Comparative transcriptomics characterized the distinct biosynthetic abilities of terpenoid and paeoniflorin biosynthesis in herbaceous peony strains

Baowei Lu [1], **Qian Gao** [2], **Xuan Wang** [3], **Yongjian Yang** [1], **Pengming Liu** [1], **Baoliang Yang** [1], **Tong Chen** [4], **Xin-Chang Li** [2], **Qinghua Chen** [1], **Fengxia An** [Corresp., 1], **Jun Liu** [Corresp. 2]

[1] Bozhou University, Bozhou, Anhui, China

[2] National Key Facility for Crop Resources and Genetic Improvement, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China

[3] Department of Biological Sciences, Xian Jiaotong-Liverpool University, Suzhou, China

[4] National Resource Center for Chinese Materia Medica, Academy of Chinese Medical Sciences, Beijing, China

Corresponding Authors: Fengxia An, Jun Liu
Email address: An_fengxia@163.com, liujun@caas.cn

The herbaceous peony (*Paeonia lactiflora* Pall.) is a perennial flowering plant of the Paeoniaceae species that is widely cultivated for medical and ornamental uses. The monoterpene glucoside paeoniflorin and its derivatives are the active compounds of the *P.lactiflora* roots. However, the gene regulation pathways associated with monoterpene and paeoniflorin biosynthesis in *P.lactiflora* are still unclear. Here, we selected three genotypes of *P.lactiflora* with distinct morphologic features and chemical compositions that were a result of long-term reproductive isolation. We performed an RNA-sequencing experiment to profile the transcriptome changes of the shoots and roots. Using *de novo* assembly analysis, we identified 36,264 unigenes, including 521 genes responsible for encoding transcription factors. We also identified 28,925 unigenes that were preferentially expressed in different organs and/or genotypes. Pathway enrichment analysis showed that the *P.lactiflora* unigenes were significantly overrepresented in several secondary metabolite biosynthesis pathways. We identified and profiled 33 genes responsible for encoding the enzymescontrolling the major catalytic reactions in the terpenoid backbone and in monoterpenoid biosynthesis. Our study identified the candidate genes in the terpenoid biosynthesis pathways, providing useful information for metabolic engineering of *P.lactiflora* intended for pharmaceutical uses and facilitating the development of strategies to improve marker-assist *P.lactiflora* in the future.

Comparative Transcriptomics Characterized the Distinct Biosynthetic Abilities of Terpenoid and Paeoniflorin Biosynthesis in Herbaceous Peony Strains

Baowei Lu[1,3], Qian Gao[2], Xuan Wang[2,3], Yongjian Yang[1], Pengming Liu[1], Baoliang Yang[1], Tong Chen[4], Xin-Chang Li[2], Qinghua Chen[1], Fengxia An[1]*and Jun Liu[2]*

[1]Bozhou University, Anhui Province, 236800, China
[2]National Key Facility for Crop Resources and Genetic Improvement, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[3]Department of Biological Sciences, Xian Jiaotong-Liverpool University, Suzhou 215123, China
[4]National Resource Center for Chinese Materia Medica, Academy of Chinese Medical Sciences, Beijing 100700, China

Corresponding Author:
FengxiaAn (An_fengxia@163.com) or Jun Liu (liujun@caas.cn)

Abstract

The herbaceous peony (*Paeonia lactiflora* Pall.) is a perennial flowering plant of the Paeoniaceae species that is widely cultivated for medical and ornamental uses. The monoterpene glucoside paeoniflorin and its derivatives are the active compounds of the *P. lactiflora* roots. However, the gene regulation pathways associated with monoterpene and paeoniflorin biosynthesis in *P. lactiflora* are still unclear. Here, we selected three genotypes of *P. lactiflora* with distinct morphologic features and chemical compositions that were a result of long-term reproductive isolation. We performed an RNA-sequencing experiment to profile the transcriptome changes of the shoots and roots. Using *de novo* assembly analysis, we identified 36,264 unigenes, including 521 genes responsible for encoding transcription factors. We also identified 28,925 unigenes that were preferentially expressed in different organs and/or genotypes. Pathway enrichment analysis showed that the *P. lactiflora* unigenes were significantly overrepresented in several secondary metabolite biosynthesis pathways. We identified and profiled 33 genes responsible for encoding the enzymes controlling the major catalytic reactions in the terpenoid backbone and in monoterpenoid biosynthesis. Our study identified the candidate genes in the terpenoid biosynthesis pathways, providing useful information for metabolic engineering of *P. lactiflora* intended for pharmaceutical uses and facilitating the development of strategies to improve marker-assist *P. lactiflora* in the future.

Introduction

The herbaceous peony (*Paeonialactiflora Pall.*) is a flowering plant in the family Paeoniaceae, which is native to Central and eastern Asia (Zhao et al. 2018; Zhao et al. 2017). Its dried root is harvested without the bark in the autumn from plants that are between 3-5 years of age; this harvested material is named Radix Paeoniae Alba or Baishao and is a well-known Chinese herb, used for over 2000 years (He & Dai 2011; Zha et al. 2012). A water/ethanol extract of Radix Paeoniae Alba, now known as Total Glucosides of Peony (TGP), was originally used in the treatment of typhoid (Li et al. 2011). Subsequently, TGP has been widely prescribed for fever, rheumatoid arthritis, hepatitis, muscle cramping and spasms, systemic lupus erythematosus, and dysmenorrhea (Fan et al. 2012; He & Dai 2011; Ji et al. 2013; Mao et al. 2012; Nam et al. 2013).

Paeoniflorin (C23H28O11, molecular weight = 480.45) is the major medicinal component in *P. lactiflora* roots. In vitro and in vivo studies in animal models have confirmed that TGP, paeoniflorin, benzoylpaeoniflorin, galloylpaeoniflorin and their derivatives, are medicinally active compounds with multiple pharmacological effects (Fan et al. 2012; He & Dai 2011; Zhou & Wink 2018). TGP can inhibit acute and subacute inflammation, an effect which is potentially mediated by the suppression of prostaglandin E2, leukotriene B4, and nitric oxide, as well as the intracellular calcium ion concentration (He & Dai 2011; Xu et al. 2016). TGP has been known to protect cells against Ca^{2+} overload and oxidative stress (Zhang et al. 2017). Moreover, the components of TGP, as important immunomodulatory effectors, can regulate the proliferation and apoptosis of lymphocytes and balance the production of proinflammatory cytokines in a dose-dependent manner (He & Dai 2011; Hu et al. 2018). In addition, paeoniflorin and its derivatives were shown to inhibit tumor growth and macrophage-mediated lung metastases (Ou et al. 2011; Wu et al. 2015).

