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

Baowei Lu [1], **Liangjing Cao** [2], **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 differentially 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

4   Baowei Lu[1,3], Liangjing Cao[2], Qian Gao[2], Xuan Wang[2,3], Yongjian Yang[1], Pengming Liu[1],

5   Baoliang Yang[1], Tong Chen[4], Xin-Chang Li[2], Qinghua Chen[1], Fengxia An[1]*and Jun Liu[2]*

7   [1]Bozhou University, Anhui Province, 236800, China

8   [2]National Key Facility for Crop Resources and Genetic Improvement, Institute of Crop Science, Chinese

9   Academy of Agricultural Sciences, Beijing 100081, China

10   [3]Department of Biological Sciences, Xian Jiaotong-Liverpool University, Suzhou 215123, China

11   [4]National Resource Center for Chinese Materia Medica, Academy of Chinese Medical Sciences, Beijing

12   100700, China

14   Corresponding Author:

15   FengxiaAn (An_fengxia@163.com) or Jun Liu (liujun@caas.cn)

25

26

27

28

29

30

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 differentially expressed among 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* breeding in the future.

46

47

48

49

50

51

52

53

54

55

56

57

Introduction

59   The herbaceous peony (*Paeonialactiflora Pall.*) is a flowering plant in the family Paeoniaceae, which is

60   native to Central and eastern Asia (Zhao et al. 2018; Zhao et al. 2017). Its dried root is harvested without the

61   bark in the autumn from plants that are between 3-5 years of age; this harvested material is named Radix

62   Paeoniae Alba or Baishao and is a well-known Chinese herb, used for over 2000 years (He & Dai 2011; Zha

63   et al. 2012). A water/ethanol extract of Radix Paeoniae Alba, now known as Total Glucosides of Peony (TGP),

64   was originally used in the treatment of typhoid (Li et al. 2011). Subsequently, TGP has been widely

65   prescribed for fever, rheumatoid arthritis, hepatitis, muscle cramping and spasms, systemic lupus

66   erythematosus, and dysmenorrhea (Fan et al. 2012; He & Dai 2011; Ji et al. 2013; Mao et al. 2012; Nam et al.

67   2013).

68   Paeoniflorin (C23H28O11, molecular weight = 480.45) is the major medicinal component in *P. lactiflora*

69   roots. In vitro and in vivo studies in animal models have confirmed that TGP, paeoniflorin,

70   benzoylpaeoniflorin, galloylpaeoniflorin and their derivatives, are medicinally active compounds with multiple

71   pharmacological effects (Fan et al. 2012; He & Dai 2011; Zhou & Wink 2018). TGP can inhibit acute and

72   subacute inflammation, an effect which is potentially mediated by the suppression of prostaglandin E2,

73   leukotriene B4, and nitric oxide, as well as the intracellular calcium ion concentration (He & Dai 2011; Xu et

74   al. 2016). TGP has been known to protect cells against Ca^{2+} overload and oxidative stress (Zhang et al. 2017).

75   Moreover, the components of TGP, as important immunomodulatory effectors, can regulate the proliferation

76   and apoptosis of lymphocytes and balance the production of proinflammatory cytokines in a dose-dependent

77   manner (He & Dai 2011; Hu et al. 2018). In addition, paeoniflorin and its derivatives were shown to inhibit

78   tumor growth and macrophage-mediated lung metastases (Ou et al. 2011; Wu et al. 2015).

79   Paeoniflorin is a monoterpene glucoside that is biosynthesized from geranyl-pyrophosphate (GPP). GPP

80   is produced via a conversion from the universal terpenoid precursor, Isopentenyl pyrophosphate (IPP). In

81   plants and bacteria, IPP is produced from the two terpene biosynthesis pathways, the mevalonate pathway

82   (MVA), and the 1-deoxy-d-xylulose-5-phosphate/ methyl-erythritol-4-phosphate (DXP/MEP) pathway

83   (Kanehisa et al. 2012; Ren et al. 2009; Xie et al. 2011). The MVA pathway reactions take place in the cytosol

84 and are catalyzed by enzymes including hydroxyl methylglutaryl-CoA synthase, acetyl-CoA *C*-

85 acetyltransferase, HMG-CoA reductase, mevalonate kinase, and phosphomevalonate kinase. The DXP/MEP

86 pathway is catalyzed in the plastids. Pyruvate and glyceraldehyde 3-phosphate are converted by 1-deoxy-D-

87 xylulose-5-phosphate synthase and 1-deoxy-D-xylulose-5-phosphate reductoisomerase to 1-deoxy-D-xylulose

88 5-phosphate and 2-C-methyl-D-erythritol 4-phosphate, respectively. The products are subsequently catalyzed

89 by 4-diphosphocytidyl-2-C-methyl-D-erythritol (CDP-ME) synthase, CDP-ME kinase, and 2-C-methyl-D-

90 erythritol2,4-cyclodiphosphate synthase to mediate the formation of 2-C-methyl-D-erythritol 2,4-

91 cyclopyrophosphate, which is then converted to (E)-4-hydroxy-3-methyl-but-2-enyl pyrophosphate (HMB-PP)

92 by HMB-PP synthase. HMB-PP is converted to IPP and dimethylallyl pyrophosphate (DMAPP) by HMB-PP

93 reductase. IPP and DMAPP are condensed by geranyl pyrophosphate synthase to produce GPP. In addition to

94 producing a monoterpene, GPP is also a precursor to sesquiterpenes and diterpenes. The conversion from GPP

95 to alpha-terpineol is critical for producing the monoterpene, which is catalyzed by (-)-alpha-terpineolsynthase

96 (EC 4.2.3.111, RLC1). Paeoniflorin can be modified by benzoic acid and gallic acid to produce

97 benzoylpaeoniflorin and galloylpaeoniflorin, respectively. Benzoic acid and gallic acid are catalyzed by 3-

98 deoxy-7-phosphoheptulonate synthase, 3-dehydroquinate synthase, and 3-

99 dehydroquinatedehydratase/shikimate dehydrogenase.

100   With a long history of domestication and selection, the *P. lactiflora* strains used for medical purposes

101 contain high levels of paeoniflorin and are nearly completely infertile due to embryo abortion in their

102 traditional planting regions, like the Bozhou area. Strains have been reproduced through the vegetative

103 propagation of shoots for hundreds of years. Thus these *P. lactiflora* accessions are reproductively isolated and

104 may serve as suitable resources in the investigation of the genetic and molecular basis of the paeoniflorin

105 biosynthesis pathways. Using sequence homology to search for the known sequences and domains, several

106 studies identified a group of genes involved in paeoniflorin biosynthesis. For example, the previous study

107 identified 24 genes, including 8 with full-length cDNA sequences and revealed transcriptional and

108 phylogenetic associations with paeoniflorin biosynthesis (Yuan et al. 2013). However, the gene alias in the

109 paeoniflorin biosynthesis pathways and their expression patterns have not been profiled in *P. lactiflora* strains

110 on a genome-wide basis.

111   High throughput sequencing technologies have revolutionized genomic and transcriptomic studies.

112   Improved algorithms are now available for *de novo* reassembly of the transcriptome of a non-model plant

113   species without a valid reference genome sequence (Luo et al. 2017b). In this study, we assembled the

114   transcriptome of roots and shoots derived from the 3 strains of *P. lactiflora* using *de novo* assembly analysis.

