Available online at www.sciencedirect.com
ScienceDirect
Chinese Journal of Natural Medicines 2022, **20**(8): 614-626
doi: 10.1016/S1875-5364(22)60198-5
Chinese Journal of Natural Medicines

•Research article•

Integrative SMRT sequencing and ginsenoside profiling analysis provide insights into the biosynthesis of ginsenoside in *Panax quinquefolium*

DI Peng[1, 2, 3], YAN Yan[3], WANG Ping[3], YAN Min[3], WANG Ying-Ping[3], HUANG Lu-Qi[1*]

[1] *State Key Laboratory Breeding Base of Dao-Di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China;*
[2] *Center for Post-doctoral Research, China Academy of Chinese Medical Sciences, Beijing 100700, China;*
[3] *State Local Joint Engineering Research Center of Ginseng Breeding and Application, Jilin Agricultural University, Changchun 130118, China*

Available online 20 Aug., 2022

[ABSTRACT] *Panax quinquefolium* is one of the most common medicinal plants worldwide. Ginsenosides are the major pharmaceutical components in *P. quinquefolium*. The biosynthesis of ginsenosides in different tissues of *P. quinquefolium* remained largely unknown. In the current study, an integrative method of transcriptome and metabolome analysis was used to elucidate the ginsenosides biosynthesis pathways in different tissues of *P. quinquefolium*. Herein, 22 ginsenosides in roots, leaves, and flower buds showed uneven distribution patterns. A comprehensive *P. quinquefolium* transcriptome was generated through single molecular real-time (SMRT) and second-generation sequencing (NGS) technologies, which revealed the ginsenoside pathway genes and UDP-glycosyltransferases (UGT) family genes explicitly expressed in roots, leaves, and flower buds. The weighted gene co-expression network analysis (WGCNA) of ginsenoside biosynthesis genes, UGT genes and ginsenoside contents indicated that three UGT genes were positively correlated to pseudoginsenoside F11, notoginsenoside R1, notoginsenoside R2 and pseudoginsenoside RT5. These results provide insights into ginsenoside biosynthesis in different tissues of *P. quinquefolium.*

[KEY WORDS] *Panax quinquefolium*; SMRT sequencing; Ginsenoside biosynthesis; UDP-glycosyltransferases
[CLC Number] R284 **[Document code]** A **[Article ID]** 2095-6975(2022)08-0614-13

Introduction

Ginseng is a highly valued medicinal plant belonging to the family *Araliaceae* (genus Panax), and consists of about 14 species. *Panax ginseng*, *Panax quinquefolium*, and *Panax notoginseng* are the most commonly used types of ginseng plants worldwide. *P. ginseng* and *P. notoginseng* originated from East Asia, while *P. quinquefolium* is native to Eastern America. *P. quinquefolium* has long been cultivated as medicine in China for hundreds of years [1], and is reported to have a wide range of therapeutic and pharmacological uses, such as neuroprotective, anti-tumor, anti-diabetic, immunomodulatory, and anti-aging effects [2-7].

Ginsenosides are triterpene saponins with diverse structures, and the main medicinal ingredient in ginseng plants. To date, about 200 ginsenosides have been identified [8]. Ginsenosides are classified into four groups on the basis of the skeleton of their aglycones: PPD (protopanaxadiol)-type, including ginsenosides Rb1, Rc, Rb2, Rd, Rg3, and Rh; PPT (protopanaxatriol)-type, including ginsenosides Rg1, Re, Rg2, and Rh1; ocotillol-type, including pseudoginsenoside F11 and pseudoginsenoside RT5, and oleanolic acid-type, including ginsenoside Ro. Ginsenoside accumulation is different in various ginseng species. For instance, the ratio of Rg1/Rb1 was widely used to differentiate *P. quinquefolium* and *P. ginseng* [9]. Some ginsenosides, such as pseudoginsenoside F11 and pseudoginsenoside RT5 were found in *P. quinquefolium* alone. In contrast, ginsenoside Rf was absent in *P. quinquefolium* [10]. Previous studies mainly focused on the characterist-

[Received on] 25-Feb.-2022
[Research funding] This work was supported by the National Natural Science Foundation of China (No. 81703635), the National Key R&D Program of China (No. 2021YFD1600900) and the Jilin Province Science and Technology Development Project (No. 20210101190JC, 20200504001YY).
[*Corresponding author] E-mail: huangluqi01@126.com
These authors have no conflict of interest to declare.

ics of ginsenoside accumulation in the roots of *P. quinquefolium* [11-14], while little attention has drew on the distribution of ginsenosides in leaves [15]. The total ginsenoside content in the aerial parts of *P. quinquefolium* was higher than that in roots [15]. Therefore, characterization of the accumulation patterns of ginsenosides in different tissues like roots, leaves and flower buds of *P. quinquefolium* is important for deep understanding the biosynthesis of ginsenosides.

Although several studies have reported the biosynthetic pathway of ginsenosides in *P. ginseng* and *P. notoginseng* [16, 17], the regulatory mechanisms of ginsenosides biosynthesis in different tissues in *P. quinquefolium* are still not completely understood [18]. The precursor of terpenoid, 2,3-oxidosqualene, is considered as the precursor of ginsenosides. It is synthesized *via* the mevalonate (MVA) in the cytosol and 2-*C*-methyl-D-erythritol-4-phosphate (MEP) pathways in plastids. Oxidosqualene cyclases (OSCs), including dammarane-diol synthase (DDS) and *β*-amyrin synthase (*β*-AS), can cyclize the 2,3-oxidosqualene to generate the dammarenediol-II and *β*-amyrin backbones, respectively. Tailoring enzymes, such as some specific cytochrome P450 (CYP450s) and UDP-glycosyltransferases (UGTs), can then modify the triterpenoids backbones to generate various ginsenosides [19]. By now, the modification steps for ginsenoside biosynthesis in different tissues remained largely unknown.

Second-generation sequencing (NGS) analysis has been used in *P. quinquefolium* studies since 2010. For instance, Sun *et al.* investigated ginsenoside biosynthesis in methyl jasmonate (MeJA)-elicited roots *via* 454 pyrosequencing [20]. Moreover, there were studies concerning ginsenoside biosynthesis in different developmental stages [21], during seed dormancy [22], in MeJA-treated adventitious roots [23], and in different root tissues [14]. These studies focused on transcriptome analysis in roots, while transcriptome data in other tissues of *P. quinquefolium*, such as leaves and flower buds, has not been investigated. The different accumulation of ginsenosides in various tissues may result from different expression patterns of ginsenoside biosynthesis genes. Inadequate transcriptome information of different tissues limits the understanding of ginsenoside biosynthesis in *P. quinquefolium*.