Paeoniflorin is a monoterpene glucoside that is biosynthesized from geranyl-pyrophosphate (GPP). GPP is produced via a conversion from the universal terpenoid precursor, Isopentenyl pyrophosphate (IPP). In plants and bacteria, IPP is produced from the two terpene biosynthesis pathways, the mevalonate pathway (MVA), and the 1-deoxy-d-xylulose-5-phosphate/ methyl-erythritol-4-phosphate (DXP/MEP) pathway (Kanehisa et al. 2012; Ren et al. 2009; Xie et al. 2011). The MVA pathway reactions take place in the cytosol and are catalyzed by enzymes including hydroxyl methylglutaryl-CoA synthase, acetyl-CoA *C*-

acetyltransferase, HMG-CoA reductase, mevalonate kinase, and phosphomevalonate kinase. The DXP/MEP pathway is catalyzed in the plastids. Pyruvate and glyceraldehyde 3-phosphate are converted by 1-deoxy-D-xylulose-5-phosphate synthase and 1-deoxy-D-xylulose-5-phosphate reductoisomerase to 1-deoxy-D-xylulose 5-phosphate and 2-C-methyl-D-erythritol 4-phosphate, respectively. The products are subsequently catalyzed by 4-diphosphocytidyl-2-C-methyl-D-erythritol (CDP-ME) synthase, CDP-ME kinase, and 2-C-methyl-D-erythritol2,4-cyclodiphosphate synthase to mediate the formation of 2-C-methyl-D-erythritol 2,4-cyclopyrophosphate, which is then converted to (E)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate (HMB-PP) by HMB-PP synthase. HMB-PP is converted to IPP and dimethylallyl pyrophosphate (DMAPP) by HMB-PP reductase. IPP and DMAPP are condensed by geranyl pyrophosphate synthase to produce GPP. In addition to producing a monoterpene, GPP is also a precursor to sesquiterpenes and diterpenes. The conversion from GPP to alpha-terpineol is critical for producing the monoterpene, which is catalyzed by (-)-alpha-terpineolsynthase (EC 4.2.3.111, RLC1). Paeoniflorin can be modified by benzoic acid and gallic acid to produce benzoylpaeoniflorin and galloylpaeoniflorin, respectively. Benzoic acid and gallic acid are catalyzed by 3-deoxy-7-phosphoheptulonate synthase, 3-dehydroquinate synthase, and 3-dehydroquinatedehydratase/shikimate dehydrogenase.

With a long history of domestication and selection, the *P. lactiflora* strains used for medical purposes contain high levels of paeoniflorin and are nearly completely infertile due to embryo abortion in their traditional planting regions, like the Bozhou area. Strains have been reproduced through the vegetative propagation of shoots for hundreds of years. Thus these *P. lactiflora* accessions are reproductively isolated and may serve as suitable resources in the investigation of the genetic and molecular basis of the paeoniflorin biosynthesis pathways. Using sequence homology to search for the known sequences and domains, several studies identified a group of genes involved in paeoniflorin biosynthesis. For example, the previous study identified 24 genes, including 8 with full-length cDNA sequences and revealed transcriptional and phylogenetic associations with paeoniflorin biosynthesis (Yuan et al. 2013). However, the gene alias in the paeoniflorin biosynthesis pathways and their expression patterns have not been profiled in *P. lactiflora* strains on a genome-wide basis.

High throughput sequencing technologies have revolutionized genomic and transcriptomic studies.

Improved algorithms are now available for *de novo* reassembly of the transcriptome of a non-model plant species without a valid reference genome sequence. In this study, we assembled the transcriptome of roots and shoots derived from the 3 strains of *P. lactiflora* using *de novo* assembly analysis. By aligning the assembled genes with public databases, we globally annotated 34,203 unigenes in *P. lactiflora*. These genes may have potential functions. For instance, our analysis identified 521 transcription factor genes. Moreover, we profiled gene expression levels and identified a group of tissue-preferential and/or strain-preferential expressed genes using differential expression analysis. Of these genes, we identified the candidate genes associated with the nearly completed terpenoid backbone biosynthesis pathway. We verified the expression pattern of a selective group of candidate genes using the qRT-PCR assay. Our study provides a valuable dataset for updating our understanding of the gene regulatory network underlying paeoniflorin biosynthesis in *P. lactiflora*.

Materials and Methods

Plant materials for RNA-Seq

The *P. lactiflora* Pu-Bang, Xian-Tiao, and Guan-Shang strains were conserved and cultivated under field conditions at Bozhou University, Bozhou, China. The shoots and roots of 3-year old plants were isolated. To avoid circadian effects, we harvested all the tissues in the afternoon of the same day. The samples were frozen in liquid nitrogen immediately after harvesting. The isolated samples and purified RNA were stored at -80 °C.

RNA extraction，library construction, and Illumina sequencing

The total RNA of individual samples was extracted and purified with the RNeasy® Plant Mini Kit (QIAGEN, Germany). RNA concentration was measured using a Nanodrop 2100 spectrophotometer. RNA Integrity values were checked using an Agilent Bioanalyzer. The samples with a RIN score >8.5 were used for library construction (Liu et al. 2014). The sequencing libraries were generated using a NEB Next Ultra RNA Library Prep Kit for Illumina (New England Biosystems), following the manufacturer's recommendations. Library sequencing was performed on a Hi-Seq X10 system with 150-cycle paired-end sequencing protocol (Illumina).

Analysis of RNA-seq datasets

Transcriptome assembly was performed using Trinity (Haas et al. 2013). Fragments per kilobase of exon

per million fragments mapped of assembled transcripts (FPKM) were calculated and normalized using DESeq2 with global normalization parameters (Anders et al. 2014; Love et al. 2014; Quinn & Chang 2016; Zhang et al. 2014). Differential expression analysis was carried out using DESeq2 (Anders & Huber 2010). Genes with normalized fold-change greater than 2, significance P-value less than 0.05, and Benjamini-Hochberg false discovery rate less than 0.1 were considered to be differentially expressed genes.

Transcriptome annotation and pathway analysis

The sequences of the assembled unigenes were annotated by Trinotate (Haas et al. 2013). Coding regions of unigenes were predicted using Transdecoder (Haas et al. 2013). BLAST v2.7.1 was performed to determine the sequence homology (e-value cutoff of $1e^{-5}$) to UniProt/SwissProt, HMMR v3.1b2, EggNOG v4.5.1, and metabolic pathways were analyzed using the Kyoto Encyclopedia of Genes and Genomes database (Kanehisa et al. 2015).

Quantitative detection of paeoniflorin

We referred to the previous method in order to measure paeoniflorin in samples (Yuan et al. 2013). The dried samples (0.50 g) were weighed and extracted with 50 mL of 50% aqueous methanol with ultrasonication for 30 min. The extracted samples were diluted with 50 mL 50% aqueous methanol and filtered with a 0.45-μm Millipore filter membrane (Millipore, MA, USA) at 25 °C. We used the Agilent 1200 LC Series (Agilent Technologies, Palo Alto, CA, USA) High Performance Liquid Chromatography (HPLC) system to measure the paeoniflorin abundance. The wavelength was set at 230 nm with a flow rate of 1.0 mL/min at a temperature of 25 °C. Standard compounds were purchased from the National Institutes for Food and Drug Control and the linearity of the standard compounds was checked at seven concentration solutions.

