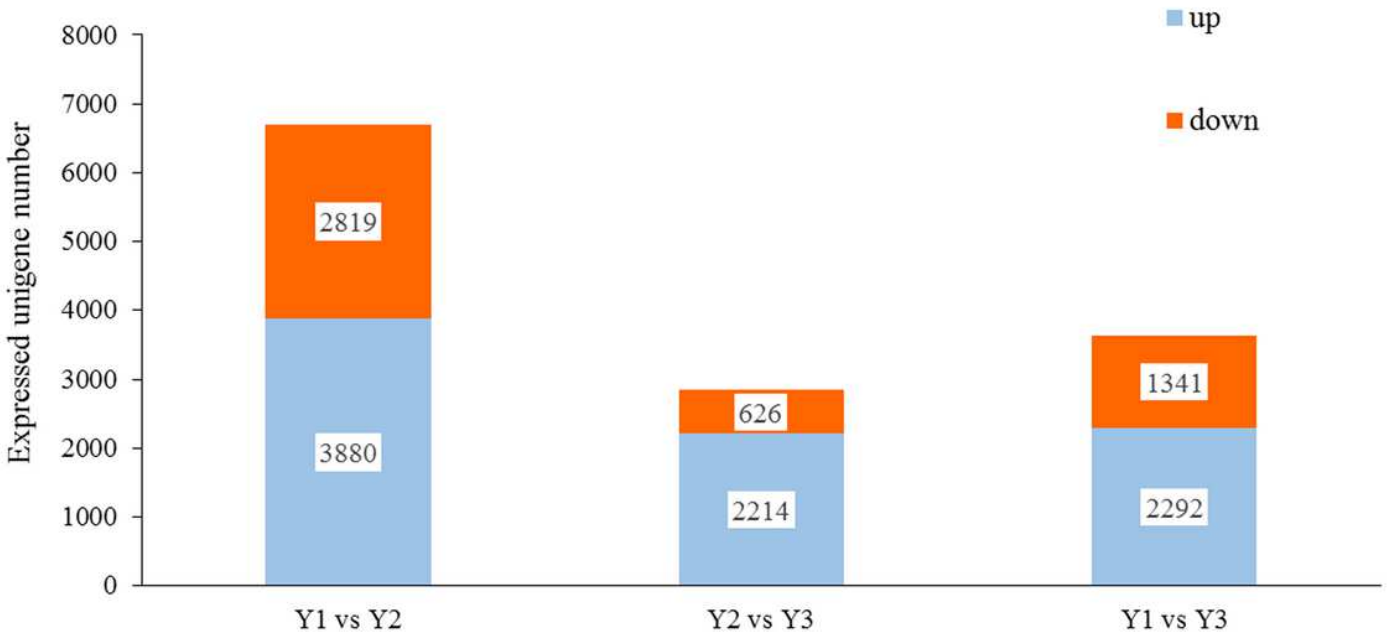


Figure 4

The number of up-down regulated DEGs of Y1 vs Y2, Y1vs Y3 and Y2 vs Y3.