The emergence of NGS technology has enhanced the research progress on the transcriptome and genome of medicinal plants. However, reads from NGS are usually shorter than 250 bp, resulting in incomplete transcript assembly and loss of important sequence information. Single-molecule real-time (SMRT) sequencing can be used to directly obtain all high-quality transcriptome information of single-RNA molecules from the 5′ end to the 3′ end. This method provides long read lengths, with high consensus accuracy and a low degree of bias transcriptome results [24]. SMRT sequencing has been used in many plant species, such as *Salvia miltiorrhiza* [25], *Panax notoginseng* [26], *Panax ginseng* [27], and *Fragaria vesca* [28], to characterize complex transcriptomes.

In the current study, SMRT and NGS sequencing technologies were used to obtain a high-quality *P. quinquefolium* transcriptome and examine the different gene expression patterns in various tissues. Furthermore, HPLC-ESI-Q-TRAP-MS/MS was used to assess the distribution of 22 ginsenosides in roots, leaves and flower buds of *P. quinquefolium*. All expressed genes (FPKM > 3) were clustered into modules by weighted gene co-expression network analysis (WGCNA) to identify the potential pathway genes and essential hub genes associated with ginsenoside biosynthesis. These findings provide valuable evidence for further metabolic engineering or genomics-assisted breeding of *P. quinquefolium*.

Materials and Methods

Plant materials

Four years old *P. quinquefolium* were collected from Wanliang Town, Fusong County (l42°26′ N, 127°17′ E), Jilin Province, China on June 9, 2018. The *P. quinquefolium* samples were authenticated by professor TIAN Yi-Xin (Jilin Agricultural University). The leaves, main root and flower buds from the same *P. quinquefolium* were labeled as L1, R1 and F1. Each tissue had three biological replicates. The tissues were washed and cut into small pieces, and immediately stored at −80 °C. Each sample was cut into two parts for transcriptomic and metabolomic analysis.

Metabolomic analysis

The standards of 22 ginsenosides used in this study were purchased from Shanghai Traditional Chinese Medicine Standardization Research Center (Shanghai, China). The metabolites of *P. quinquefolium* tissues were analyzed in Wuhan MetWare Biotechnology Company (http://www.metware.cn) using a widely targeted metabolomics strategy. The extraction, detection, identification, and quantification of the metabolites were conducted using the methods described by Chen *et al.* [29]. Briefly, the freeze-dried samples were crushed in a mixer mill (MM 400, Retsch) with a zirconia bead at 30 Hz for 1.5 min. The dried powder (100 mg) was extracted using 1 mL of 70% aqueous methanol. A liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS/MS) system was used to assess the extracts. A Waters ACQUITY UPLC HSS T3 C_{18} (1.8 μm, 2.1 mm × 100 mm) column was used for separation. The procedure was as follows: the solvent system consisted of water (0.04% acetic acid) and acetonitrile (0.04% acetic acid) with the gradient program of 95 : 5, *V/V* at 0 min, 5 : 95, *V/V* at 11.0 min, 5 : 95, *V/V* at 12.0 min, 95 : 5, *V/V* at 12.1 min and 95 : 5, *V/V* at 15.0 min (flow rate, 0.40 mL·min^{-1}; temperature, 40 °C and injection volume, 2 μL). An ESI-triple quadrupole-linear ion trap (Q TRAP)-MS was used for further analysis. The ginsenosides were identified by the reference standards, and quantified through multiple reaction monitoring.

RNA isolation and transcriptome sequencing

Total RNA was extracted from the samples using the

RNAprep Pure Plant Plus Kit (Tiangen Biotech, Beijing, China) in accordance with the manufacturer's instructions. RNA degradation and contamination were monitored on 1% agarose gels, the purity was measured with a NanoPhotometer spectrophotometer (IMPLEN, CA, USA), and the integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system. NGS libraries were generated using the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, USA), according to the manufacturer's instructions. The total RNA from the nine tissues was mixed in equal amounts for constructing libraries for Isoform Sequencing (Iso-Seq). The Iso-Seq library was prepared according to the Iso-Seq protocol using the Clontech SMARTer PCR cDNA Synthesis Kit and the BluePippin Size Selection System. The protocol was carried out based on the content of Pacific Biosciences (PN 100-092-800-03). NGS sequencing was performed on an Illumina Hiseq 2000 platform, and SMRT sequencing was performed on a PacBio RSII instrument of Novogene company (www.novogene.com).

Sequencing data were processed using the SMRTlink 4.0 software (Pacific Biosciences, USA). The circular consensus sequence (CCS) was generated from subread BAM files with the following parameters: Min_length 200, max_drop_fraction 0.8, no_polish TRUE, min_zscore −999, min_passes 1, min_predicted_accuracy 0.8, and max_length 18 000. The output CCS.BAM files were then classified into full-length and non-full-length reads. Then, the non-full length and full-length fasta files were sent to isoform-level clustering (ICE) and arrow polishing.

Functional annotation of transcripts

All the full-length transcripts were aligned with the NR (NCBI non-redundant protein sequences); NT (NCBI non-redundant nucleotide sequence database); Pfam (Protein family database); KOG/COG (Clusters of orthologous groups of protein database); Swiss-Prot (A manually annotated and reviewed protein sequence database); KO (KEGG Ortholog) and GO (Gene Ontology) database for functional annotation. The BLAST was used for NT database analysis with the e-value of '1e-10' [30]. The Diamond BLASTX was used for NR, KOG, Swiss-Prot, and KEGG analyses with the e-value of '1e-10'. The Hmmscan was used for Pfam database analysis.

Analysis of differentially expressed genes

The fragments per kilobase of transcript per million mapped reads (FPKM) was used to calculate the expression abundance of the unigenes. The *R* package of DESeq (1.18.0) was used for differential expression analysis [31]. DESeq provides statistical routines for digital gene expression analysis, determining differentially expressed genes by the negative binomial distribution-based model. The resultant *P*-values were adjusted using Benjamini and Hochberg's approach to correct the false discovery rate. $P < 0.05$ indicated differentially expressed genes.

Gene co-expression network analysis

The co-expression network was constructed using the WGCNA V1.69 *R* package [32]. A soft power threshold (16) was used to transform the correlation matrix into a signed weighted adjacency matrix, obtaining approximate scale-free topology for the obtained network. *R* package dynamic TreeCut (v. 1.63) was used for module detection. The minimum module size was set at 30 to prevent the generation of small modules, and the deep split was set at 2 to yield fine clusters. The modules with high similarity of eigengenes (dissimilarity, 0.25) were merged to prevent the generation of over-splitting clusters.

Statistical analysis

Standard error was calculated from three biological replicates. One-way analysis of variance (ANOVA) followed by Newman−Keuls multiple comparison were used to analyze the results. The mean values with the same letters are not significantly different from each other, while those with different letters are significantly different.