Quantitative RT-PCR

A total of 1 to 2 μg RNA samples were treated by DNase I (RNeasy plant mini kit) and were reverse transcribed with oligo (dT) primer and SuperScript III (Invitrogen). cDNA samples were analyzed using quantitative PCR with SYBR Premix Ex Taq (Takara) and a Biorad CFX96 real-time PCR system. The conserved glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) transcript sequences were used as endogenous reference genes to normalize the expression levels among samples (Qi et al. 2018). The qRT-PCR reactions were carried out under 60°C annealing temperature and 40 cycles of amplification. We used three technical replicates to produce the average expression levels of the genes relative to that of the reference gene

using the $2^{-\Delta\Delta CT}$ method (Livak & Schmittgen 2001). The primers are listed in Supplemental Table 1.

Accession numbers

The RNA-seq datasets (accession number PRJCA001310) are available on the Genome Sequence Archive database (http://gsa.big.ac.cn/) in the BIG Data Center (BIG.Members. 2018; Wang et al. 2017).

Results

RNA-Seq and *de novo* assembly of the *P. lactiflora* transcriptome

The *P. lactiflora* Pu-Bang (PB) and Xian-Tiao (XT) accessions are the most widely used herbaceous peony stains for medical uses due to their high levels of paeoniflorin, which is derived mainly from their roots; whereas the Guan-Shang (GS) accession contains less paeoniflorin and is usually cultivated for ornamental uses. The morphological features of the 1- and 3-year old plants of the 3 strains were shown in Figure 1, respectively. We determined the paeoniflorin levels of the isolated root samples using High Performance Liquid Chromatography (HPLC). The results confirmed the accumulated levels of paeoniflorin in PB and XT roots compared with those of GS (Supplemental Fig. 1).

To systematically identify genes and explore the gene expression network underlying paeoniflorin biosynthesis in *P. lactiflora*, we purified the RNA samples derived from the shoots and roots of 3-year old PB, XT, and GS plants, and carried out the RNA-sequencing analysis with three biological replicates using the Illumina paired-end 150 bp protocol. After filtering out the low quality reads, we obtained 775.73 million reads in total (Supplemental table 2). Using Trinity (Haas et al. 2013), we performed *de novo* transcriptome assembly and obtained 36,264 unigenes encoding 72,910 transcripts with 986 nt contig N50 length and 42.7% average GC content.

Functional annotation of expressed genes in the three *P. lactiflora* accessions

Due to embryo abortion and vegetative propagation, the *P. lactiflora* accessions have been undergoing reproductive isolation with a long cultivation history in the Bozhou region and were thought to be genetically distinct (Zhou et al. 2002). However, the genomic evidence supporting this point is still lacking. We measured the expression levels of unigenes by calculating normalized Fragments Per Kilobase of exon per million fragments Mapped (FPKM). The unigenes with an FPKM value higher than 1 in at least one were used to perform hierarchical clustering analysis based on the Pearson correlation efficiency. We analyzed the

193 hierarchical structure of the gene expression levels on a genome-wide basis (Fig. 2A). Most of the biological
194 replicate samples belonged to the same clusters. Principal component analysis results also showed a similar
195 result confirming a high level of reproducibility of the biological replicate samples (Fig. 2B). However, the
196 tissues and strains were distributed in different cluster clades. It was noted that all the clades derived from PB
197 and XT were separated from those of GS, suggesting that the strains for medicinal uses are genetically
198 divergent from the strains used for ornamental purposes, possibly due to the selection and reproductive
199 isolation among the strains.

200 Next, we predicted the protein-coding potential for the unigenes using the Transdecoder and searched for
201 the annotation for the unigenes by aligning the assembled transcripts and predicted peptide sequences to the
202 protein sequences annotated by Swiss-Prot, TrEMBL, Pfam, and KEGG databases using Trinotate, BLASTP,
203 and BLASTX (Boutet et al. 2016; Camon et al. 2003; El-Gebali et al. 2019; Haas et al. 2013; Kanehisa et al.
204 2017). In total, we identified 34,203 unigenes containing significant matches to the annotated genes/proteins in
205 at least one database (Fig. 3B). Of them, 28,083 (82%) were reproducibly detected by at least 2 data resources.
206 The annotation information, predicted protein sequences, and FASTA-formatted sequences of these genes
207 were provided in supplemental materials that could serve as a reference annotation for future studies
208 (Supplemental table 3).

209 Transcription factors (TFs) with DNA-binding domains are the major regulators controlling the activity
210 and specificity of the gene transcription process. We predicted genes encoding TFs in our assembled unigene
211 dataset and identified 521 TF-encoding unigenes belonging to 32 TF families (Fig. 3C). Of these, AP2/ERF,
212 bHLH, FAR1, bZIP, and HB are the most abundant TF genes. The detailed information of the TF genes was
213 provided in supplemental Table 4.

214 **Identification of tissue- and/or strain- preferentially expressed unigenes**

215 Next, we searched for the differentially expressed unigenes using paired-wise comparison between
216 different samples (fold change of expression level > 2 and false discovery rate < 0.05). Our analysis identified
217 10,125 up-regulated unigenes in roots and 11,911 in shoots. We found only 332 (3%) and 886 (7%) unigenes
218 were up-regulated in the roots and shoots of the three *P. lactiflora* strains, respectively; whereas there are 1906
219 to 3737 unigenes specifically up-regulated in the roots and/or shoots of each strain (Fig. 4). This result is

consistent with the fact that the three strains have been undergoing genetic separation and selection during the last several centuries.

Identification of genes in terpenoid and paeoniflorin biosynthesis pathway

To further dissect the regulation pathways of the unigenes, we analyzed the gene list enrichment using the Kyoto Encyclopedia of Genes and Genomes (KEGG) datasets and KOBAS3.0 with hypergeometric testing and Benjamini and Hochberg correction (Kanehisa et al. 2012; Xie et al. 2011). In total, we identified 71 significantly enriched KEGG pathways in *P. lactiflora* (Supplemental Table 5). The list included several KEGG terms of secondary metabolite biosynthesis, such as: Terpenoid backbone biosynthesis (*P*-value < 0.0152), Glycerophospholipid metabolism (*P*-value < 0.0001), Inositol phosphate metabolism (*P*-value < 0.0001), Pyruvate metabolism (*P*-value < 0.0002), Seleno compound metabolism (*P*-value < 0.0298), Ascorbate and aldarate metabolism (*P*-value < 0.0172) and Butanoate metabolism (*P*-value < 0.0352). *P. lactiflora* are generally known to have abundant secondary metabolites (Li et al. 2016a; Liu et al. 2017; Ma et al. 2016). Our transcriptome and pathway enrichment results are consistent with the metabolism profiling studies.