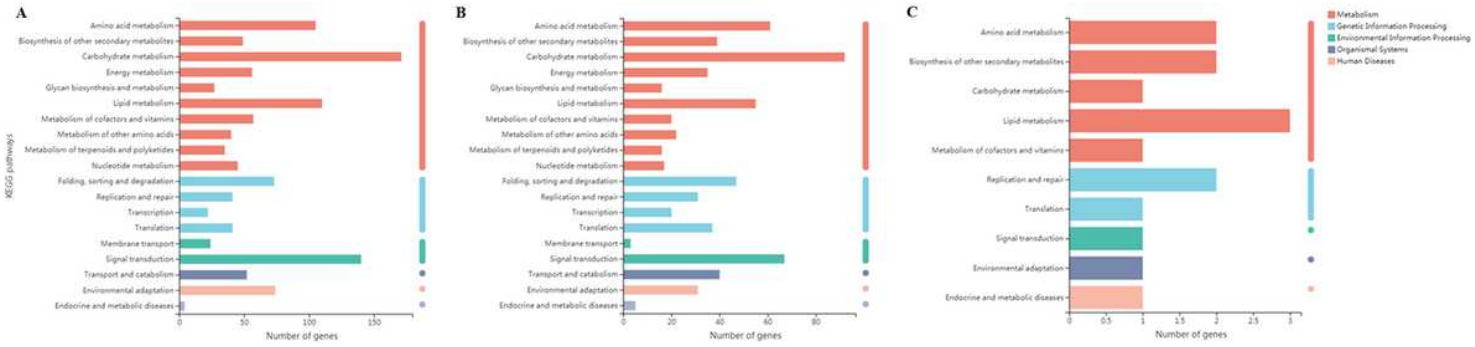


Figure 5

Functional classification and pathway assignment of DEGs by KEGG in Y1 vs Y2, Y1 vs Y3 and Y2 vs Y3. (A) Y1 vs Y2; (B) Y1 vs Y3; (C) Y2 vs Y3.