Results

Metabolomic analysis of ginsenosides in P. quinquefolium tissues

HPLC-ESI-Q-TRAP-MS/MS was used to determine the accumulation of 22 ginsenosides in the roots, leaves, and flower buds of *P. quinquefolium* (Table 1 and Table S1). A total of 21 ginsenosides were detected in all the tissues, and only notoginsenoside R1 was not detected in roots. The four types of ginsenosides exhibited different distribution patterns. Ginsenoside Ro, the oleanolic acid-type saponins, showed significantly higher production in roots than that in the aerial part. The ocotillol-type saponins pseudoginsenoside RT5 and pseudoginsenoside F11 showed a reverse pattern to ginsenoside Ro; both showed high accumulation in the aerial parts, especially in flower buds. Dammarane-type saponin content was different in various tissues. PPD-type ginsenosides Rg3, F2 and Rb3, and PPT-type ginsenosides Rf, notoginsenoside R1 and notoginsenoside R2 were significantly accumulated in the aerial parts. PPD-type ginsenosides compound K and Rb1, and PPT-type Rg2 were significantly accumulated in roots (Fig. 1). The structure and potential biosynthesis pathway of these ginsenosides are shown in Fig. S1. Overall, these results indicated that the patterns of ginsenoside accumulation in *P. quinquefolium* are tissue-specific. Therefore, the tissue-specific accumulation results combined with transcriptomic analysis can help to explore the synthetic mechanism of different ginsenosides.

Transcriptomic analysis and functional annotation of P. quinquefolium

SMRT sequencing was used to generate the full-length transcriptome of *P. quinquefolium*. NGS sequencing was used to verify the SMRT results and analyze the expression of transcripts in various tissues. SMRT sequencing yielded 18.76 gigabytes (Gb) subreads, and NGS sequencing yielded 62.2 Gb clean reads (Table S2). A total of 512 835 full-length non-chimeric reads with two primers and poly-A tail (average length: 2753 bp) were identified from the 752 712 insert

Table 1 Ginsenosides indentified from *P. quinquefolium*

No.	Compounds	Q1 (Da)	Q3 (Da) MS only	t_R/min	Molecular Weight (Da)	Ionization model	KEGG ID
1	Ginsenoside Rd	945.5501	783.40	6.00	946.55	$[M-H]^-$	C20725
2	Notoginsenoside R1	931.5345	475.40	4.27	932.53	$[M-H]^-$	C08961
3	Ginsenoside Rg1	799.4922	637.40	4.41	800.49	$[M-H]^-$	C08946
4	Pseudoginsenoside RT5	655.4343	457.20	5.60	654.43	$[M+H]^+$	-
5	Ginsenoside F2	783.4973	621.50	7.09	784.50	$[M-H]^-$	C20779
6	Ginsenoside Ro	955.4981	793.40	5.60	956.50	$[M-H]^-$	C17543
7	Pseudoginsenoside F11	801.4922	439.60	5.41	800.49	$[M+H]^+$	-
8	20(*S*)-Ginsenosides Rg3	783.4973	459.50	7.47	784.50	$[M-H]^-$	C20778
9	Ginsenoside Re	945.5501	783.50	4.37	946.55	$[M-H]^-$	C08944
10	20(*R*)-Ginsenoside Rg3	783.4973	459.50	7.56	784.50	$[M-H]^-$	C20778
11	Ginsenosides Rg2	783.4973	475.40	5.68	784.50	$[M-H]^-$	-
12	Compound K	621.4445	459.30	8.97	622.44	$[M-H]^-$	-
13	Ginsenoside Rf	799.4900	637.50	5.34	800.49	$[M-H]^-$	C08945
14	Notoginsenoside R2	769.4816	475.30	5.54	770.48	$[M-H]^-$	G00520
15	20(*S*)-Ginsenoside Rh1	637.4394	475.40	5.79	638.44	$[M-H]^-$	-
16	Ginsenosides Rb2	1077.5924	945.60	6.11	1078.59	$[M-H]^-$	-
17	Protopanaxdiol	459.3916	401.10	5.92	460.39	$[M-H]^-$	C20715
18	Ginsenoside F1	637.4394	475.10	6.11	638.44	$[M-H]^-$	C20780
19	20(*S*)-Ginsenosides Rb3	1077.5924	783.60	6.11	1078.59	$[M-H]^-$	-
20	Ginsenoside Rk1	765.4868	603.20	8.56	766.49	$[M-H]^-$	-
21	Ginsenoside Rb1	1107.6000	945.60	5.84	1108.60	$[M-H]^-$	C20713
22	Protopanaxtriol	475.3900	391.40	8.05	476.39	$[M-H]^-$	C20716

reads of the SMRT results. The SMRT non-chimeric reads were then corrected to reduce high sub-reading error rates *via* clean NGS reads. The final 241 213 non-redundant unique full-length transcripts were obtained (average length was 3221 bp, and N50 was 4851 bp) after CD-HIT was applied to remove the redundant transcripts (Table S2). Most of SMRT transcripts (91%) were longer than 600 bp. Moreover, 66.9% of the transcripts were 600−5000 bp in length and 23.3% were larger than 5000 bp.

SMRT data were annotated using NR (non-redundant), GO (Gene Ontology), COG (clusters of orthologous groups), Swiss-Port, KEGG (Kyoto Encyclopedia of Genes and Genomes), and NT (nucleotide) databases. A total of 227 752 transcripts (94.41%) were functionally annotated, of which 219 230 transcripts (90.88%) were annotated in the NR database. A total of 178 825 transcripts (74.13%) were annotated in the GO database and assigned to three functional categories, namely biological processes (BPs), cell components (CCs), and molecular functions (MFs) (Fig. 2A). Transcripts in the KEGG database were assigned to five categories, and "metabolism" was the most enriched (47 588). A total of 1867 transcripts were annotated in terpenoid and polyketide metabolism, and 1611 transcripts were annotated in the biosynthesis of other secondary metabolites (Fig. 2B). A total of 79 664 (33.02%), 183 070 (75.89%), and 193 790 (80.33%) transcripts were annotated in the KOG, Swiss-Port, and NT databases, respectively (Fig. 2C). The sequencing reads of *P. quinquefolium* were deposited in the CNGBdb (https://db.cngb.org, Accession Number, CNP0002014 and CNP0001680).