The previous studies have identified 19 EST sequences in the terpenoid backbone biosynthesis pathways in *P. lactiflora*, including 7 with full-length cDNA sequences (Yuan et al. 2013). However, the genes in the terpenoid backbone biosynthesis pathway have not been globally profiled and the enzyme catalyzing the initiation step from GPP to monoterpenoid biosynthesis has not been identified in *P. lactiflora* yet. In our datasets, we identified 32 genes with full-length CDSs encoding the enzymes controlling the major catalytic reactions in MVA and MEP pathways (Fig. 5). Moreover, we identified the unigene (E_H33980_c1_g4) encoding (-)-alpha-terpineol synthase (EC 4.2.3.111, RLC1) that can catalyze the conversion from GPP to alpha-terpineol (Kulkarni et al. 2013b), a monoterpene precursor of paeoniflorin. Our analysis identified the genes encoding the enzymes that almost completely catalyzed the reactions from the glycolysis products to the terpenoid backbone and monoterpenoid biosynthesis in *P. lactiflora*.

We analyzed the expression patterns of the aforementioned 33 unigenes (Fig 6). With the exception of *AACT1*, the genes associated with the reactions of the MVP did not clearly reveal XT- and/or PB-root preferential expression patterns; whereas the genes encoding ISPF in the MEP pathway and ISPH, GGR,

CDPMEK, IPI2, FPS1, and RLC1 associated with I-PP to monoterpene conversion reactions showed XT- and/or PB-root preferential expression patterns. To confirm the expression levels determined by the RNA-seq experiment, we used a quantitative Real-time PCR (qRT-PCR) assay to measure the expression levels of several unigenes (Supplemental Fig. 2). These results indicated that the different biosynthetic abilities of terpenoid biosynthesis among the strains may be largely contributed to by several master genes.

Discussion

The herbaceous peony has a large and complex genome. It was estimated that the genome size of the tree peony, a closely related species of *P. lactiflora*, is about 12.5 Gbp. Thus, assembling the high-quality genome of *P. lactiflora* remains a challenge. Recently, transcriptome studies on a large number of plant species with complex genomes, such as smooth cordgrass (*Spartinaalterniflora*) (Bedre et al. 2016), buckwheat (Lu et al. 2018), and *Caraganakor shinskii* (Li et al. 2016b), have been carried out. , Transcriptome assembly is currently a feasible and cost-efficient technology to globally identify genes in the *P. lactiflora* genome because genome information is unavailable. In our study, we identified 36,264 unigenes. Using the homolog alignment analysis, we identified a large number of these assembled unigenes encoded in known regulatory domains/motifs, suggesting their molecular functions. Of these, we identified 521 transcription factor genes belonging to 32 families and profiled their expression patterns. Our study identified a large number of previously un-annotated genes in the *P. lactiflora* genome, which could supply valuable molecular information for future functional studies.

Plant terpenoids are widely used as traditional herbal remedies and for their aromatic qualities. Terpenoids are highly abundant in several accessions of *P. lactiflora,* suggesting a high potential for terpenoid biosynthesis in them. However, the terpenoid biosynthetic pathways are not yet fully understood in *P. lactiflora*. Previous studies have identified the 19 EST sequences in the terpenoid backbone biosynthesis pathway in *P. lactiflora* (Yuan et al. 2013). In our study, we identify 32 unigenes in the pathway and profiled their expression patterns that covered most of the reactions in the terpenoid backbone biosynthesis pathway. With this data, the other unidentified gene members and the complete terpenoid backbone biosynthesis pathway in P. lactiflora strains can be deciphered in the future, using homologous cloning, phylogenetic analysis, and catalytic kinetics experiments.

Our transcriptome profiling analysis uncovered the tissue- and/or strain-preferential expression patterns in the three *P. lactiflora* strains, suggesting the genes in the terpenoid biosynthesis pathway were diversified during the selection process. These results could supply valuable molecular information for future functional studies in *P. lactiflora*. Geranyldiphosphate synthases and farnesyldiphosphate synthase catalyze the important branch-point reactions from GPP to monoterpenes and sesquiterpenes. The catalytic activities of the two enzymes are sensitive to temperature and metal ion concentration (Kulkarni et al. 2013a). Under high temperature and Mg^{2+} rich condition, the monoterpene concentration was significant higher compared with those of sesquiterpenes. In our experiment, we identified several unigenes encoding Geranyldiphosphate synthases and farnesyldiphosphate synthase in *P. lactiflora*. Their catalytic activities could be investigated under different temperature and metal ion conditions in follow-up studies.

It has been known that *P. lactiflora* is highly sensitive to the photoperiod and temperature changes. The accessions grown in the Bozhou region have been undergoing reproductive isolation for a long period. Thus, the genetic background of each individual might be fixed. In our study, we identified a large number of genes in *P. lactiflora* and found gene expression patterns in the terpenoid pathway are highly diversified among different accessions. With the sequences of these genes, the phylogenetic structures and population genetics backgrounds of germplasm resources could be further investigated to elucidate the domestication and selection process of *P. lactiflora*.

Conclusion

Using the RNA-sequencing protocol, we assembled the exon structures of 36,264 genes in the shoots and roots of three 3-year-old *P. lactiflorais* accessions, including 28,925 differentially expressed genes. We systematically annotated their molecular functions by aligning their sequences with those from multiple data resources. 521 transcription factor genes were identified. We identified 32 genes with full-length CDSs encoding the enzymes controlling the major catalytic reactions in MVA and MEP pathways and one gene encoding (-)-alpha-terpineol synthase. These genes contributed nearly complete terpenoid backbone biosynthesis pathways. By profiling gene expression patterns associated with MVP, MEP pathways, and I-PP to monoterpene conversation reactions, we uncovered several unigenes that were highly expressed in the roots of high-paeoniflorin accessions, indicating that the different biosynthetic abilities of terpenoid biosynthesis

among the accessions may be largely contributed to by several master genes.

Author Contributions

F. A., B. L., and J. L. designed the experiment. B. L., Y. Y., P. L., Q. C., Y. Z. and B. Y. sampled the biological materials and extracted the RNA and sequenced the libraries. T. C., F. A., J.L. and Y. W. analyzed the datasets. B.L. J.L. and Y.W. prepared the figures and wrote the manuscript. All authors reviewed and approved the final version of the manuscript.

REFERENCE

Anders S, and Huber W. 2010. Differential expression analysis for sequence count data. *Genome biology* 11:R106. 10.1186/gb-2010-11-10-r106

Anders S, Pyl PT, and Huber W. 2014. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*:btu638.

Bedre R, Mangu VR, Srivastava S, Sanchez LE, and Baisakh N. 2016. Transcriptome analysis of smooth cordgrass (Spartina alterniflora Loisel), a monocot halophyte, reveals candidate genes involved in its adaptation to salinity. *BMC genomics* 17:657-657. 10.1186/s12864-016-3017-3

BIG.Members. 2018. Database Resources of the BIG Data Center in 2018. *Nucleic acids research* 46:D14-D20. 10.1093/nar/gkx897

Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, Poux S, Bougueleret L, and Xenarios I. 2016. UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. In: Edwards D, ed. *Plant Bioinformatics: Methods and Protocols*. New York, NY: Springer New York, 23-54.