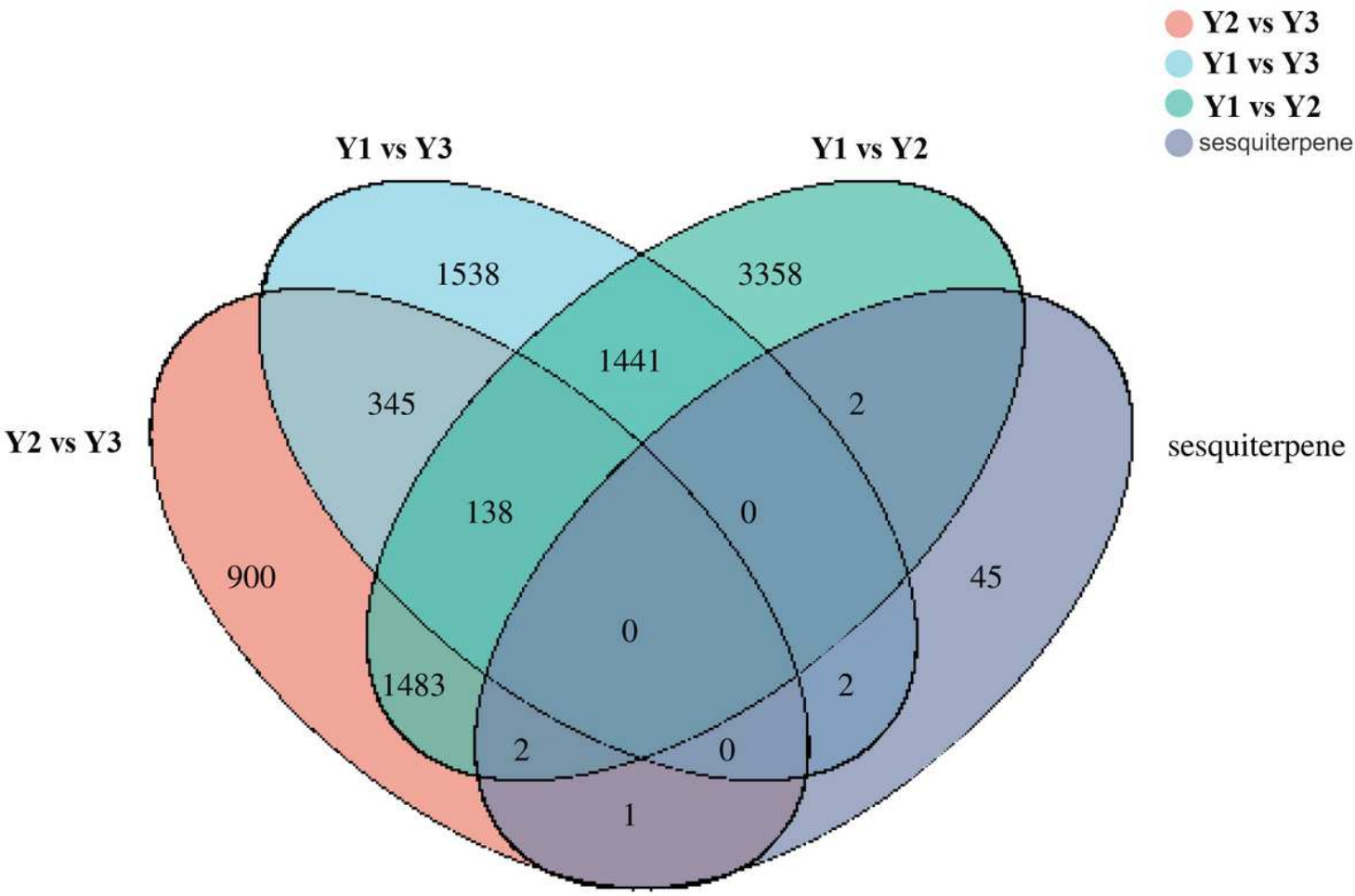


Figure 6

A venn diagram of DEG statistics from Y1 vs Y2, Y1 vs Y3, Y2 vs Y3 and sesquiterpene genes.