115   By aligning the assembled genes with public databases, we globally annotated 34,203 unigenes in *P. lactiflora*.

116   These genes may have potential functions. For instance, our analysis identified 521 transcription factor genes.

117   Moreover, we profiled gene expression levels and identified a group of tissue- and/or strain-differential

118   expressed genes using differential expression analysis. Of these genes, we identified the candidate genes

119   associated with the nearly completed terpenoid backbone biosynthesis pathway. We verified the expression

120   pattern of a selective group of candidate genes using the qRT-PCR assay. Our study provides a valuable

121   dataset for updating our understanding of the gene regulatory network underlying paeoniflorin biosynthesis in

122   *P. lactiflora*.

Materials and Methods

123

Plant materials for RNA-Seq

124

125   The *P. lactiflora* Pu-Bang, Xian-Tiao, and Guan-Shang strains were conserved and cultivated under

126   field conditions at   Bozhou University, Bozhou, China. The shoots and roots of 3-year old plants were

127   isolated. To avoid circadian effects, we harvested all the tissues in the afternoon of the same day. The samples

128   with three biological replicates were frozen in liquid nitrogen immediately after harvesting. The isolated

129   samples and purified RNA were stored at -80 °C.

RNA extraction,library construction, and Illumina sequencing

130

131   The total RNA of individual samples was extracted and purified with the RNeasy® Plant Mini Kit

132   (QIAGEN, Germany). RNA concentration was measured using a Nanodrop 2100 spectrophotometer. RNA

133   Integrity values were checked using an Agilent Bioanalyzer. The samples with a RIN score >8.5 were used for

134   library construction (Liu et al. 2014). The sequencing libraries were generated using a NEB Next Ultra RNA

135   Library Prep Kit for Illumina (New England Biosystems), following the manufacturer's recommendations.

136   Library sequencing was performed on a Hi-Seq X10 system with 150-cycle paired-end sequencing protocol

137   (Illumina).

Bioinformatics analysis of RNA-seq datasets

138

139   Referring the methods used in recent studies with modification (Bedre et al. 2016; Lu et al. 2018), we

140 assembled the transcriptome using Trinity (Haas et al. 2013). We aligned read sequences using HISAT2 (Kim

141 et al. 2015). Fragments per kilobase of exon per million fragments mapped of assembled transcripts (FPKM)

142 were calculated and normalized using DESeq2 with global normalization parameters (Anders et al. 2014; Love

143 et al. 2014; Quinn & Chang 2016; Zhang et al. 2014). The sequences of the assembled unigenes were

144 annotated by Trinotate   (Haas et al. 2013). Coding regions of unigenes were predicted using Transdecoder

145 (Haas et al. 2013). BLAST v2.7.1 was performed to determine the sequence homology (e-value cutoff of $1e^{-5}$)

146 to UniProt/SwissProt, HMMR v3.1b2, EggNOG v4.5.1, and metabolic pathways were analyzed using the

147 Kyoto Encyclopedia of Genes and Genomes database (Kanehisa et al. 2015). For the unigene with annotated

148 homolog, we used the criterion of FPKM more than 0.8 to filter out the lowly expressed genes; for those

149 without homolg, we used FPKM more than 1 as the cut-off criterion. Differential expression analysis was

150 carried out using DESeq2 (Anders & Huber 2010). Genes with normalized fold-change greater than 2,

151 significance P-value less than 0.05, and Benjamini-Hochberg false discovery rate less than 0.1 were considered

152 to be differentially expressed genes.

Quantitative detection of paeoniflorin

153

154   We referred to the previous method in order to measure paeoniflorin in samples (Yuan et al. 2013). The

155 dried samples (0.50 g) were weighed and extracted with 50 mL of 50% aqueous methanol with ultrasonication

156 for 30 min. The extracted samples were diluted with 50 mL 50% aqueous methanol and filtered with a 0.45-

157 µm Millipore filter membrane (Millipore, MA, USA) at 25 °C. We used the Agilent 1200 LC Series (Agilent

158 Technologies, Palo Alto, CA, USA) High Performance Liquid Chromatography (HPLC) system to measure the

159 paeoniflorin abundance. The wavelength was set at 230 nm with a flow rate of 1.0 mL/min at a temperature of

160 25 °C. Standard compounds were purchased from the National Institutes for Food and Drug Control and the

161 linearity of the standard compounds was checked at seven concentration solutions.

Quantitative RT-PCR

162

163   A total of 1 to 2 µg RNA samples were treated by DNase I (RNeasy plant mini kit) and were reverse

164 transcribed with oligo (dT) primer and SuperScript III (Invitrogen). cDNA samples were analyzed using

165 quantitative PCR with SYBR Premix Ex Taq (Takara) and a Biorad CFX96 real-time PCR system. The

166   *P. lactiflora* Actin transcript sequence (Available at https://www.ncbi.nlm.nih.gov/nuccore/JN105299) was

167   used as endogenous reference genes to design primers to normalize the expression levels among samples (Qi et

168   al. 2018; Yuan et al. 2013). The qRT-PCR reactions were carried out with three- step cycles (95℃

169   denaturation, 60℃ annealingm, 72℃ amplification, 95℃ denaturation, 60℃ annealing and 95℃ denaturation)

170   and 40 cycles of amplification. We used three technical replicates to produce the average expression levels of

171   the genes relative to that of the reference gene using the $2^{-\Delta\Delta CT}$ method (Livak & Schmittgen 2001). The

172   primers are listed in Supplemental Table 1.

173   **Accession numbers**

174   The fastq-formatted RNA-seq datasets and the sampling information (accession number CRA001881,

175   https://bigd.big.ac.cn/gsa/browse/CRA001881) are publically available on the Genome Sequence Archive

176   database (BIG.Members. 2018; Wang et al. 2017).

177   # Results

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

179   The *P. lactiflora* Pu-Bang (PB) and Xian-Tiao (XT) accessions are the most widely used herbaceous peony

180   stains for medical uses due to their high levels of paeoniflorin, which is derived   mainly from their roots; whereas

181   the Guan-Shang (GS) accession contains less paeoniflorin and is usually cultivated for ornamental uses. The

182   morphological features of the 1- and 3-year old plants of the 3 strains were shown in Figure 1, respectively. We

183   determined the paeoniflorin levels of the isolated root samples using High Performance Liquid Chromatography

184   (HPLC). The results confirmed the accumulated levels of paeoniflorin in PB and XT roots compared with

185   those of GS (Supplemental Fig. 1).

186   To systematically identify genes and explore the gene expression network underlying paeoniflorin

187   biosynthesis in *P. lactiflora*, we purified the RNA samples with three biological replicates derived from the

188   shoots and roots of 3-year old PB, XT, and GS plants, and carried out the RNA-sequencing analysis using the

189   Illumina paired-end 150 bp protocol. After filtering out the low quality reads, we obtained 775.73 million

190   reads in total (Supplemental table 2). Using Trinity (Haas et al. 2013), we performed *de novo* transcriptome

191   assembly and obtained 36,264 unigenes encoding 72,910 transcripts with 986 nt contig N50 length and 42.7%

192   average GC content. We also checked the completeness of *de novo* assembled Unigenes by aligning the

193   sequences with SwissPort database (TheUniProtConsortium. 2017). The results showed that 7,131 (50%) out

194   of the 14,165 aligned sequences have more than 40% coverage of known transcripts (Supplemental table 3 &

195   4).