Camon E, Barrell D, Brooksbank C, Magrane M, and Apweiler R. 2003. The Gene Ontology Annotation (GOA) Project--Application of GO in SWISS-PROT, TrEMBL and InterPro. *Comparative and functional genomics* 4:71-74. 10.1002/cfg.235

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, and Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Research* 47:D427-D432. 10.1093/nar/gky995

Fan YF, Xie Y, Liu L, Ho HM, Wong YF, Liu ZQ, and Zhou H. 2012. Paeoniflorin reduced acute toxicity of aconitine in rats is associated with the pharmacokinetic alteration of aconitine. *Journal of ethnopharmacology* 141:701-708. 10.1016/j.jep.2011.09.005

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, and Regev A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* 8:1494-1512. 10.1038/nprot.2013.084

He DY, and Dai SM. 2011. Anti-inflammatory and immunomodulatory effects of paeonia lactiflora pall., a traditional chinese herbal medicine. *Frontiers in pharmacology* 2:10. 10.3389/fphar.2011.00010

Hu B, Xu G, Zhang X, Xu L, Zhou H, Ma Z, Shen X, Zhu J, and Shen R. 2018. Paeoniflorin Attenuates Inflammatory Pain by Inhibiting Microglial Activation and Akt-NF-kappaB Signaling in the Central Nervous System. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* 47:842-850. 10.1159/000490076

Ji Y, Wang T, Wei ZF, Lu GX, Jiang SD, Xia YF, and Dai Y. 2013. Paeoniflorin, the main active constituent of Paeonia lactiflora roots, attenuates bleomycin-induced pulmonary fibrosis in mice by suppressing the synthesis of type I collagen. *Journal of ethnopharmacology* 149:825-832. 10.1016/j.jep.2013.08.017

Kanehisa M, Furumichi M, Tanabe M, Sato Y, and Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research* 45:D353-D361. 10.1093/nar/gkw1092

Kanehisa M, Goto S, Sato Y, Furumichi M, and Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research* 40:D109-114. 10.1093/nar/gkr988

Kanehisa M, Sato Y, Kawashima M, Furumichi M, and Tanabe M. 2015. KEGG as a reference resource for gene and protein annotation. *Nucleic acids research* 44:D457-D462.

Kulkarni R, Pandit S, Chidley H, Nagel R, Schmidt A, Gershenzon J, Pujari K, Giri A, and Gupta V. 2013a. Characterization of three novel isoprenyl diphosphate synthases from the terpenoid rich mango fruit. *Plant physiology and biochemistry : PPB* 71:121-131. 10.1016/j.plaphy.2013.07.006

Kulkarni R, Pandit S, Chidley H, Nagel R, Schmidt A, Gershenzon J, Pujari K, Giri A, and Gupta V. 2013b. Characterization of three novel isoprenyl diphosphate synthases from the terpenoid rich mango fruit. *Plant Physiology and Biochemistry* 71:121-131. https://doi.org/10.1016/j.plaphy.2013.07.006

Li B, Bhandari DR, Rompp A, and Spengler B. 2016a. High-resolution MALDI mass spectrometry imaging of gallotannins and monoterpene glucosides in the root of Paeonia lactiflora. *Scientific reports* 6:36074. 10.1038/srep36074

Li J, Chen CX, and Shen YH. 2011. Effects of total glucosides from paeony (Paeonia lactiflora Pall) roots on experimental atherosclerosis in rats. *Journal of ethnopharmacology* 135:469-475. 10.1016/j.jep.2011.03.045

Li S, Fan C, Li Y, Zhang J, Sun J, Chen Y, Tian C, Su X, Lu M, Liang C, and Hu Z. 2016b. Effects of drought and salt-stresses on gene expression in Caragana korshinskii seedlings revealed by RNA-seq. *BMC genomics* 17:200-200. 10.1186/s12864-016-2562-0

Liu D, Wang D, Qin Z, Zhang D, Yin L, Wu L, Colasanti J, Li A, and Mao L. 2014. The SEPALLATA MADS‐box protein SLMBP21 forms protein complexes with JOINTLESS and MACROCALYX as a transcription activator for development of the tomato flower abscission zone. *The Plant Journal* 77:284-296.

Liu P, Xu Y, Yan H, Chen J, Shang E-X, Qian D-W, Jiang S, and Duan J-A. 2017. Characterization of molecular signature of the roots of Paeonia lactiflora during growth. *Chinese Journal of Natural Medicines* 15:785-793. https://doi.org/10.1016/S1875-5364(17)30110-3

Livak KJ, and Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *methods* 25:402-408.

Love MI, Huber W, and Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15:550.

Lu Q-H, Wang Y-Q, Song J-N, and Yang H-B. 2018. Transcriptomic identification of salt-related genes and de novo assembly in common buckwheat (F. esculentum). *Plant Physiology and Biochemistry* 127:299-309. https://doi.org/10.1016/j.plaphy.2018.02.001

Ma X, Chi YH, Niu M, Zhu Y, Zhao YL, Chen Z, Wang JB, Zhang CE, Li JY, Wang LF, Gong M, Wei SZ, Chen C, Zhang L, Wu MQ, and Xiao XH. 2016. Metabolomics Coupled with Multivariate Data and Pathway Analysis on Potential Biomarkers in Cholestasis and Intervention Effect of Paeonia lactiflora Pall. *Frontiers in pharmacology* 7:14. 10.3389/fphar.2016.00014

Mao QQ, Ip SP, Xian YF, Hu Z, and Che CT. 2012. Anti-depressant-like effect of peony: a mini-review. *Pharmaceutical biology* 50:72-77. 10.3109/13880209.2011.602696

Nam KN, Yae CG, Hong JW, Cho DH, Lee JH, and Lee EH. 2013. Paeoniflorin, a monoterpene glycoside, attenuates lipopolysaccharide-induced neuronal injury and brain microglial inflammatory response. *Biotechnology letters* 35:1183-1189. 10.1007/s10529-013-1192-8

Ou TT, Wu CH, Hsu JD, Chyau CC, Lee HJ, and Wang CJ. 2011. Paeonia lactiflora Pall inhibits bladder cancer growth involving phosphorylation of Chk2 in vitro and in vivo. *Journal of ethnopharmacology* 135:162-172. 10.1016/j.jep.2011.03.011

Qi Z, Zhang Z, Wang Z, Yu J, Qin H, Mao X, Jiang H, Xin D, Yin Z, Zhu R, Liu C, Yu W, Hu Z, Wu X, Liu J, and Chen Q. 2018. Meta-analysis and transcriptome profiling reveal hub genes for soybean seed storage composition during seed development. *Plant, cell & environment*. 10.1111/pce.13175

Quinn JJ, and Chang HY. 2016. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet* 17:47-62. 10.1038/nrg.2015.10

Ren ML, Zhang X, Ding R, Dai Y, Tu FJ, Cheng YY, and Yao XS. 2009. Two new monoterpene glucosides from Paeonia lactiflora Pall. *Journal of Asian natural products research* 11:670-674. 10.1080/10286020902980087

Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, Tang B, Dong L, Ding N, Zhang Q, Bai Z, Dong X, Chen H, Sun M, Zhai S, Sun Y, Yu L, Lan L, Xiao J, Fang X, Lei H, Zhang Z, and Zhao W. 2017. GSA: Genome Sequence Archive<sup/>. *Genomics, proteomics & bioinformatics* 15:14-18. 10.1016/j.gpb.2017.01.001