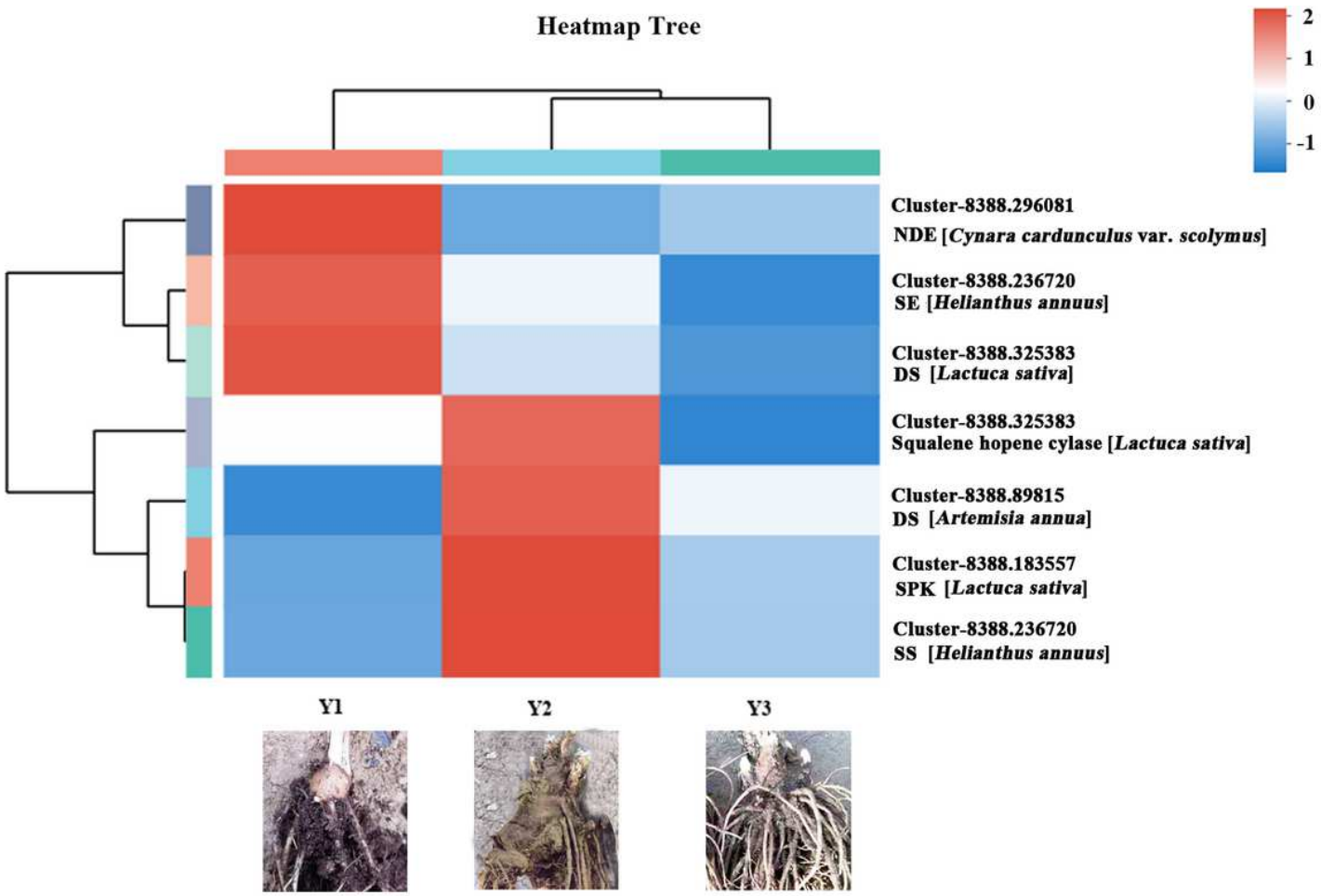


Figure 7

Heatmap of expressions for DEGs related to sesquiterpene biosynthetic pathway

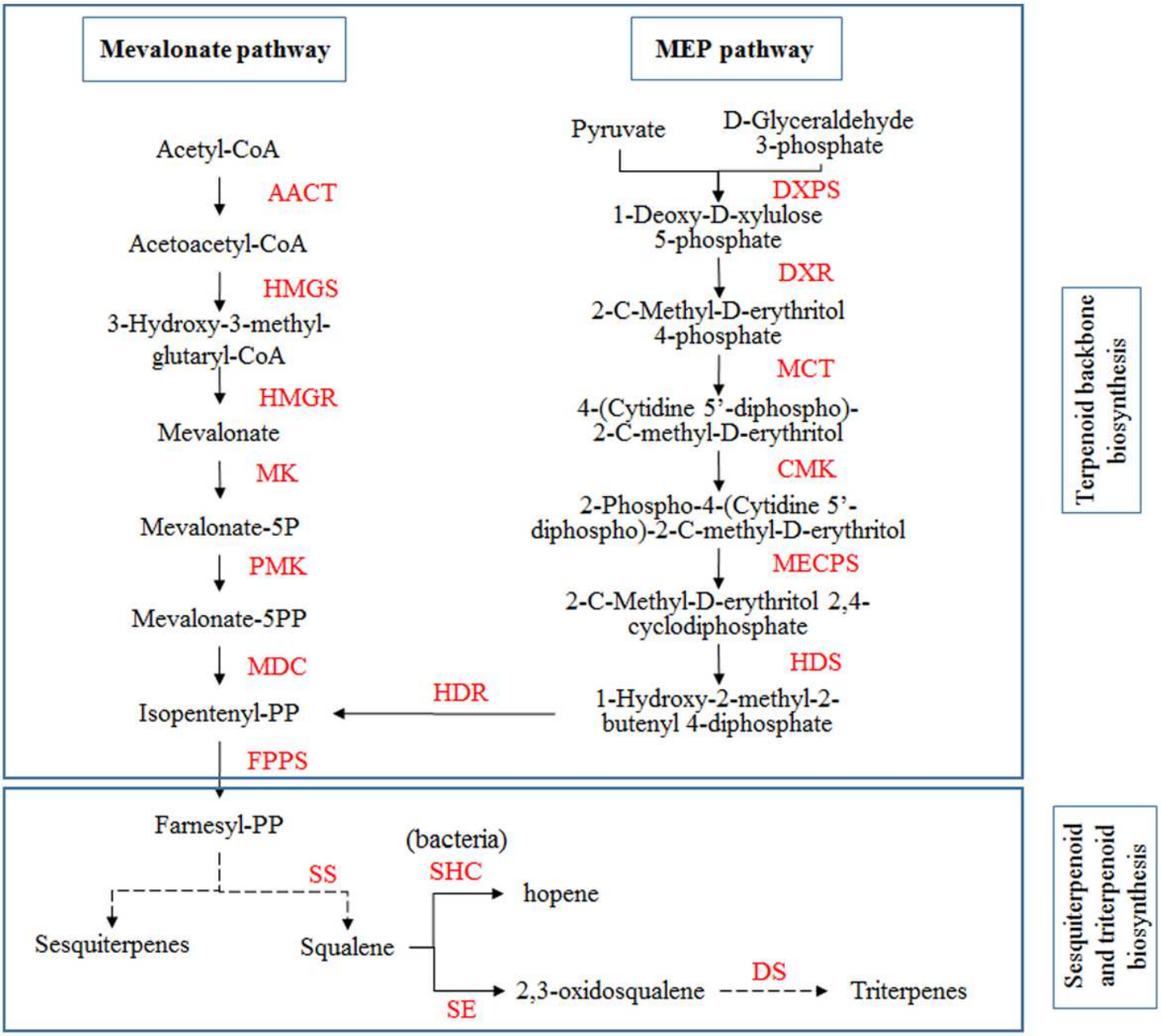


Figure 8

QRT-PCR validation of transcriptome sequencing analysis. Heat map showed the mean value of transcript levels detected in three biological replicates. Relative transcript levels as detected by RNA-Seq (top) or by qRT-PCR (bottom) were shown by color scales. R, correlation coefficient value between RNA-seq data and qRT-PCR data.