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

197   Due to embryo abortion and vegetative propagation, the *P. lactiflora* accessions have been undergoing

198   reproductive isolation with a long cultivation history in the Bozhou region and were thought to be genetically

199   distinct (Zhou et al. 2002). However, the genomic evidence supporting this point is still lacking. We measured

200   the expression levels of unigenes by calculating normalized Fragments Per Kilobase of exon per million

201   fragments Mapped (FPKM). The unigenes with an FPKM value higher than 1 in at least one were used to

202   perform hierarchical clustering analysis based on the Pearson correlation efficiency. We analyzed the

203   hierarchical structure of the gene expression levels on a genome-wide basis (Fig. 2A). Most of the biological

204   replicate samples belonged to the same clusters. Principal component analysis results also showed a similar

205   result confirming a high level of reproducibility of the biological replicate samples (Fig. 2B). However, the

206   tissues and strains were distributed in different cluster clades. It was noted that all the clades derived from PB

207   and XT were separated from those of GS, suggesting that the strains for medicinal uses are genetically

208   divergent from the strains used for ornamental purposes, possibly due to the selection and reproductive

209   isolation among the strains.

210   Next, we predicted the protein-coding potential for the unigenes using the Transdecoder and searched for

211   the annotation for the unigenes by aligning the assembled transcripts and predicted peptide sequences to the

212   protein sequences annotated by Swiss-Prot, TrEMBL, Pfam, and KEGG databases using Trinotate, BLASTP,

213   and BLASTX (Boutet et al. 2016; Camon et al. 2003; El-Gebali et al. 2019; Haas et al. 2013; Kanehisa et al.

214   2017). In total, we identified 34,203 unigenes containing significant matches to the annotated genes/proteins in

215   at least one database (Fig. 3B). Of them, 28,083 (82%) were reproducibly detected by at least 2 data resources.

216   The annotation information, predicted protein sequences, and FASTA-formatted sequences of these genes

217   were provided in supplemental materials that could serve as a reference annotation for future studies

218   (Supplemental table 3).

219   Transcription factors (TFs) with DNA-binding domains are the major regulators controlling the activity

220   and specificity of the gene transcription process. We predicted genes encoding TFs in our assembled unigene

221   dataset and identified 521 TF-encoding unigenes belonging to 32 TF families (Fig. 3C). Of these, AP2/ERF,

222   bHLH, FAR1, bZIP, and HB are the most abundant TF genes. The detailed information of the TF genes was

223   provided in supplemental Table 5.

224   **Identification of tissue- and/or strain-differentially expressed unigenes**

225   Next, we searched for the differentially expressed unigenes. For each strain, we compared shoots and

226   roots and extracted lists of root/shoot specific upregulated genes (fold change of expression level > 2 and false

227   discovery rate < 0.05). Our analysis identified 10,125 up-regulated unigenes in roots and 11,911 in shoots.

228   There were 1906 to 3737 unigenes specifically up-regulated in the roots and/or shoots of each strain; whereas

229   only 332 (3%) and 886 (7%) unigenes were up-regulated in the roots and shoots of all the three *P. lactiflora*

230   strains, respectively (Fig. 4). This result showed that a large number of the tissue-specific genes were also

231   strain-specific, which is consistent with the fact that the three *P. lactiflora* strains have been undergoing genetic

232   separation and selection during the last several centuries.

233   We performed the correlation analysis between the paeoniflorin levels and the expression levels of all the

234   genes in shoots and roots of the three strains (supplemental table 3). Our analysis identified 3952 genes with

235   the Pearson correlation coefficients higher than 0.6. Among the 161 gene with Pearson correlation coefficients

236   higher than 0.9, we found several genes encoding transcription factors, E3 ubiquitin-protein ligase and

237   oxidation-reduction related process, such as GT-2, ABCF4, SKP2A, FREE, GOR, TLP, RMA1H1, HAT5,

238   XBAT31 and DELLA, suggesting transcriptional and post-translational regulation may be highly associated

239   with paeoniflorin accumulation.

240

241   **Identification of genes in terpenoid and paeoniflorin biosynthesis pathway**

242   To further dissect the regulation pathways of the unigenes, we analyzed the gene list enrichment using the

243   Kyoto Encyclopedia of Genes and Genomes (KEGG) datasets and KOBAS3.0 with hypergeometric testing

244   and Benjamini and Hochberg correction (Kanehisa et al. 2012; Xie et al. 2011). In total, we identified 71

245   significantly enriched KEGG pathways in *P. lactiflora* (Supplemental Table 6). The list included several

246   KEGG terms of secondary metabolite biosynthesis, such as: Terpenoid backbone biosynthesis (P-value <

247   0.0152), Glycerophospholipid metabolism (P-value < 0.0001), Inositol phosphate metabolism (P-value <

248   0.0001), Pyruvate metabolism (P-value < 0.0002), Seleno compound metabolism (P-value < 0.0298),

249   Ascorbate and aldarate metabolism (P-value < 0.0172) and Butanoate metabolism (P-value < 0.0352).

250   *P. lactiflora* are generally known to have abundant secondary metabolites (Li et al. 2016a; Liu et al. 2017; Ma

251   et al. 2016). Our transcriptome and pathway enrichment results are consistent with the metabolism profiling

252   studies.

253   The previous studies have identified 19 EST sequences in the terpenoid backbone biosynthesis pathways in

254   *P. lactiflora*, including 7 with full-length cDNA sequences (Yuan et al. 2013). However, the genes in the

255   terpenoid backbone biosynthesis pathway have not been globally profiled and the enzyme catalyzing the

256   initiation step from GPP to monoterpenoid biosynthesis has not been identified in *P. lactiflora* yet. In our

257   datasets, we identified 32 genes with full-length CDSs encoding the enzymes controlling the major catalytic

258   reactions in MVA and MEP pathways (Fig. 5). Compared with the previous study (Yuan et al. 2013), the genes

259   encoding 10 previously reported enzymes were also identified in our study (Fig. 5). Moreover, we identified

260   the unigene (E_H33980_c1_g4, *RLC1*) encoding (-)-alpha-terpineol synthase (EC 4.2.3.111) that can catalyze

261   the conversion from GPP to alpha-terpineol (Kulkarni et al. 2013b), a monoterpene precursor of paeoniflorin.

262   Our analysis identified the genes encoding the enzymes that almost completely catalyzed the reactions from

263   the glycolysis products to the terpenoid backbone and monoterpenoid biosynthesis in *P. lactiflora*.

264   The high accumulation of paeoniflorin in XT and/or BP roots suggests some genes in the paeoniflorin

265   biosynthesis pathway may be highly expressed in XT and/or PB roots. We analyzed the expression patterns of

266   the aforementioned 33 unigenes (Fig 6). In XT roots, The expression levels of *HMGS*, *MK*, *PMK*, *GGR*, *FPS1*

267   were higher than those in GS roots; whereas the expression levels of *PMK*, *MVD*, *GGR* in PB roots were

268   higher than those in GS roots.   The other genes did not clearly reveal XT- and/or PB-root differential

269   expression patterns. Note that XT and PB presented the distinct expression pattern of these genes, which was

270   consistent with the fact that the two strains were genetically isolated with a long history (Zhou et al. 2002). To

271   confirm the expression levels determined by the RNA-seq experiment, we used a quantitative Real-time PCR

272   (qRT-PCR) assay to measure the expression levels of several unigenes (Supplemental Fig. 2). These results

273   indicated that the different biosynthetic abilities of terpenoid biosynthesis among the strains may be largely

274   contributed to by several master genes, also suggesting that the heterosis in crosses of PB and XT should be

275   considered as a potential approach to breed *P. lactiflora* strains with high paeoniflorin contant.

276

277

Discussion
278

279   The herbaceous peony has a large and complex genome. It was estimated that the genome size of the tree

280   peony, a closely related species of *P. lactiflora*, is about 12.5 Gbp. Thus, assembling the high-quality genome

281   of *P. lactiflora* remains a challenge. Recently, transcriptome studies on a large number of plant species with

282   complex genomes, such as smooth cordgrass (*Spartinaalterniflora*) (Bedre et al. 2016), buckwheat (Lu et al.