Wu Q, Chen GL, Li YJ, Chen Y, and Lin FZ. 2015. Paeoniflorin inhibits macrophage-mediated lung cancer metastasis. *Chinese journal of natural medicines* 13:925-932. 10.1016/S1875-5364(15)30098-4

Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, and Wei L. 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic acids research* 39:W316-322. 10.1093/nar/gkr483

Xu W, Zhao Y, Qin Y, Ge B, Gong W, Wu Y, Li X, Xu P, and Xue M. 2016. Enhancement of Exposure and Reduction of Elimination for Paeoniflorin or Albiflorin via Co-Administration with Total Peony Glucosides and Hypoxic Pharmacokinetics Comparison. *Molecules* 21. 10.3390/molecules21070874

Yuan Y, Yu J, Jiang C, Li M, Lin S, Wang X, and Huang L. 2013. Functional diversity of genes for the biosynthesis of paeoniflorin and its derivatives in Paeonia. *International Journal of Molecular Sciences* 14:18502-18519. 10.3390/ijms140918502

Zha LP, Cheng ME, and Peng HS. 2012. Identification of ages and determination of paeoniflorin in roots of Paeonia lactiflora Pall. From four producing areas based on growth rings. *Microscopy research and technique* 75:1191-1196. 10.1002/jemt.22048

Zhang Y, Qiao L, Xu W, Wang X, Li H, Chu K, and Lin Y. 2017. Paeoniflorin Attenuates Cerebral Ischemia-Induced Injury by Regulating Ca(2+)/CaMKII/CREB Signaling Pathway. *Molecules* 22. 10.3390/molecules22030359

Zhang YC, Liao JY, Li ZY, Yu Y, Zhang JP, Li QF, Qu LH, Shu WS, and Chen YQ. 2014. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol* 15:512. 10.1186/s13059-014-0512-1

Zhao D, Wang R, Liu D, Wu Y, Sun J, and Tao J. 2018. Melatonin and Expression of Tryptophan Decarboxylase Gene (TDC) in Herbaceous Peony (Paeonia lactiflora Pall.) Flowers. *Molecules* 23. 10.3390/molecules23051164

Zhao D, Wei M, Shi M, Hao Z, and Tao J. 2017. Identification and comparative profiling of miRNAs in herbaceous peony (Paeonia lactiflora Pall.) with red/yellow bicoloured flowers. *Scientific reports* 7:44926. 10.1038/srep44926

Zhou H, Hu S, Guo B, Feng X, Yan Y, and Li JJYxxbApS. 2002. A study on genetic variation between wild and cultivated populations of Paeonia lactiflora Pall. 37:383-388.

Zhou JX, and Wink M. 2018. Reversal of Multidrug Resistance in Human Colon Cancer and Human Leukemia Cells by Three Plant Extracts and Their Major Secondary Metabolites. *Medicines* 5. 10.3390/medicines5040123

Figure 1

The developmental features of the 3 *P.lactiflora*s strains

(A) shows the 30-day-old plants of XT, PB, and GS strains, which were grown from the shoots (B) isolated from 3-year-old plants. (C) and (D) show the leaves and roots of 3 year-old plants.

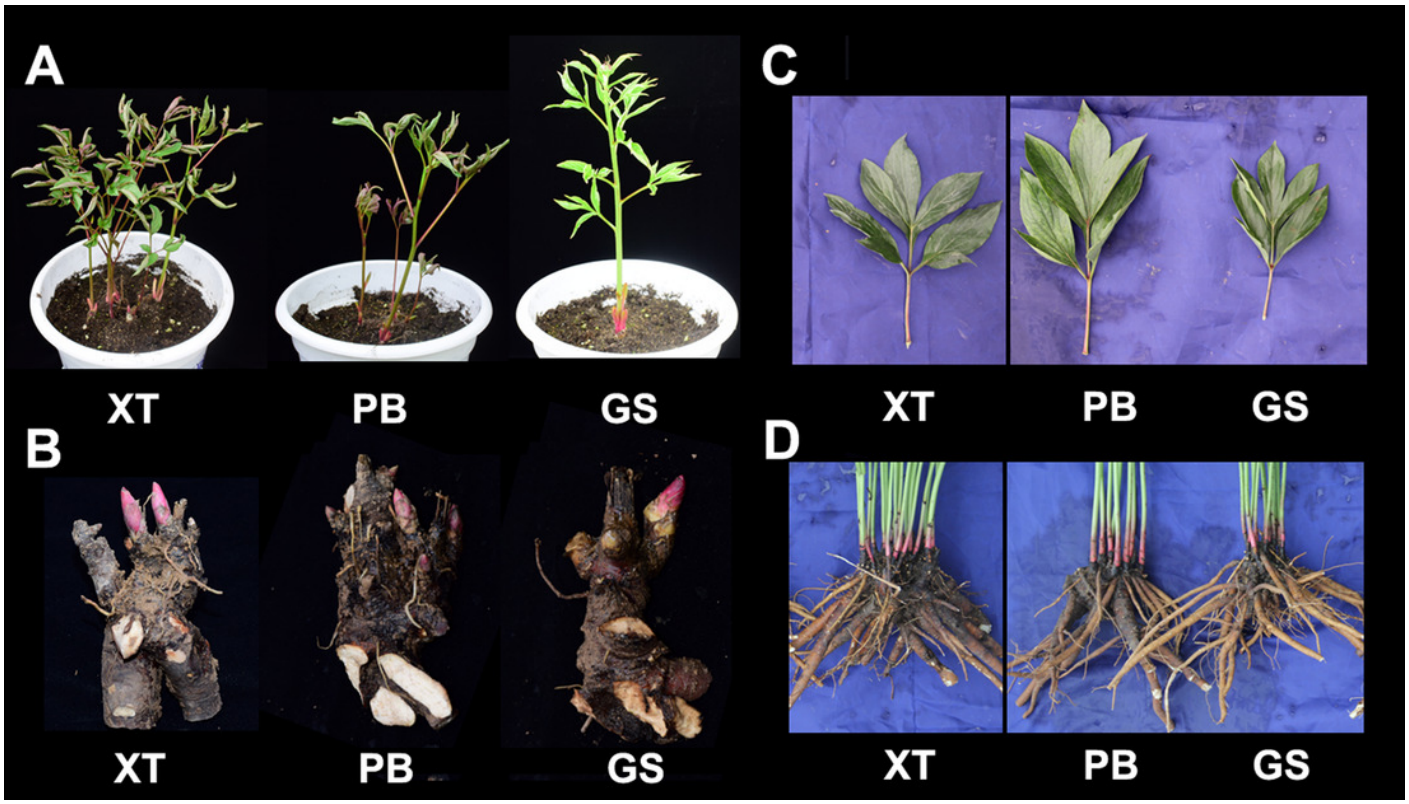

Figure 2

Figure 2. Hierarchical structure of data-sets

(A) The hierarchical structure of the gene expression levels. Genes with the FPKM values more than 1 were used for the analysis. (B) Principal component analysis for the genes.