Tissue-specific gene expression of P. quinquefolium

The gene expression patterns in various tissues of *P. quinquefolium* were assessed. The cluster dendrograms showed that the above-ground and underground parts of *P. quinquefolium* were distinguished. The above-ground parts (flower buds and leaves) were also differentiated (Fig. 3). The Venn diagram showed that 54 762 transcripts were expressed in all three tissues of *P. quinquefolium*; 6246, 5107 and 10 823 transcripts were uniquely expressed in roots, leaves and flower buds, respectively (Fig. 3). The differentially expressed genes (DEG) were screened under the condition, log fold change (FC) > 1, and false discovery rate (FDR) < 0.05. The comparison between roots *vs* leaves, roots *vs* flower buds, and leaves *vs* flower buds revealed 12 036,

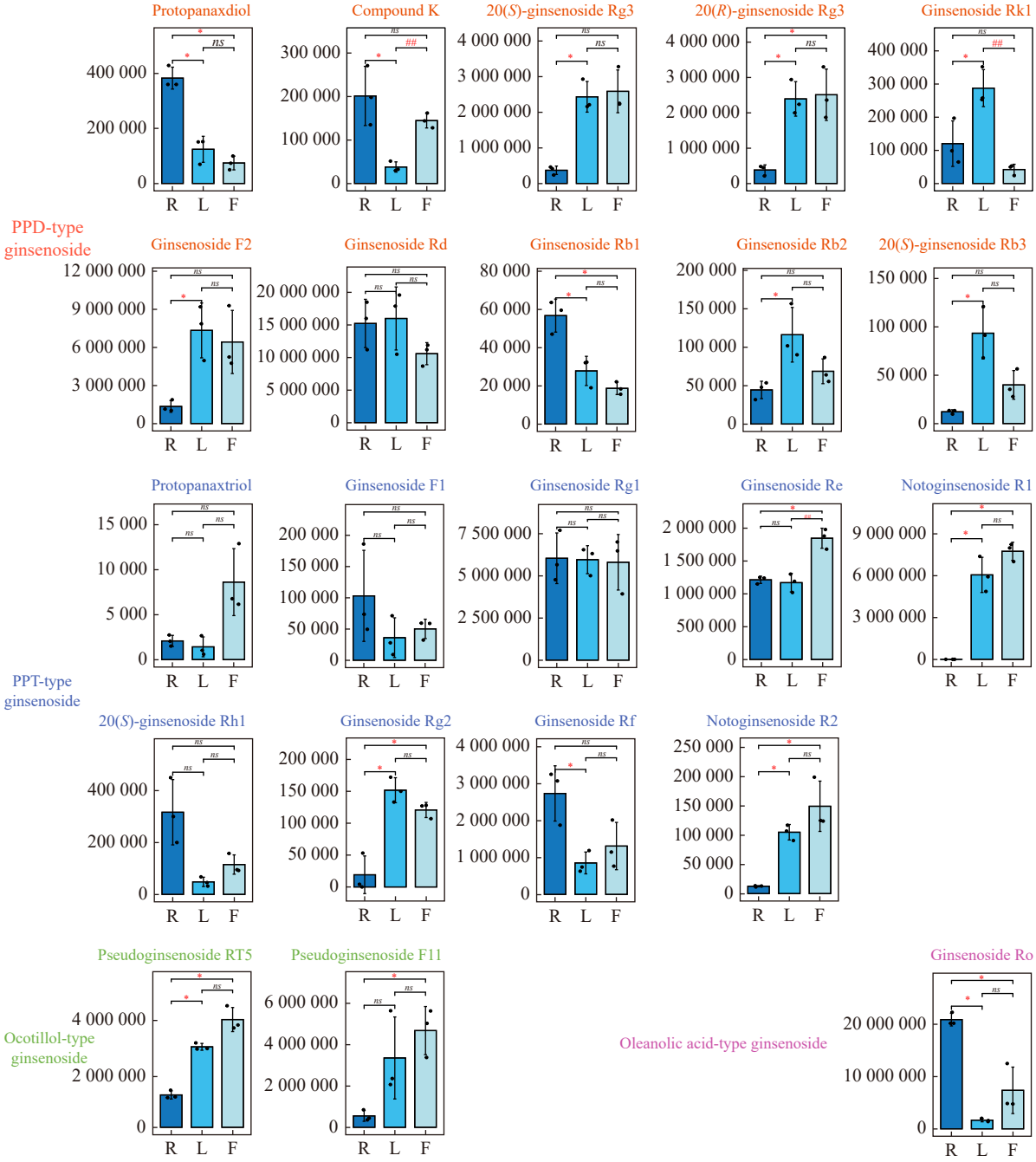

Fig. 1 HPLC-ESI-MS/MS analysis of 22 ginsenosides in roots (R), leaves (L), and flower buds (F) of *P. quinquefolium*. Data are expressed as the mean ± SD ($n = 3$). $^{*}P < 0.05$ *vs* group R; $^{\#\#}P < 0.05$ *vs* group L; ns: no significance

8841, and 4124 DEGs, respectively.

Co-expression analysis between gene expression and ginsenoside accumulation in P. quinquefolium

WGCNA analysis can identify gene modules and hub genes associated with the specific traits by analyzing gene expression patterns in various samples. A total of 31 modules containing 48 217 transcripts (FPKM > 3) were constructed (Fig. 4A). The size of each module ranged from 53 to 15 183 transcripts, and several modules showed a positive correlation to ginsenoside accumulation ($R > 0.8$, $P < 0.05$). For example, the transcripts in mediumblue module (249) were correlated with the accumulation of ginsenoside Rd ($R = 0.82$, $P < 0.05$) and ginsenoside Rb1 ($R = 0.72$, $P < 0.05$). The transcripts in darkorchid2 module (393) were positively correlated with ginsenoside Rg3 ($R = 0.8$, $P < 0.05$) and ginsenoside Rf ($R = 0.77$, $P < 0.05$). A total of 880 transcripts in module lightgoldenrod3 were correlated with ginsenoside Rb2 ($R = 0.95$, $P < 0.05$) and ginsenoside Rb3 ($R = 0.91$, $P < 0.05$). A total of 168 transcripts in module mediumorchid4 showed a positive correlation to pseudoginsenoside RT5 ($R = 0.89$, $P < 0.05$), pseudoginsenoside F11 ($R = 0.7$, $P < 0.05$), notoginsenoside R2 ($R = 0.78$, $P < 0.05$) and notoginsenos-

ide R1 ($R = 0.71$, $P < 0.05$) (Fig. 4B). These results indicated the complexity of gene regulation of ginsenoside biosynthesis in *P. quinquefolium*.