283   2018), and *Caraganakor shinskii* (Li et al. 2016b), have been carried out.   Transcriptome assembly is

284   currently a feasible and cost-efficient technology to globally identify genes in the *P. lactiflora* genome because

285   genome information is unavailable (Luo et al. 2017a; Luo et al. 2017b). In our study, we identified 36,264

286   unigenes. Using the homolog alignment analysis, we identified a large number of these assembled unigenes

287   encoded in known regulatory domains/motifs, suggesting their molecular functions. Of these, we identified

288   521 transcription factor genes belonging to 32 families and profiled their expression patterns. Our study

289   identified a large number of previously un-annotated genes in the *P. lactiflora* genome, which could supply

290   valuable molecular information for future functional studies.

291   Plant terpenoids are widely used as traditional herbal remedies and for their aromatic qualities.

292   Terpenoids are highly abundant in several accessions of *P. lactiflora,* suggesting a high potential for terpenoid

293   biosynthesis in them. However, the terpenoid biosynthetic pathways are not yet fully understood in

294   *P. lactiflora*. Previous studies have identified the 19 EST sequences in the terpenoid backbone biosynthesis

295   pathway in *P. lactiflora* (Yuan et al. 2013). Paeoniflorin is accumulating in roots of 3-year-old plants and

296   lacking in leaves and its accumulation levels are highly differed in different strains (Li et al. 2011; Yuan et al.

297   2013). The tissue- and strain-specific accumulation pattern of paeoniflorin facilitates us to explore the

298   molecular basis of paeoniflorin biosynthesis pathways. In our study, we carry out a transcriptome analysis and

299   identify 32 unigenes in the pathway and profiled their expression patterns that covered most of the reactions in

300 the terpenoid backbone biosynthesis pathway. The expression levels of *AACT1* and several transcription factor

301 genes showed specific accumulation in XT/PB roots; whereas most of the genes in the terpene backbone

302 biosynthesis showed differential expression patterns between strains rather between tissues. This result raised

303 two testable hypotheses: 1) The several genes with *AACT1* expression patterns are master genes in the pathway;

304 2) The root-specific expression patterns of these genes associated with paeoniflorin levels could be found in

305 other strains. The gene sequences and their expression patterns identified in our study could serve as a

306 reference dataset for follow up studies to test their expression specificities in more P. lactiflora strains

307 associated with paeoniflorin levels. With this data, the other unidentified gene members and the complete

308 terpenoid backbone biosynthesis pathway in *P. lactiflora* strains can be deciphered in the future, using

309 homologous cloning, phylogenetic analysis, and catalytic kinetics experiments.

310

311  Our transcriptome profiling analysis uncovered the tissue- and/or strain-differential expression patterns in

312 the three *P. lactiflora* strains, suggesting the genes in the terpenoid biosynthesis pathway were diversified

313 during the selection process. These results could supply valuable molecular information for future functional

314 studies in *P. lactiflora*. Geranyldiphosphate synthases and farnesyldiphosphate synthase catalyze the important

315 branch-point reactions from GPP to monoterpenes and sesquiterpenes. The catalytic activities of the two

316 enzymes are sensitive to temperature and metal ion concentration (Kulkarni et al. 2013a). Under high

317 temperature and Mg^{2+} rich condition, the monoterpene concentration was significant higher compared with

318 those of sesquiterpenes. In our experiment, we identified several unigenes encoding Geranyldiphosphate

319 synthases and farnesyldiphosphate synthase in *P. lactiflora*. Their catalytic activities could be investigated

320 under different temperature and metal ion conditions in follow-up studies.

321  It has been known that *P. lactiflora* is highly sensitive to the photoperiod and temperature changes. The

322 accessions grown in the Bozhou region have been undergoing reproductive isolation for a long period. Thus,

323 the genetic background of each individual might be fixed. In our study, we identified a large number of genes

324 in *P. lactiflora* and found gene expression patterns in the terpenoid pathway are highly diversified among

325 different accessions. With the sequences of these genes, the phylogenetic structures and population genetics

326   backgrounds of germplasm resources could be further investigated to elucidate the domestication and selection

327   process of *P. lactiflora.*

Conclusion

329   Using the RNA-sequencing protocol, we assembled the exon structures of 36,264 unigenes in the shoots and

330   roots of three 3-year-old *P. lactiflorais* accessions, including 28,925 differentially expressed genes. We

331   systematically annotated their molecular functions by aligning their sequences with those from multiple data

332   resources. 521 transcription factor genes were identified. Of 3,952 genes, the Pearson Correlation Coefficients

333   between the expression levels and the paeoniflorin levels were higher than 0.6. We identified 32 genes

334   encoding the enzymes controlling the major catalytic reactions in MVA and MEP pathways and one gene

335   encoding (-)-alpha-terpineol synthase. These genes contributed nearly complete terpenoid backbone

336   biosynthesis pathways. By profiling gene expression patterns associated with MVP, MEP pathways, and I-PP

337   to monoterpene conversation reactions, we uncovered several unigenes that were highly expressed in the roots

338   of high-paeoniflorin accessions, indicating that the different biosynthetic abilities of terpenoid biosynthesis

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

Author Contributions

341   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
342   biological materials and extracted the RNA and sequenced the libraries. T. C., F. A., J.L. and Y. W. analyzed
343   the datasets. B.L. J.L. and Y.W. prepared the figures and wrote the manuscript. All authors reviewed and
344   approved the final version of the manuscript.
345
346

REFERENCE

348   Anders S, and Huber W. 2010. Differential expression analysis for sequence count data. *Genome biology* 11:R106.
349   10.1186/gb-2010-11-10-r106
350   Anders S, Pyl PT, and Huber W. 2014. HTSeq–a Python framework to work with high-throughput sequencing data.
351   *Bioinformatics*:btu638.
352   Bedre R, Mangu VR, Srivastava S, Sanchez LE, and Baisakh N. 2016. Transcriptome analysis of smooth cordgrass
353   (Spartina alterniflora Loisel), a monocot halophyte, reveals candidate genes involved in its adaptation to
354   salinity. *BMC genomics* 17:657-657. 10.1186/s12864-016-3017-3
355   BIG.Members. 2018. Database Resources of the BIG Data Center in 2018. *Nucleic acids research* 46:D14-D20.
356   10.1093/nar/gkx897