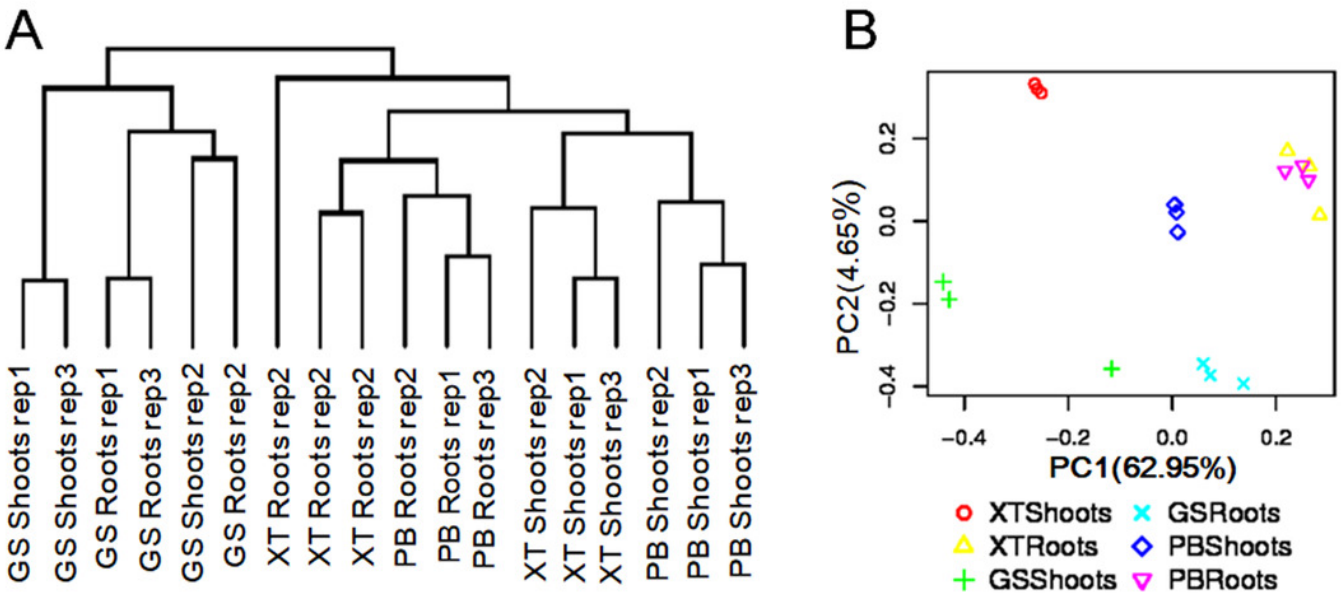

Figure 3

Figure 3. Functional annotation of unigenes

(A) The length distribution of the assembled unigenes. (B) Functional annotation of unigenes using the 4 databases. (C) Transcription factor genes identified in *P. lactiflora*.

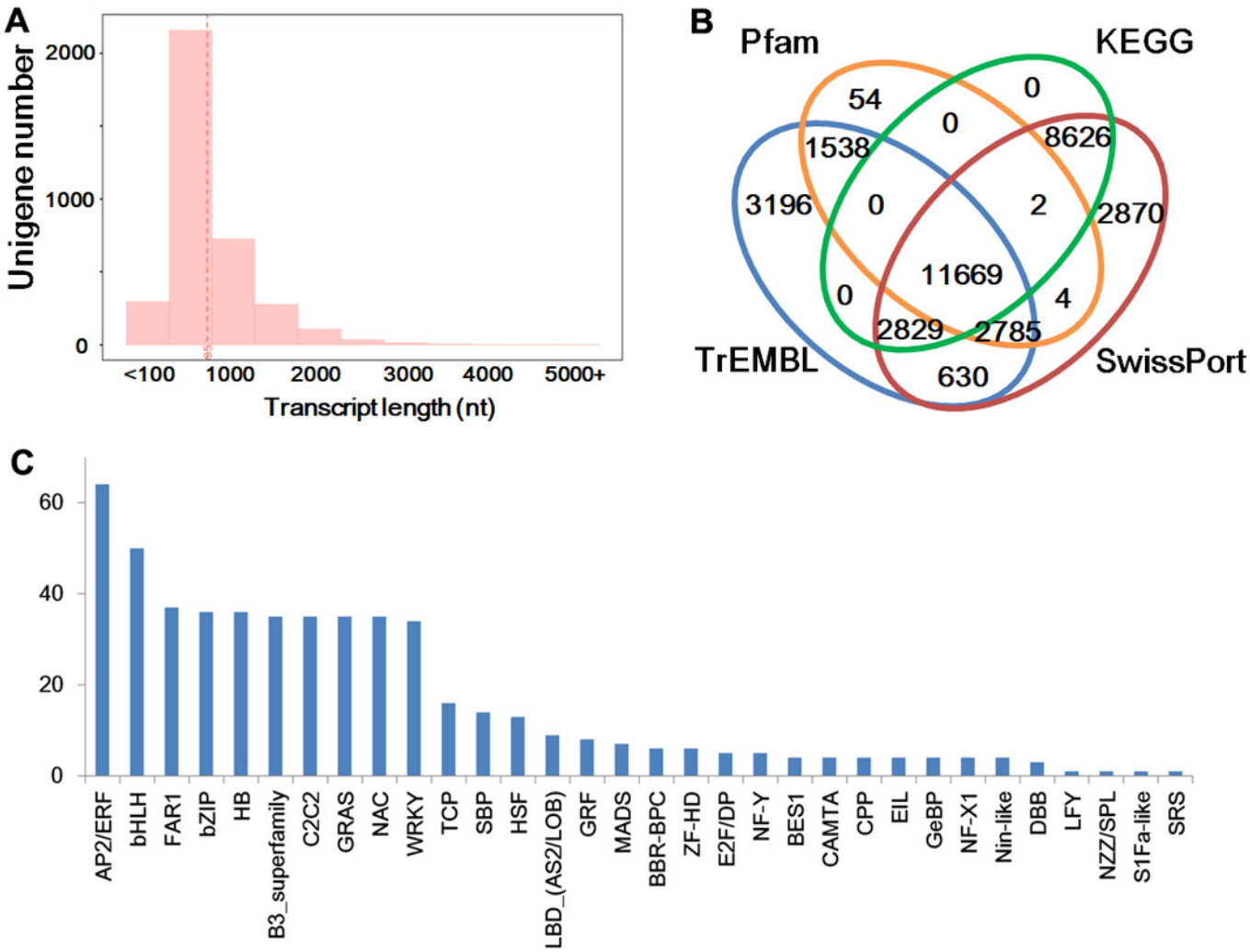

Figure 4

Figure 4. differential expression analysis of unigenes

The differentially expressed genes in roots and shoots of the three *P. lactiflora* accessions are given.