Analysis of the candidate genes involved in ginsenoside biosynthesis in different tissues

The function annotation method was used to determine

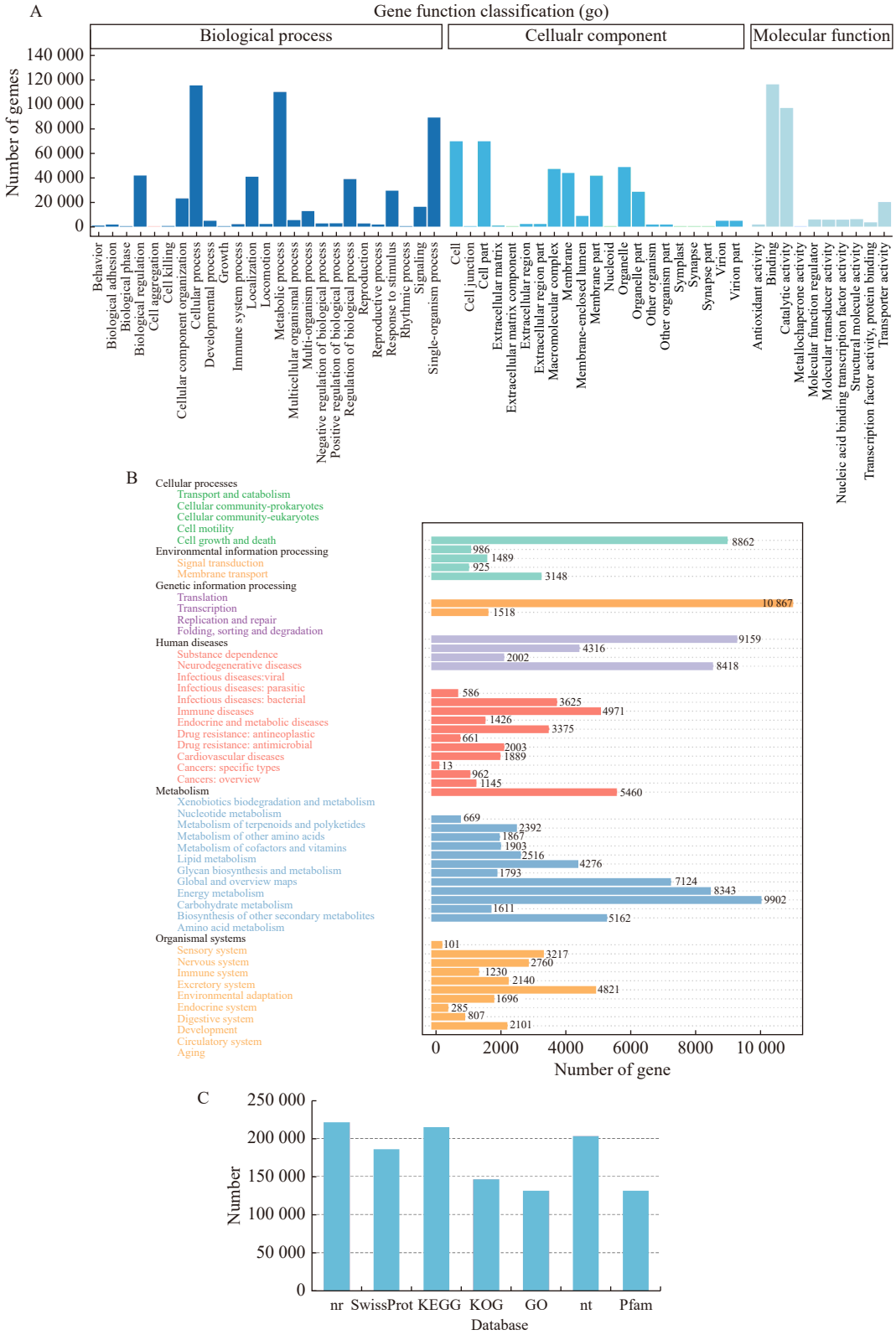

Fig. 2 Full-length transcriptome annotation of *P. quinquefolium*. (A) Gene ontology classification of transcripts; (B) KEGG annotation of transcripts; and (C) Transcripts annotated in different databases

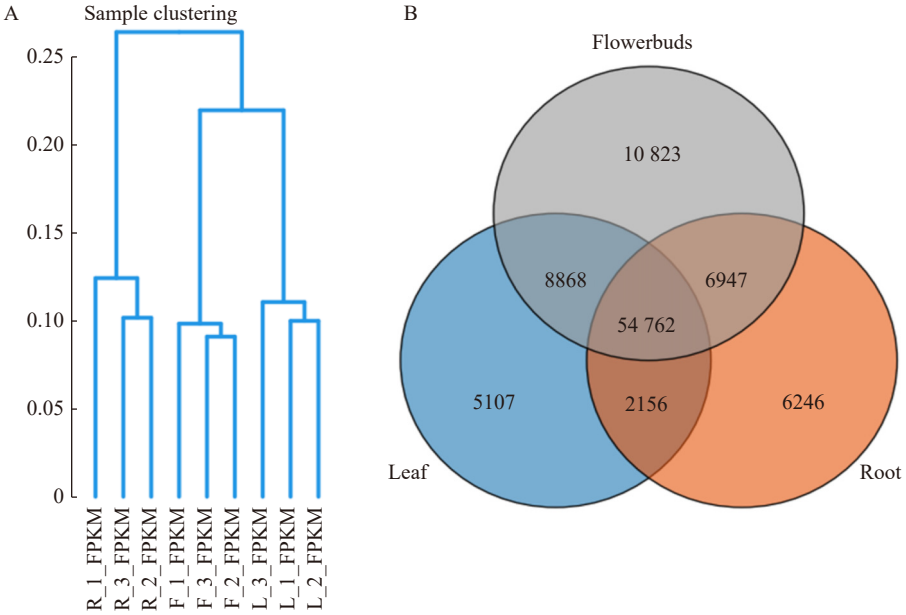

Fig. 3 Cluster analysis of gene expression in roots, leaves, and flower buds of *P. quinquefolium*. (A) Cluster tree analysis of gene expression in the samples of *P. quinquefolium*; (B) A venn diagram of gene expression among different tissues of *P. quinquefolium*

the candidate genes in the ginsenoside biosynthetic pathway. A total of 102 genes were related to the biosynthesis of the triterpene backbone, including the MVA and MEP pathways and the 2,3-oxidized squalene pathway (Fig. 5). According to the KEGG pathway annotation, 33 unigenes encodings six gene families (AACT, HMGS, HMGR, MVK, PMVK, and MVD) were related to the MVA pathway. Five unigenes were annotated to HMGR, eight AACT, seven HMGS, seven HMGR, five MK, two PMK, and four MVD. A total of 29 unigenes encoding DXS, DXR, IspD, IsPE, IspF, and IDI were enriched in the MEP pathway (Fig. 5). Moreover, 31 genes encoding five enzymes (FPPS, SS, SQE, bAS, and DDS) were associated with triterpene backbone synthesis (Fig. 5, Table S3). All the annotated genes in the pathway had more than one homologous gene. These results indicated the complex ginsenoside biosynthetic pathways in *P. quinquefolium*.