357 Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, Poux S, Bougueleret L, and Xenarios I. 2016.
358   UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the
359   Entry View. In: Edwards D, ed. *Plant Bioinformatics: Methods and Protocols*. New York, NY: Springer New
360   York, 23-54.
361 Camon E, Barrell D, Brooksbank C, Magrane M, and Apweiler R. 2003. The Gene Ontology Annotation (GOA)
362   Project--Application of GO in SWISS-PROT, TrEMBL and InterPro. *Comparative and functional genomics*
363   4:71-74. 10.1002/cfg.235
364 El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A,
365   Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, and Finn RD. 2019. The Pfam protein
366   families database in 2019. *Nucleic Acids Research* 47:D427-D432. 10.1093/nar/gky995
367 Fan YF, Xie Y, Liu L, Ho HM, Wong YF, Liu ZQ, and Zhou H. 2012. Paeoniflorin reduced acute toxicity of aconitine in
368   rats is associated with the pharmacokinetic alteration of aconitine. *Journal of ethnopharmacology*
369   141:701-708. 10.1016/j.jep.2011.09.005
370 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M,
371   MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN,
372   Henschel R, LeDuc RD, Friedman N, and Regev A. 2013. De novo transcript sequence reconstruction from
373   RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* 8:1494-1512.
374   10.1038/nprot.2013.084
375 He DY, and Dai SM. 2011. Anti-inflammatory and immunomodulatory effects of paeonia lactiflora pall., a
376   traditional chinese herbal medicine. *Frontiers in pharmacology* 2:10. 10.3389/fphar.2011.00010
377 Hu B, Xu G, Zhang X, Xu L, Zhou H, Ma Z, Shen X, Zhu J, and Shen R. 2018. Paeoniflorin Attenuates Inflammatory
378   Pain by Inhibiting Microglial Activation and Akt-NF-kappaB Signaling in the Central Nervous System.
379   *Cellular physiology and biochemistry : international journal of experimental cellular physiology,*
380   *biochemistry, and pharmacology* 47:842-850. 10.1159/000490076
381 Ji Y, Wang T, Wei ZF, Lu GX, Jiang SD, Xia YF, and Dai Y. 2013. Paeoniflorin, the main active constituent of Paeonia
382   lactiflora roots, attenuates bleomycin-induced pulmonary fibrosis in mice by suppressing the synthesis of
383   type I collagen. *Journal of ethnopharmacology* 149:825-832. 10.1016/j.jep.2013.08.017
384 Kanehisa M, Furumichi M, Tanabe M, Sato Y, and Morishima K. 2017. KEGG: new perspectives on genomes,
385   pathways, diseases and drugs. *Nucleic Acids Research* 45:D353-D361. 10.1093/nar/gkw1092
386 Kanehisa M, Goto S, Sato Y, Furumichi M, and Tanabe M. 2012. KEGG for integration and interpretation of large-
387   scale molecular data sets. *Nucleic acids research* 40:D109-114. 10.1093/nar/gkr988
388 Kanehisa M, Sato Y, Kawashima M, Furumichi M, and Tanabe M. 2015. KEGG as a reference resource for gene and
389   protein annotation. *Nucleic acids research* 44:D457-D462.
390 Kim D, Langmead B, and Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat*
391   *Methods* 12:357-360. 10.1038/nmeth.3317
392 nmeth.3317 [pii]
393 Kulkarni R, Pandit S, Chidley H, Nagel R, Schmidt A, Gershenzon J, Pujari K, Giri A, and Gupta V. 2013a.
394   Characterization of three novel isoprenyl diphosphate synthases from the terpenoid rich mango fruit.
395   *Plant physiology and biochemistry : PPB* 71:121-131. 10.1016/j.plaphy.2013.07.006
396 Kulkarni R, Pandit S, Chidley H, Nagel R, Schmidt A, Gershenzon J, Pujari K, Giri A, and Gupta V. 2013b.
397   Characterization of three novel isoprenyl diphosphate synthases from the terpenoid rich mango fruit.

398   *Plant Physiology and Biochemistry* 71:121-131. https://doi.org/10.1016/j.plaphy.2013.07.006

399 Li B, Bhandari DR, Rompp A, and Spengler B. 2016a. High-resolution MALDI mass spectrometry imaging of

400   gallotannins and monoterpene glucosides in the root of Paeonia lactiflora. *Scientific reports* 6:36074.

401   10.1038/srep36074

402 Li J, Chen CX, and Shen YH. 2011. Effects of total glucosides from paeony (Paeonia lactiflora Pall) roots on

403   experimental atherosclerosis in rats. *Journal of ethnopharmacology* 135:469-475.

404   10.1016/j.jep.2011.03.045

405 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-

406   stresses on gene expression in Caragana korshinskii seedlings revealed by RNA-seq. *BMC genomics*

407   17:200-200. 10.1186/s12864-016-2562-0

408 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

409   SLMBP21 forms protein complexes with JOINTLESS and MACROCALYX as a transcription activator for

410   development of the tomato flower abscission zone. *The Plant Journal* 77:284-296.

411 Liu P, Xu Y, Yan H, Chen J, Shang E-X, Qian D-W, Jiang S, and Duan J-A. 2017. Characterization of molecular

412   signature of the roots of Paeonia lactiflora during growth. *Chinese Journal of Natural Medicines* 15:785-

413   793. https://doi.org/10.1016/S1875-5364(17)30110-3

414 Livak KJ, and Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and

415   the 2− ΔΔCT method. *methods* 25:402-408.

416 Love MI, Huber W, and Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with

417   DESeq2. *Genome biology* 15:550.

418 Lu Q-H, Wang Y-Q, Song J-N, and Yang H-B. 2018. Transcriptomic identification of salt-related genes and de novo

419   assembly in common buckwheat (F. esculentum). *Plant Physiology and Biochemistry* 127:299-309.

420   https://doi.org/10.1016/j.plaphy.2018.02.001

421 Luo J, Duan J, Huo D, Shi Q, Niu L, and Zhang Y. 2017a. Transcriptomic Analysis Reveals Transcription Factors

422   Related to Leaf Anthocyanin Biosynthesis in Paeonia qiui. *Molecules* 22. 10.3390/molecules22122186

423 Luo J, Shi Q, Niu L, and Zhang Y. 2017b. Transcriptomic Analysis of Leaf in Tree Peony Reveals Differentially

424   Expressed Pigments Genes. *Molecules* 22. 10.3390/molecules22020324

425 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,

426   Wu MQ, and Xiao XH. 2016. Metabolomics Coupled with Multivariate Data and Pathway Analysis on

427   Potential Biomarkers in Cholestasis and Intervention Effect of Paeonia lactiflora Pall. *Frontiers in*

428   *pharmacology* 7:14. 10.3389/fphar.2016.00014

429 Mao QQ, Ip SP, Xian YF, Hu Z, and Che CT. 2012. Anti-depressant-like effect of peony: a mini-review.

430   *Pharmaceutical biology* 50:72-77. 10.3109/13880209.2011.602696

431 Nam KN, Yae CG, Hong JW, Cho DH, Lee JH, and Lee EH. 2013. Paeoniflorin, a monoterpene glycoside, attenuates

432   lipopolysaccharide-induced neuronal injury and brain microglial inflammatory response. *Biotechnology*

433   *letters* 35:1183-1189. 10.1007/s10529-013-1192-8

434 Ou TT, Wu CH, Hsu JD, Chyau CC, Lee HJ, and Wang CJ. 2011. Paeonia lactiflora Pall inhibits bladder cancer growth

435   involving phosphorylation of Chk2 in vitro and in vivo. *Journal of ethnopharmacology* 135:162-172.

436   10.1016/j.jep.2011.03.011

437 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.

438   2018. Meta-analysis and transcriptome profiling reveal hub genes for soybean seed storage composition

439   during seed development. *Plant, cell & environment*. 10.1111/pce.13175

440   Quinn JJ, and Chang HY. 2016. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet*

441   17:47-62. 10.1038/nrg.2015.10

442   Ren ML, Zhang X, Ding R, Dai Y, Tu FJ, Cheng YY, and Yao XS. 2009. Two new monoterpene glucosides from Paeonia

443   lactiflora Pall. *Journal of Asian natural products research* 11:670-674. 10.1080/10286020902980087

444   TheUniProtConsortium. 2017. UniProt: the universal protein knowledgebase. *Nucleic acids research* 45:D158-D169.

445   10.1093/nar/gkw1099

446   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

447   S, Sun Y, Yu L, Lan L, Xiao J, Fang X, Lei H, Zhang Z, and Zhao W. 2017. GSA: Genome Sequence

448   Archive<sup/>. *Genomics, proteomics & bioinformatics* 15:14-18. 10.1016/j.gpb.2017.01.001

449   Wu Q, Chen GL, Li YJ, Chen Y, and Lin FZ. 2015. Paeoniflorin inhibits macrophage-mediated lung cancer metastasis.

450   *Chinese journal of natural medicines* 13:925-932. 10.1016/S1875-5364(15)30098-4

451   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

452   annotation and identification of enriched pathways and diseases. *Nucleic acids research* 39:W316-322.