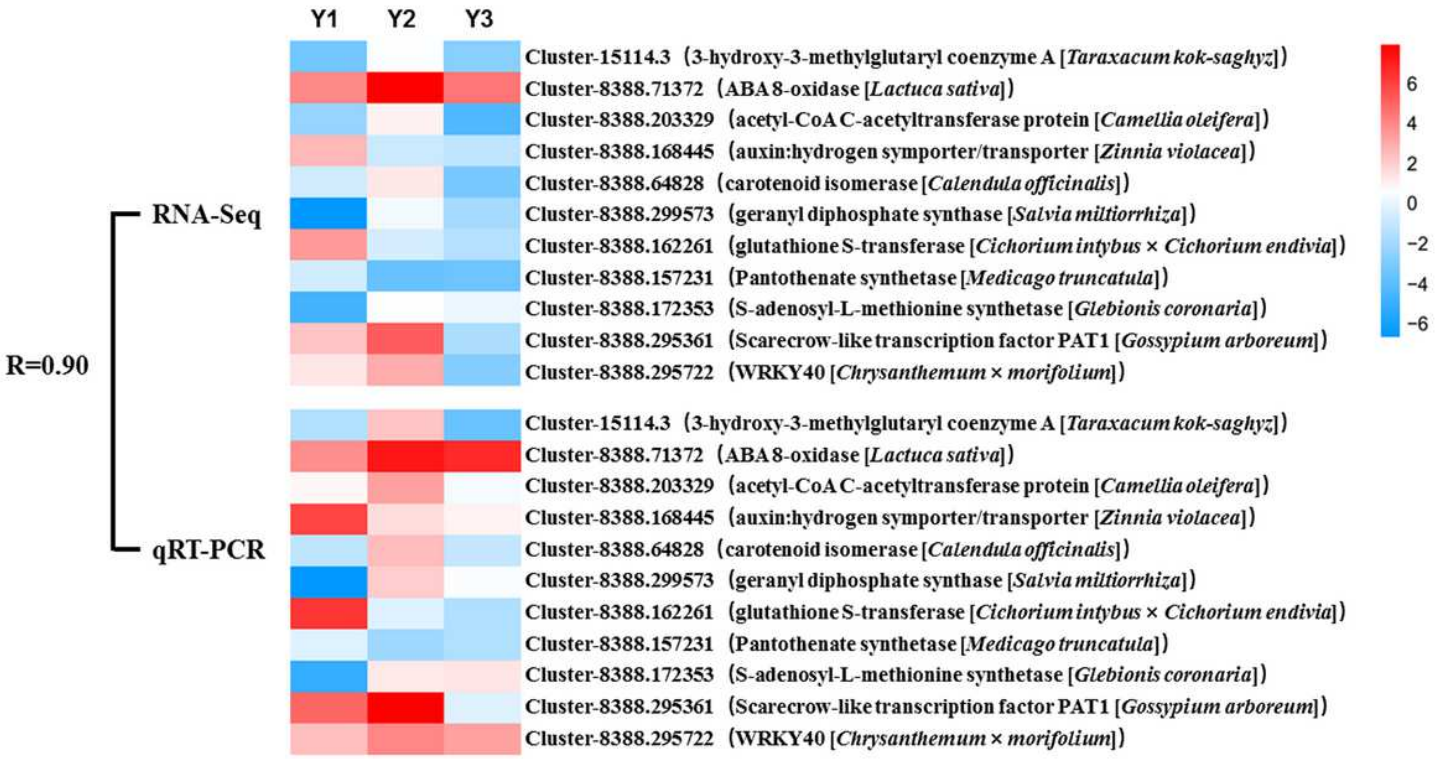


Figure 9

Flow diagram of Putative sesquiterpenoid and triterpenoid biosynthetic pathway. The red letters represent key enzymes for the action of sesquiterpenoid and triterpenoid biosynthetic pathway. Solid line represented directly catalytic reaction, and dotted line for indirectly catalytic reaction.