A heatmap was further constructed using the expression of candidate genes involved in the ginsenoside biosynthetic pathway to assess the tissue-specific expression pattern of the ginsenoside biosynthetic pathway (Fig. 5). The expression of the candidate ginsenoside biosynthetic pathway genes was tissue-specific. MVA pathway genes were highly expressed in flower buds, while MEP pathway genes were highly expressed in leaves compared with other tissues. DXR genes were highly expressed in flower buds, while IDI genes were highly expressed in roots compared with other tissues. Genes related to triterpene skeleton biosyntheses, such as FPPS and SS, were highly expressed in flower buds. Some SQE genes were highly expressed in leaves. Two unigenes annotated as β-AS related to oleanolic acid type saponin biosynthesis were highly expressed in flower buds and roots. Seven genes annotated as PPDS, the critical enzymes for PPD-type ginsenoside biosynthesis, were highly expressed in flower buds. Two genes annotated as PPTS were highly expressed in flower buds and roots. These results indicated that ginsenoside accumulation in various tissues may be related to the different expression patterns of those biosynthetic pathway genes.

Identification and phylogenetic analysis of UGTs in P. quinquefolium

UDP-glucosyltransferase (UGT) catalyzes ginsenoside glycosylation, which is essential for the formation of various ginsenosides. A total of 274 unigenes were annotated as UGTs from the full-length transcriptome of *P. quinquefolium*. A total of 140 UGTs were identified after filtering unigenes encoding proteins less than 300 amino acids and comparing with UGTs in other species. The protein sequences with more than 40% similarity were clustered with *Arabidopsis thaliana* into the same family (Table S4).

The 140 UGTs were divided into 14 UGT families based on sequence similarity. The UGT88 family comprised 34 genes and was the most common family. The UGT71 (27 members) and UGT85 families (19 genes) were the second and third-largest families, respectively. The UGT amino acid sequences of *P. quinquefolium* were then aligned with the UGT sequences of *A. thaliana*, *Barbarea vulgaris*, *Cicer arietinum*, *Crocus sativus*, *Glycine max*, *Glycyrrhiza uralensis*, *Medicago truncatula*, *Oryza sativa*, *P. ginseng*, *Panax japonicus* var. major, and *Zea mays* to construct a maximum likelihood (ML) phylogenetic tree (Table S5). The reported UGTs were grouped into 14 conservative groups (A to N) as in arabidopsis and two new groups as in maize (O and P) (Fig. 6) [33, 34]. UGTs from *P. quinquefolium* were divided into 11 groups. No *P. quinquefolium* UGTs was in groups I, N, H, K and C, and four *P. quinquefolium* UGTs were not in any group. Group E had the most UGTs (64), including the

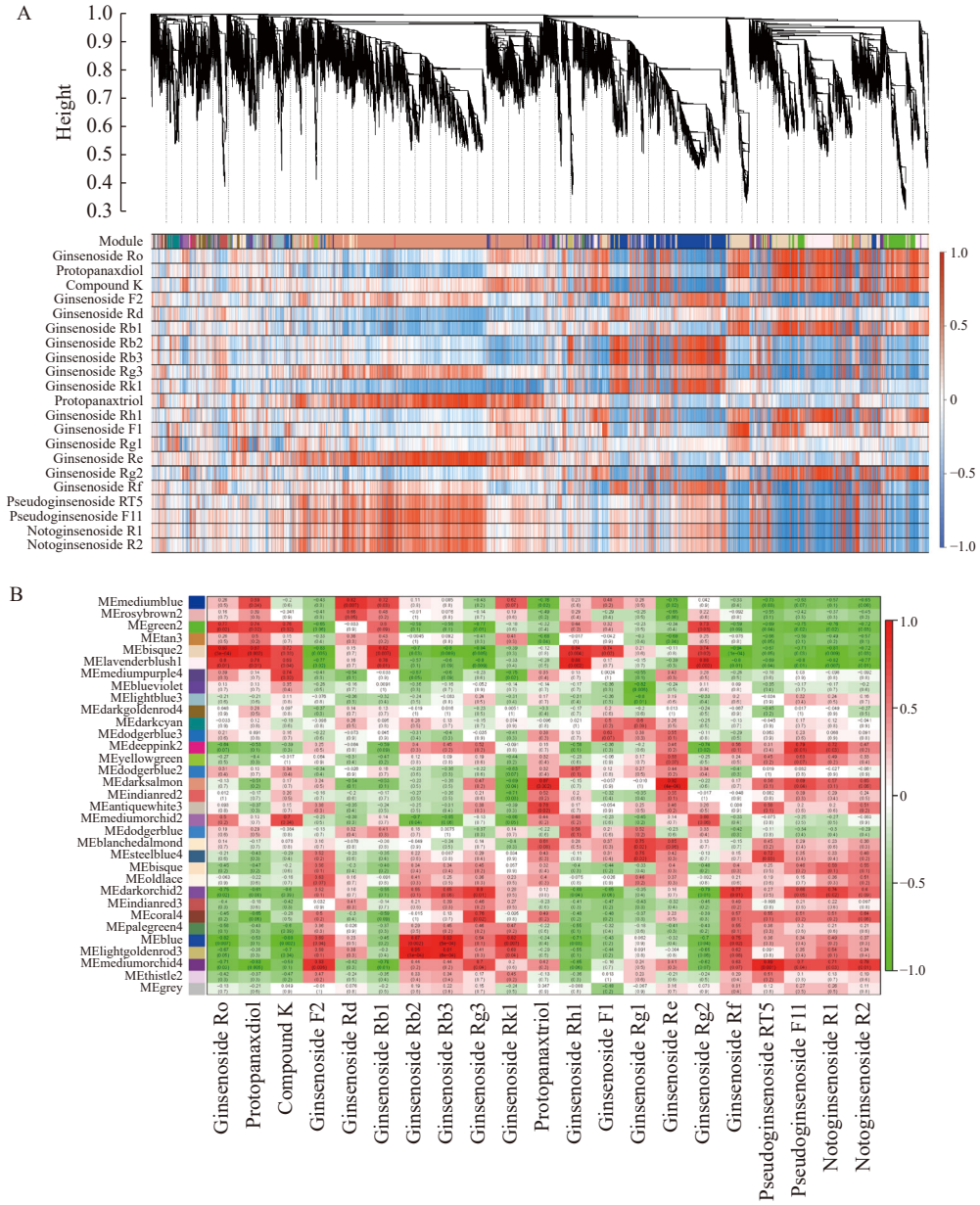

Fig. 4 Co-expression profiles of all transcripts and ginsenoside contents in *P. quinquefolium*. (A) The gene-trait association diagram in *P. quinquefolium*; (B) The relation of modules and ginsenosides in *P. quinquefolium*

UGT71 family (26) and UGT88 family (34). Group F, group O, and group M had only one UGT (Fig. 6).