453   10.1093/nar/gkr483

454   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

455   Elimination for Paeoniflorin or Albiflorin via Co-Administration with Total Peony Glucosides and Hypoxic

456   Pharmacokinetics Comparison. *Molecules* 21. 10.3390/molecules21070874

457   Yuan Y, Yu J, Jiang C, Li M, Lin S, Wang X, and Huang L. 2013. Functional diversity of genes for the biosynthesis of

458   paeoniflorin and its derivatives in Paeonia. *International Journal of Molecular Sciences* 14:18502-18519.

459   10.3390/ijms140918502

460   Zha LP, Cheng ME, and Peng HS. 2012. Identification of ages and determination of paeoniflorin in roots of Paeonia

461   lactiflora Pall. From four producing areas based on growth rings. *Microscopy research and technique*

462   75:1191-1196. 10.1002/jemt.22048

463   Zhang Y, Qiao L, Xu W, Wang X, Li H, Chu K, and Lin Y. 2017. Paeoniflorin Attenuates Cerebral Ischemia-Induced

464   Injury by Regulating Ca(2+)/CaMKII/CREB Signaling Pathway. *Molecules* 22. 10.3390/molecules22030359

465   Zhang YC, Liao JY, Li ZY, Yu Y, Zhang JP, Li QF, Qu LH, Shu WS, and Chen YQ. 2014. Genome-wide screening and

466   functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of

467   rice. *Genome Biol* 15:512. 10.1186/s13059-014-0512-1

468   Zhao D, Wang R, Liu D, Wu Y, Sun J, and Tao J. 2018. Melatonin and Expression of Tryptophan Decarboxylase Gene

469   (TDC) in Herbaceous Peony (Paeonia lactiflora Pall.) Flowers. *Molecules* 23. 10.3390/molecules23051164

470   Zhao D, Wei M, Shi M, Hao Z, and Tao J. 2017. Identification and comparative profiling of miRNAs in herbaceous

471   peony (Paeonia lactiflora Pall.) with red/yellow bicoloured flowers. *Scientific reports* 7:44926.

472   10.1038/srep44926

473   Zhou H, Hu S, Guo B, Feng X, Yan Y, and Li JJYxxbApS. 2002. A study on genetic variation between wild and

474   cultivated populations of Paeonia lactiflora Pall.   37:383-388.

475   Zhou JX, and Wink M. 2018. Reversal of Multidrug Resistance in Human Colon Cancer and Human Leukemia Cells

476   by Three Plant Extracts and Their Major Secondary Metabolites. *Medicines* 5. 10.3390/medicines5040123

477

478

479

Figure 1

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

(A-C) shows the 30-day-old plants of Xian-Tiao (XT), Pu-Bang (PB), and Guan-Shang (GS) strains, which were grown from the shoots (D-F) isolated from 3-year-old plants. (G-I) and (J-L) show the leaves and roots of 3 year-old plants, repectively.
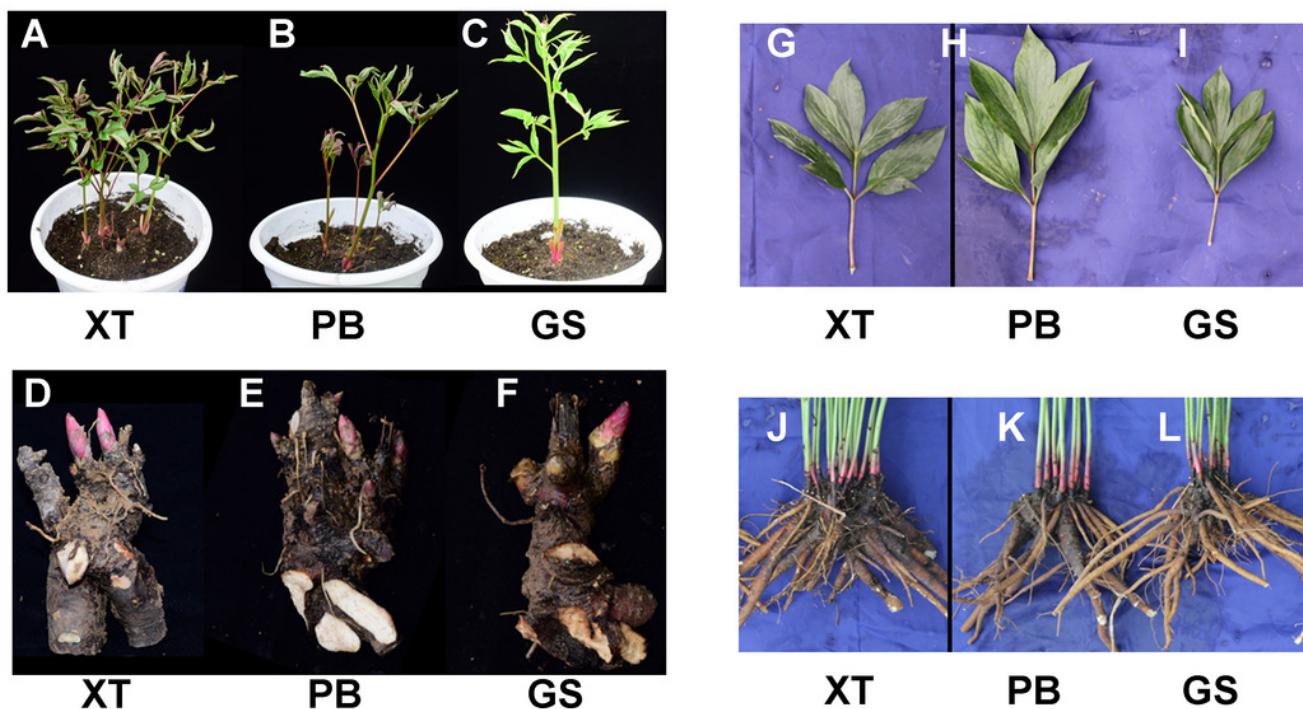
Figure 2