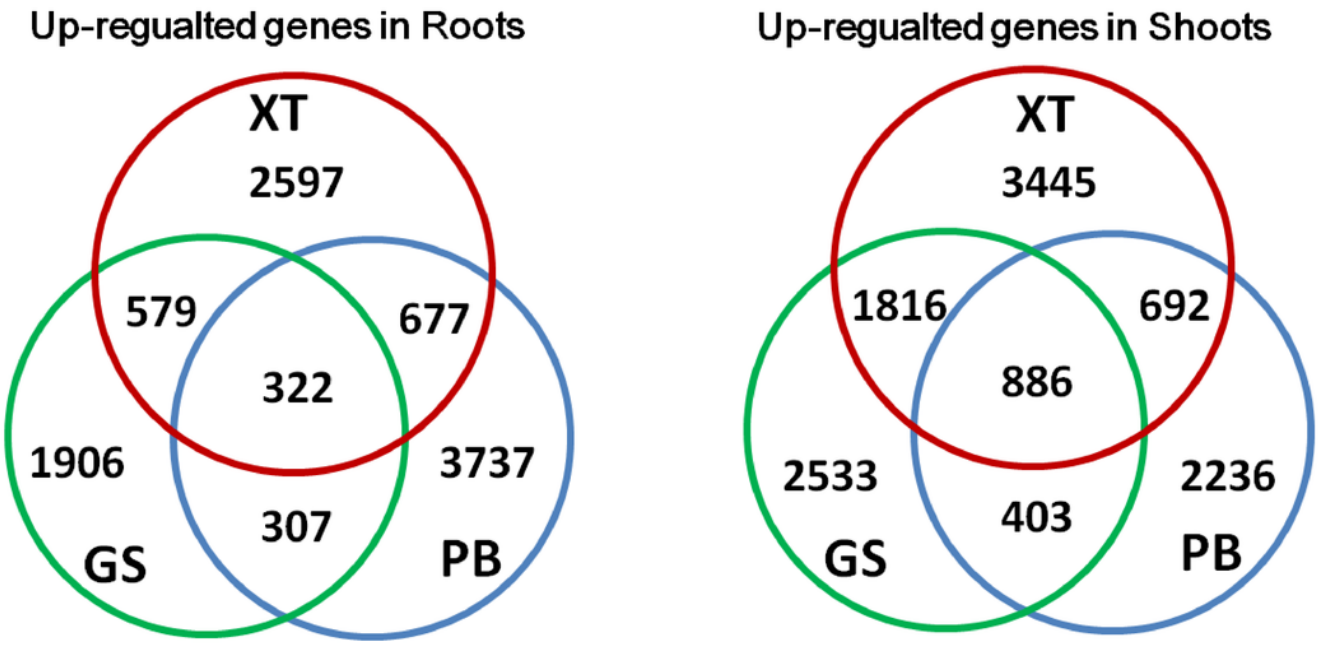

Figure 5

Figure 5. The assembled genes encoding enzymes in terpenoid backbone and monoterpenoid biosynthesis pathways

The identified *P. lactiflora* enzymes are highlighted by red rectangles.

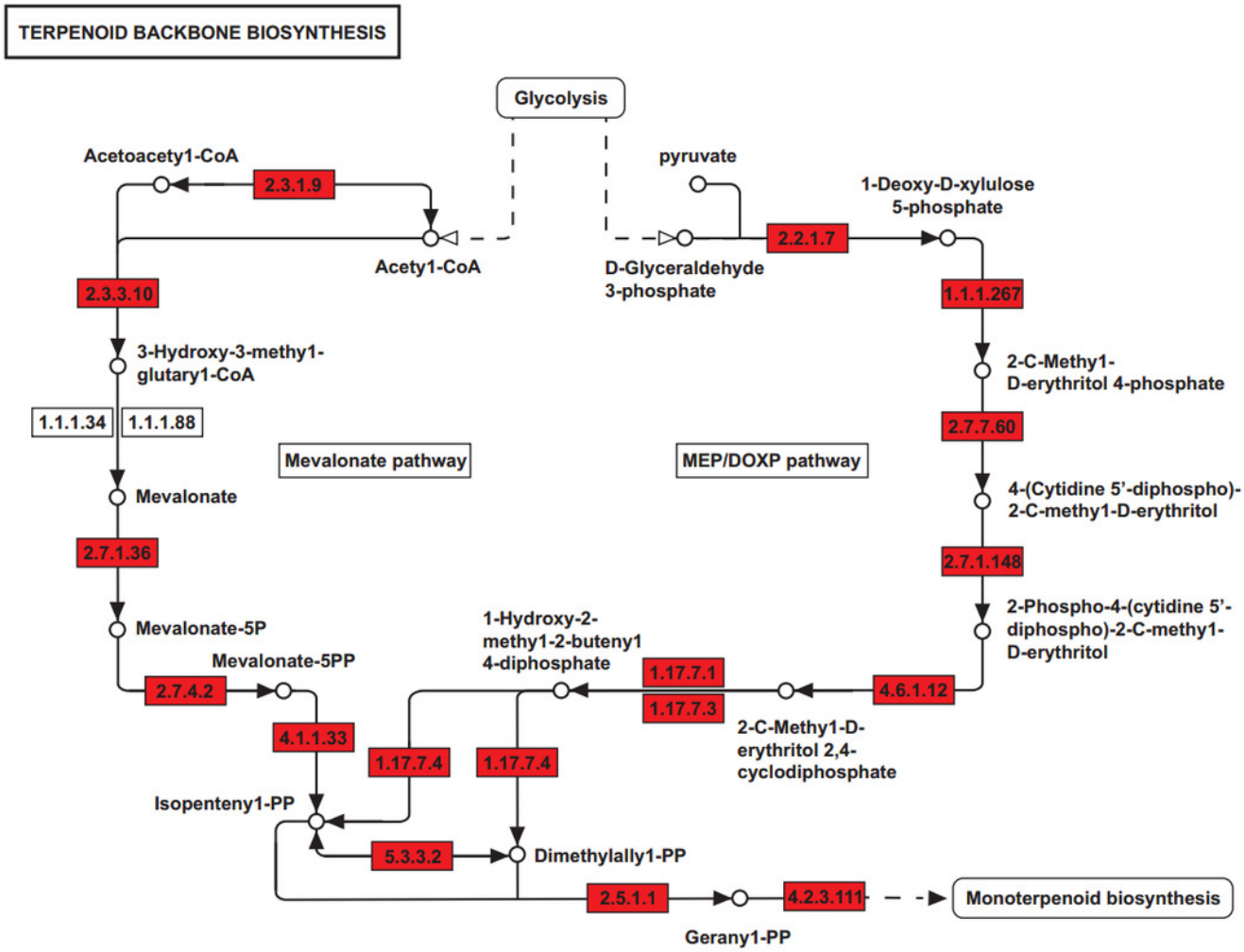

Figure 6

Figure 6. The expression levels of the genes encoding enzymes in terpenoid backbone and monoterpenoid biosynthesis pathways in shoot and root samples of the three *P. lactiflora* accessions.

The heatmap plot presents the mean value of expression levels of the three biological replicates.The unigene ID, enzyme id and symbols are given according to their expression levels. S and R give shoots and roots respectively.