Hierarchical cluster analysis was used to assess the PqUGT tissue-specific expression patterns in *P. quinquefolium*. Then, the expression profile heatmap of the UGTs was constructed based on their FPKM-normalized expression values. After filtering the non-expressed genes, a total of 140 UGTs were identified and grouped into seven clusters (P1−P7; Fig. S2). UGTs in clusters P6 (25 genes), P1 (15 genes), and P3 (32 genes) were highly expressed in roots, leaves, and flower buds, respectively. In contrast, UGTs in clusters P2 (18 genes), P4 (12 genes), and P7 (17 genes) were lowly expressed in roots, leaves, and flower buds, respectively. According to UGT expression analysis, UGT expression was highest in flower buds (26 genes, FPKM > 10), followed by leaves (25 genes, FPKM > 10) and roots (20 genes, FPKM > 10) (Fig. S2). These results indicated that UGT modification mainly occurs in the aerial parts, which explains the presence of various secondary metabolites in leaves and flower buds.

Co-expression analysis of ginsenoside accumulation and PqUGT genes

WGCNA analysis showed that many modules were positively correlated to the accumulation of ginsenosides. For instance, oleanolic acid-type saponin ginsenoside Ro and module bisque2 showed a positive correlation (R = 0.93, P < 0.05), while ginsenosides Rb2 (R = 0.95, P < 0.05) and Rb3 (R = 0.91, P < 0.05) showed positive correlations to module

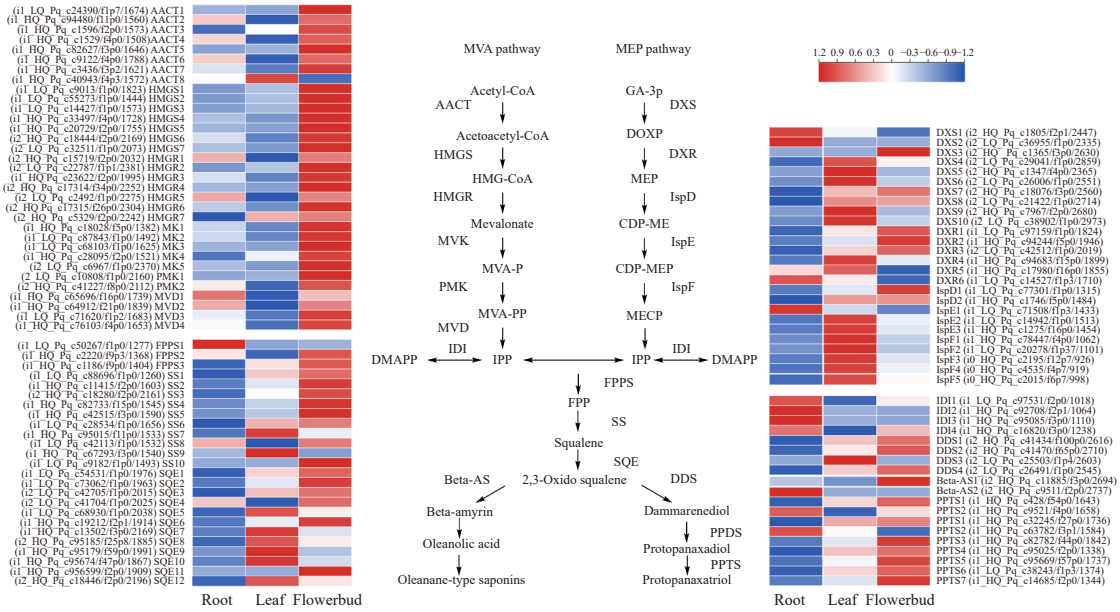

Fig. 5 The expression of the annotated genes in the MVA and MEP pathways in *P. quinquefolium*. The expression patterns of the candidate MVA and MEP pathway genes in three tissues were illustrated on the heatmap. Protopanaxatriol synthase (PPTS); Protopanaxadiol synthase (PPDS); Dammarenediol II synthase (DDS); *β*-Amyrin synthase (*β*-AS); Squalene epoxidase (SQE); Squalene synthase (SS); Farnesyl diphosphate synthase (FPPS); Isopentenyl-diphosphate delta-isomerase (IDI); Mevalonate diphosphate decarboxylase (MVD); Phosphomevalonate kinase (PMK); Mevalonate kinase (MVK); 3-Hydroxy-3-methylglutaryl-CoA synthase (HMGS); 3-Hydroxy-3- methylglutaryl-CoA reductase (HMGR); Acetyl-CoA *C*-acetyltransferase (AACT); 1-Deoxy-D-xylulose-5-phosphate synthase (DXS); DXP reductoisomerase (DXR); 2-*C*-methyl-D-erythritol 4-phosphate cytidylyltransferase (IspD); 4-(Cytidine 5-diphospho)-2-*C*-methyl-Derythritol kinase (IspE); and 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF)

lightgoldenrod3, and ginsenoside Re showed a positive correlation to module darksalmon (R = 0.92, P < 0.05). The annotation and identification of functional genes in these highly correlated modules provide insights into ginsenoside biosynthesis.

The 140 UGTs were screened in the modules, and highly related to ginsenoside accumulation. Finally, 50 UGTs (FPKM > 3) were selected in the WGCNA network (Fig. 7). The 50 UGTs were related to various patterns of ginsenoside accumulation. Genes in UGT families 71, 73, 74, and 91 showed a significantly positive correlation to several ginsenosides (R > 0.9, P < 0.05). The correlations of UGT713 to ginsenoside Ro and PPD were positive (R > 0.9, P < 0.05), UGT715 showed positive correlations to PPD, ginsenoside Rb1 and Rh1 (R > 0.9, P < 0.05), UGT714 was positively correlated to ginsenoside Rb3 (R = 0.901, P < 0.05), and UGT719 was positively correlated to pseudoginsenoside RT5 (R = 0.91, P < 0.05). Moreover, UGT731 and UGT735 showed significantly positive correlations to ginsenoside Rb3 (R = 0.94, P < 0.05), UGT739 and UGT743 showed positive associations with ginsenoside Ro (R = 0.90, P < 0.05), and UGT748 was positively correlated to ginsenoside Ro and PPD (R > 0.9, P < 0.05). UGT913 and UGT914 were positively correlated to pseudoginsenoside F11, notoginsenoside R1, and notoginsenoside R2 (R > 0.9, P < 0.05), UGT913 was positively related to ginsenoside Rg3, and UGT914 was positively associated with ginsenosides Rb2 and Rb3 (Table S6). These complex correlations revealed the important role of UGTs which are related to the formation of various ginsenosides.