Paeoniflorin levels in the three *P.lactiflora* accessions

Bars show standard deviations of five biological replicates by HPLC system.
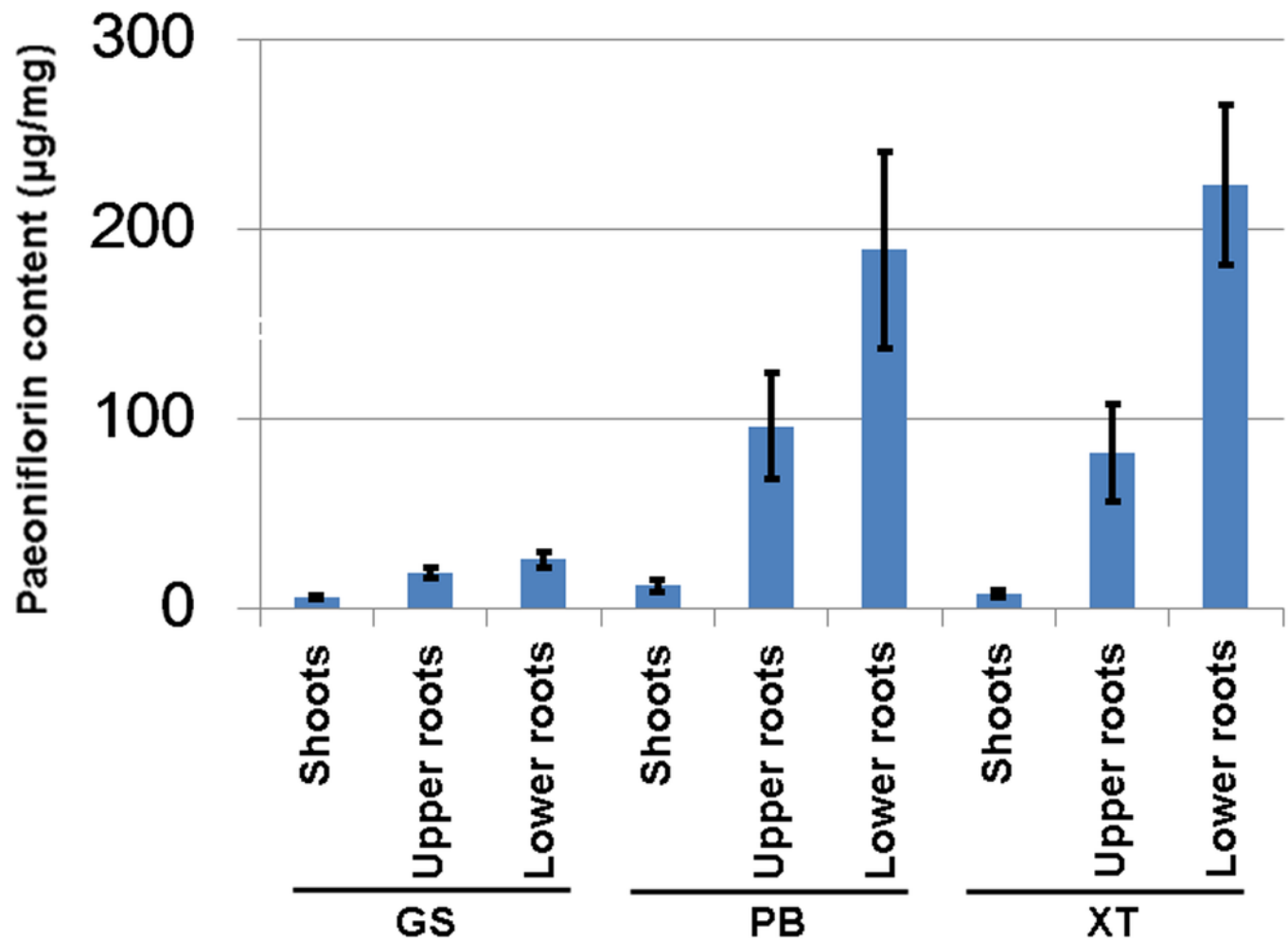
Figure 3

Functional annotation of the assembled 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*.
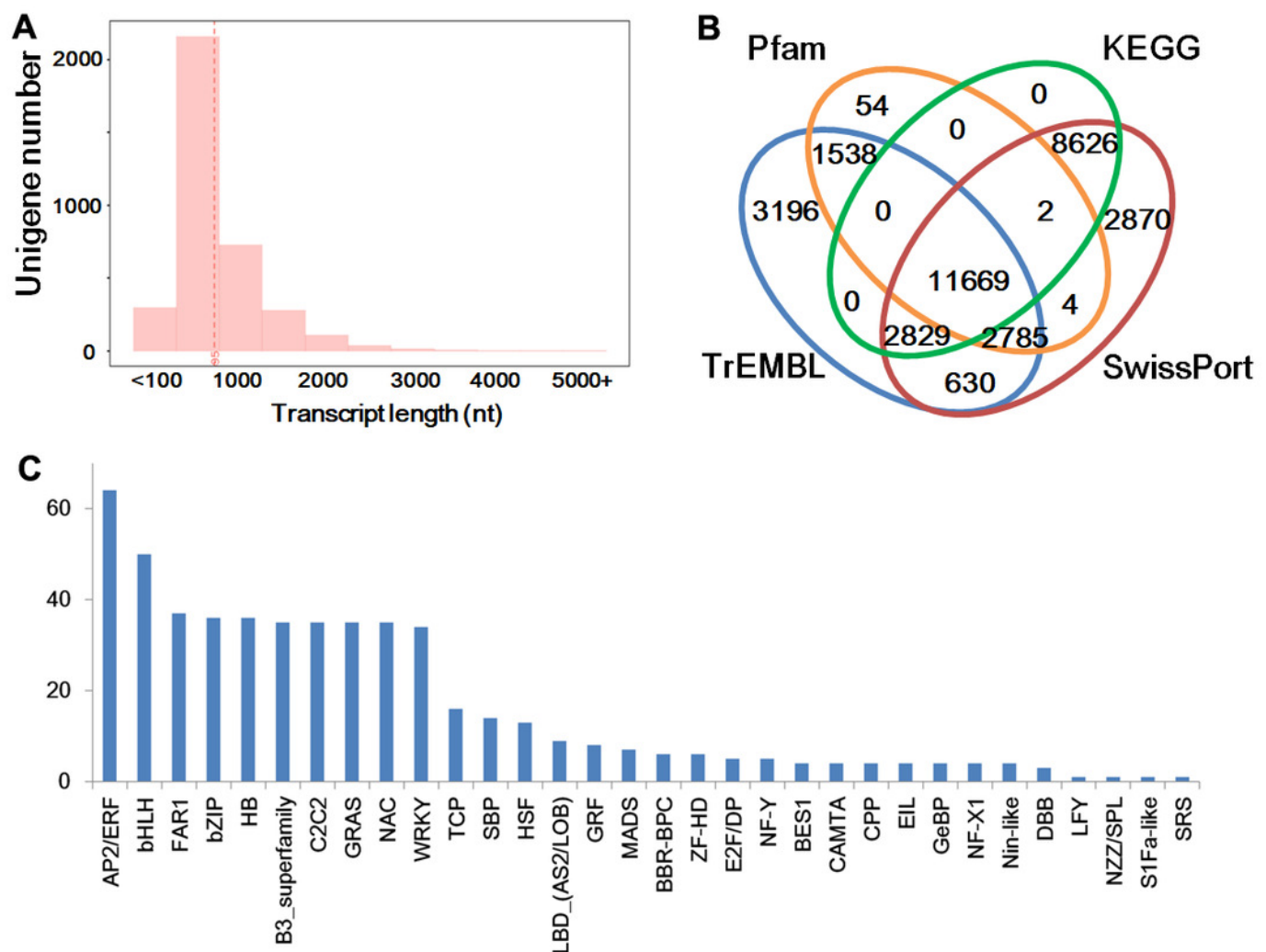
Figure 4

Identification of the differentially expressed genes

The differentially expressed genes in roots (A) and shoots (B) of the three *P. lactiflora* accessions.
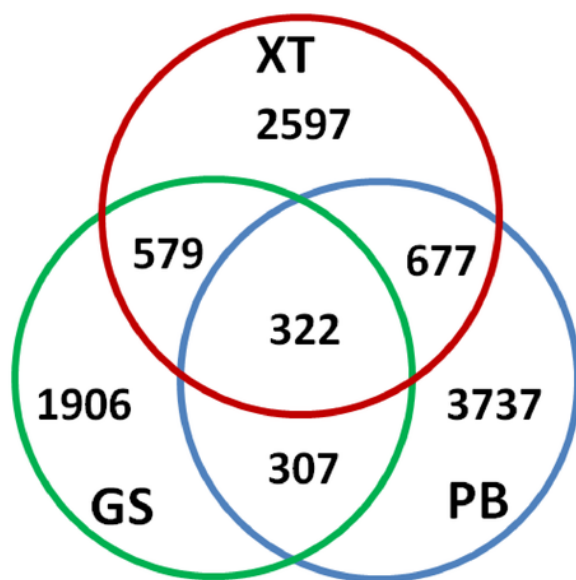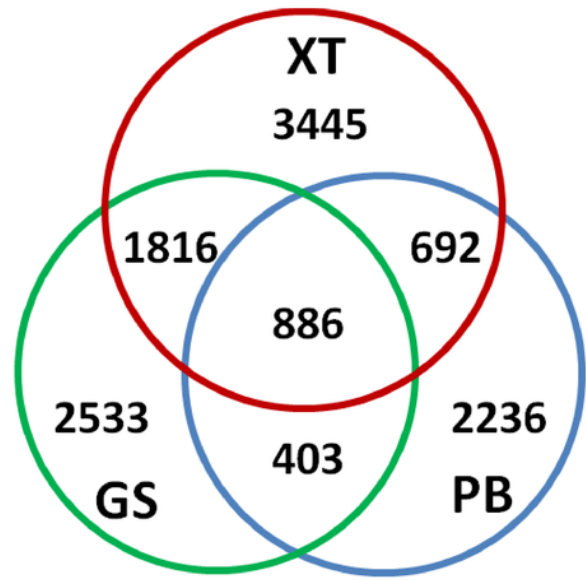
Figure 5