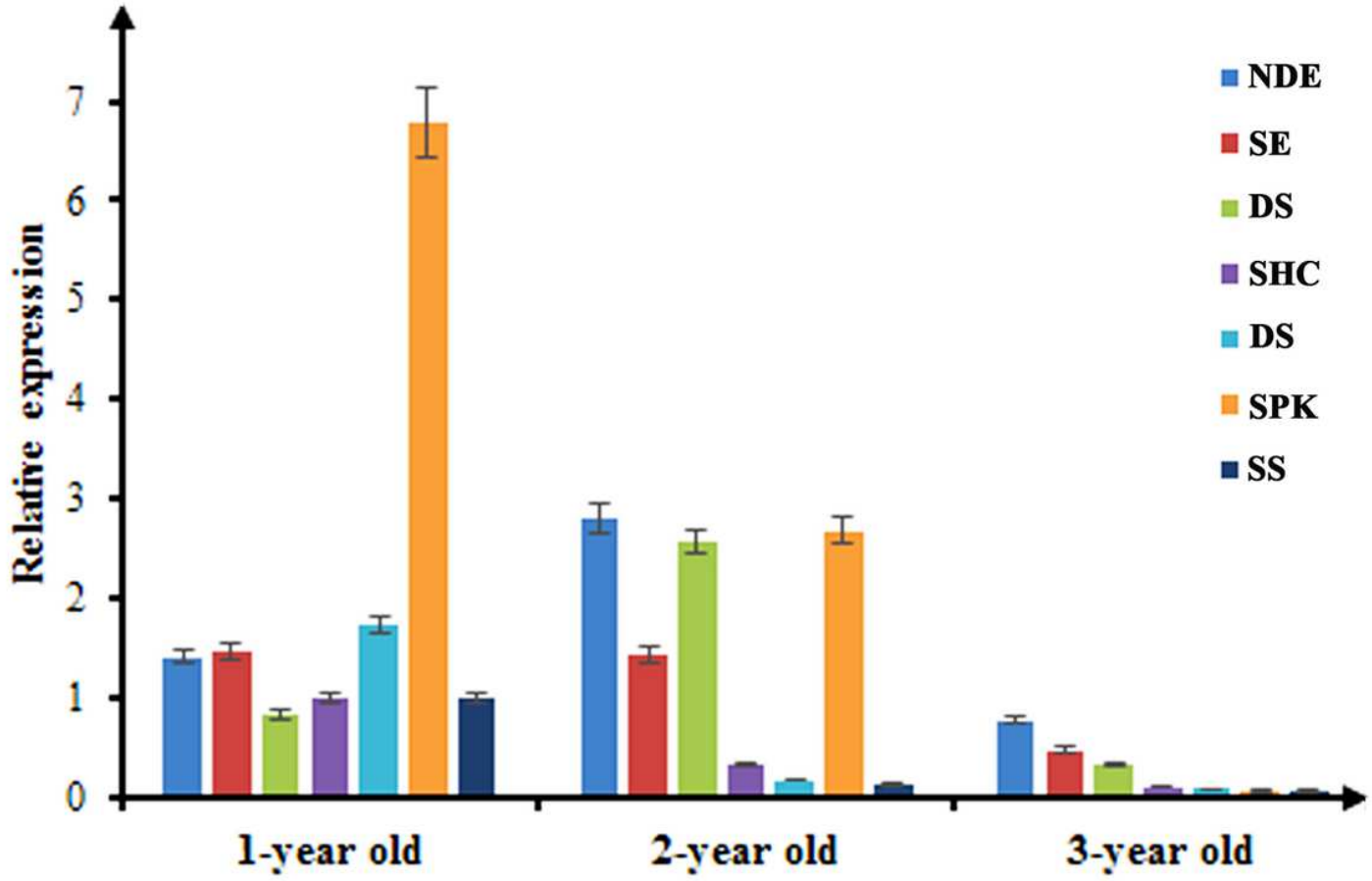


Figure 10

QRT-PCR analysis of sven DEGs involved in sesquiterpenoid and triterpenoid biosynthesis pathway.

Supplementary Files

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

- TableS.docx