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

The identified *P. lactiflora* enzymes are highlighted by red rectangles. The identified enzymes with cDNA and/or Expressed Sequence Tag evidence are indicated by arrows.
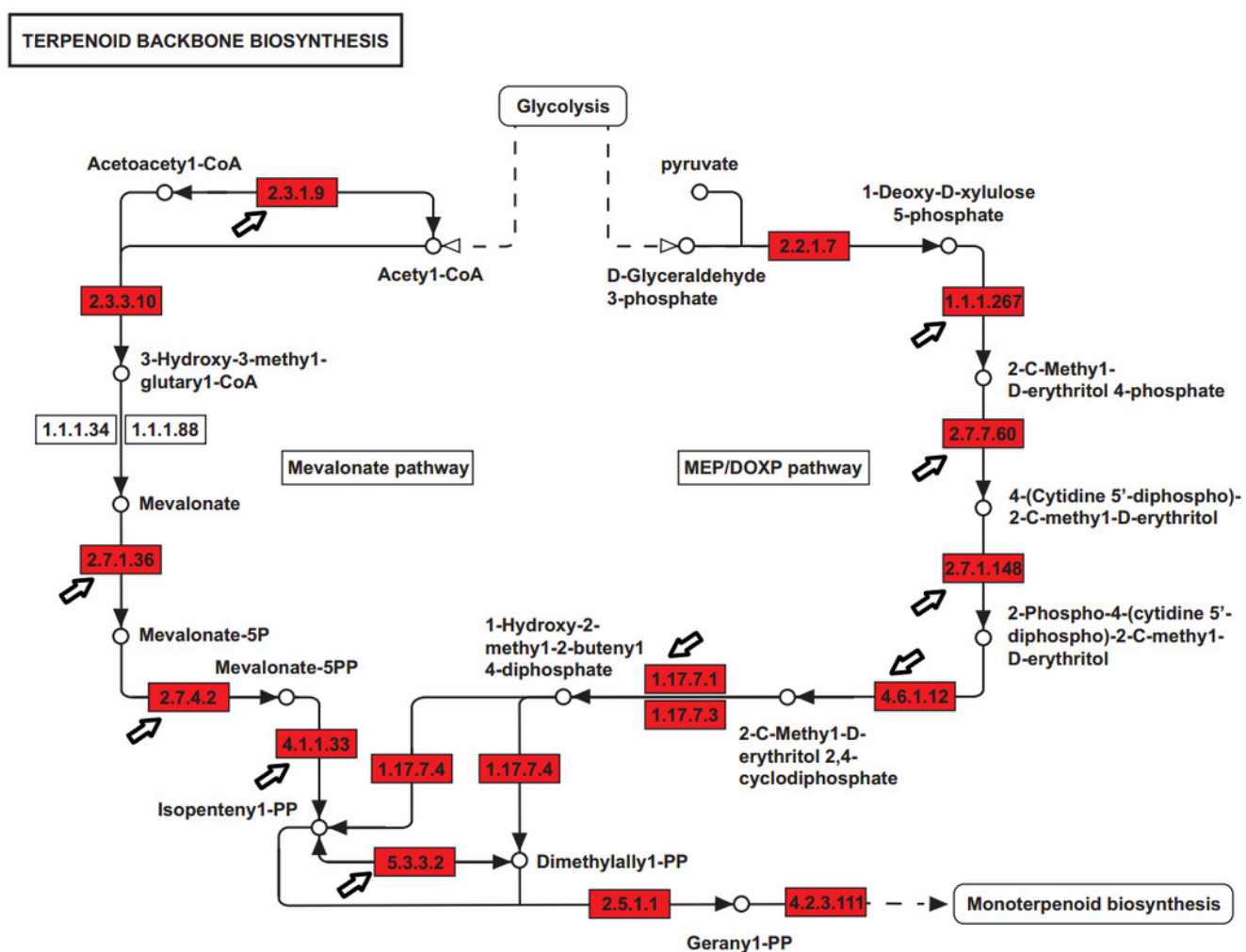
Figure 6

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.

Symbol	Enzyme	Unigene ID
AACT1	EC:2.3.1.9	E_H38048_c1_g4_i4
AACT1	EC:2.3.1.9	E_H41621_c2_g1_i2
HMGS	EC:2.3.3.10	E_H36833_c0_g1_i6
HMGS	EC:2.3.3.10	E_H36833_c0_g2_i3
HMGS	EC:2.3.3.10	E_H36833_c0_g3_i1
HMGS	EC:2.3.3.10	E_H37031_c1_g1_i10
HMGS	EC:2.3.3.10	E_H37031_c1_g3_i1
MK	EC:2.7.1.36	E_H30868_c0_g1_i1
MK	EC:2.7.1.36	E_H36964_c0_g1_i12
MK	EC:2.7.1.36	E_H41789_c1_g3_i8
PMK	EC:2.7.4.2	E_H37235_c1_g1
PMK	EC:2.7.4.2	E_H37235_c1_g3
PMK	EC:2.7.4.2	E_H37235_c1_g5
MVD	EC:4.1.1.33	E_H35912_c0_g1_i1
MVD	EC:4.1.1.33	E_H35912_c0_g10_i1
MVD	EC:4.1.1.33	E_H35912_c0_g2_i1
MVD	EC:4.1.1.33	E_H35912_c0_g5_i1
MVD	EC:4.1.1.33	E_H35912_c0_g9_i1
ISPH/HDR	EC:1.17.7.4	E_H34631_c1_g1_i6
ISPH/HDR	EC:1.17.7.4	E_H37133_c0_g2_i3
DXS	EC:2.2.1.7	E_H40099_c0_g2_i4
DXS	EC:2.2.1.7	E_H40099_c0_g3_i4
ISPD	EC:2.7.7.60	E_H33362_c0_g1_i4
ISPF	EC:4.6.1.12	E_H31492_c0_g1_i5
GPS1	EC:2.5.1.1	E_H38119_c0_g1_i23
GGR	EC:2.5.1.1	E_H36480_c0_g2_i2
GGR	EC:2.5.1.1	E_H36792_c1_g1_i3
GGR	EC:2.5.1.1	E_H36792_c1_g4_i4
FPS1	EC:2.5.1.1	E_H35482_c0_g3_i5
CDPMEK	EC:2.7.1.148	E_H37093_c0_g11_i1
CDPMEK	EC:2.7.1.148	E_H37093_c0_g7_i4
IPI2	EC:5.3.3.2	E_H38745_c1_g3_i3
RLC1	EC:4.2.3.111	E_H33980_c1_g4_i1