Discussion

P. quinquefolium and *P. ginseng* are the most widely used herbal medicines in the ginseng genus worldwide. Ginsenosides are the main bioactive compounds in the ginseng genus and have various structures. Herein, 22 ginsenosides, including one oleanolic acid-type, seven dammarane PPD type, eight dammarane PPT type, and two dammarane ocotillol-type, were identified in *P. quinquefolium* using HPLC-ESI-Q-TRAP-MS/MS. Analyzing the accumulation patterns of these ginsenosides in different tissues can provide insights into the biosynthesis of ginsenoside in *P. quinquefolium*. Many studies have been conducted to investigate ginsenosides in *P. quinquefolium* [11, 15, 35, 36]. However, the ginsenoside profiles in flower buds and ginsenoside accumulation patterns in various tissues have not been well studied [37]. The comparative study of ginsenoside accumulation in various tissues indicated that ginsenosides showed diverse accumulation patterns. The ocotillol-type ginsenosides were highly accumulated in the aerial parts, and the oleanolic acid-type ginsenosides were highly accumulated in roots. The PPT-type and the PPD-type ginsenosides did not show a single accumulation trend in roots or the aerial parts. These diverse accumulation patterns might be related to the complex regulatory

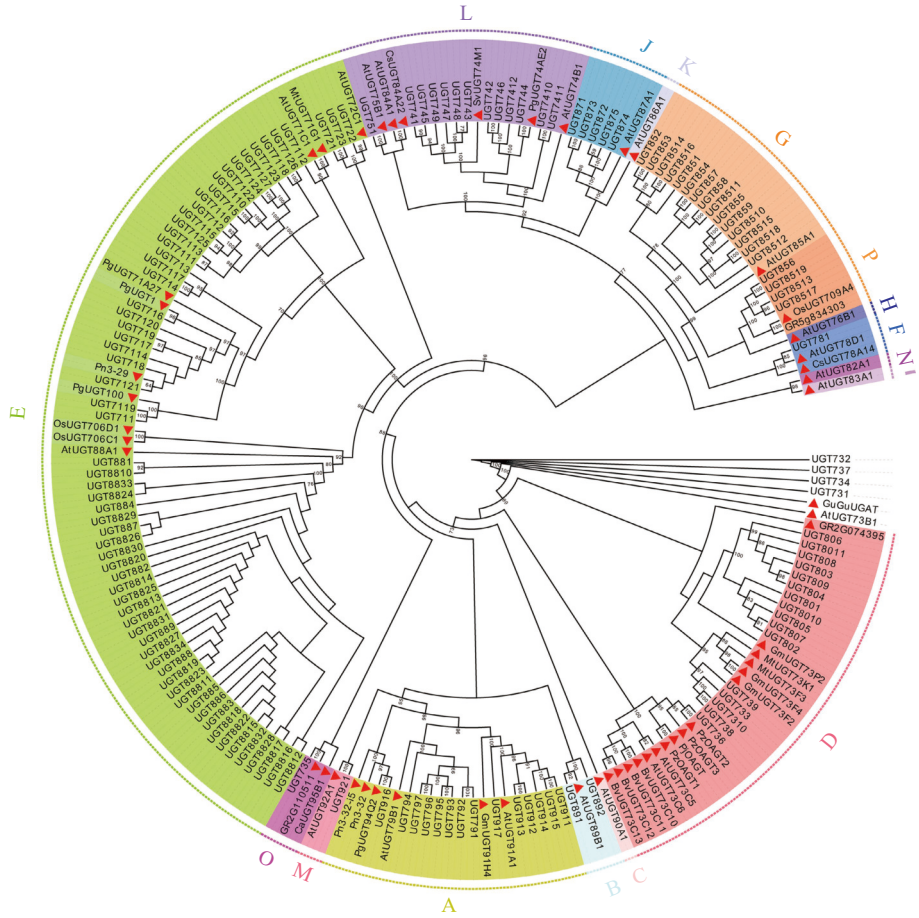

Fig. 6 Phylogenetic analysis of *P. quinquefolium* UGT genes. Red triangles indicate the UGTs from other plants

mechanism of ginsenoside biosynthesis.

The reported *P. quinquefolium* transcriptome studies were conducted based on NGS sequencing [20-23]. Due to the short reads of NGS sequencing, some important information of biosynthesis genes may be lost. Compared with NGS sequencing, SMRT sequencing can achieve a read length of 20 kb and is ideal for accurate sequencing of full-length transcripts [38, 39]. Our SMRT sequencing results of *P. quinquefolium* annotated more abundant transcripts (227 752) than the previous NGS sequencing results. Based on the SMRT sequencing results, we identified 105 full-length candidate ginsenoside pathway genes and 140 UGTs. All these findings will accelerate the elucidation of the ginsenoside biosynthesis pathway in *P. quinquefolium*.

Ginsenosides are unevenly distributed in ginseng plants [40, 41]. For instance, saponin content is significantly higher in flowers than in roots [42], and ginsenoside content is higher in wild *P. quinquefolium* leaves than in roots [43]. Such tissue-specific ginsenoside accumulation was also reflected here. The tissue specificity may be related to the differential expression of essential biosynthetic genes. Herein, the unigenes annotated on the MVA pathway were highly expressed in flower buds, and only ACCT8 and MVD1 were highly expressed in leaves and roots, respectively. Compared with leaves, some of the MVA pathway genes were relatively higher expressed in roots, indicating that the MVA pathway is the primary source of ginsenoside biosynthesis precursors in roots and flower buds. The MEP pathway genes were mainly expressed in leaves, indicating that the MEP pathway is the main source of ginsenoside biosynthesis precursors in leaves, which is similar to previous studies [44]. The genes involved in the synthesis of 2,3 oxide squalene were mainly expressed in the aerial parts. *β*-AS1 and *β*-AS2, which may control the synthesis of oleanolic acid type ginsenosides, were highly expressed in roots and flower buds, respectively. The different levels of the two genes might be associated with the accumulation of the oleanolic acid type ginsenosides in different tissues. The metabolomic analysis (Fig. 1) also showed that the total ginsenoside Ro was mainly distributed in roots and flower buds. Four transcripts were annotated as DDS, the key enzymes for dammarane-type ginsenoside biosynthesis primarily expressed in the aerial parts. For instance, DDS3 was highly expressed in leaves, whereas other genes were highly expressed in flower buds, indicating that the skeleton of dammarane-type ginsenosides might be biosynthesized in the above-ground part. PPDS and PPTS control the formation of the diol and triol type ginsenosides and are highly expressed in roots. Herein, PPTS1 and the other six PPDS were

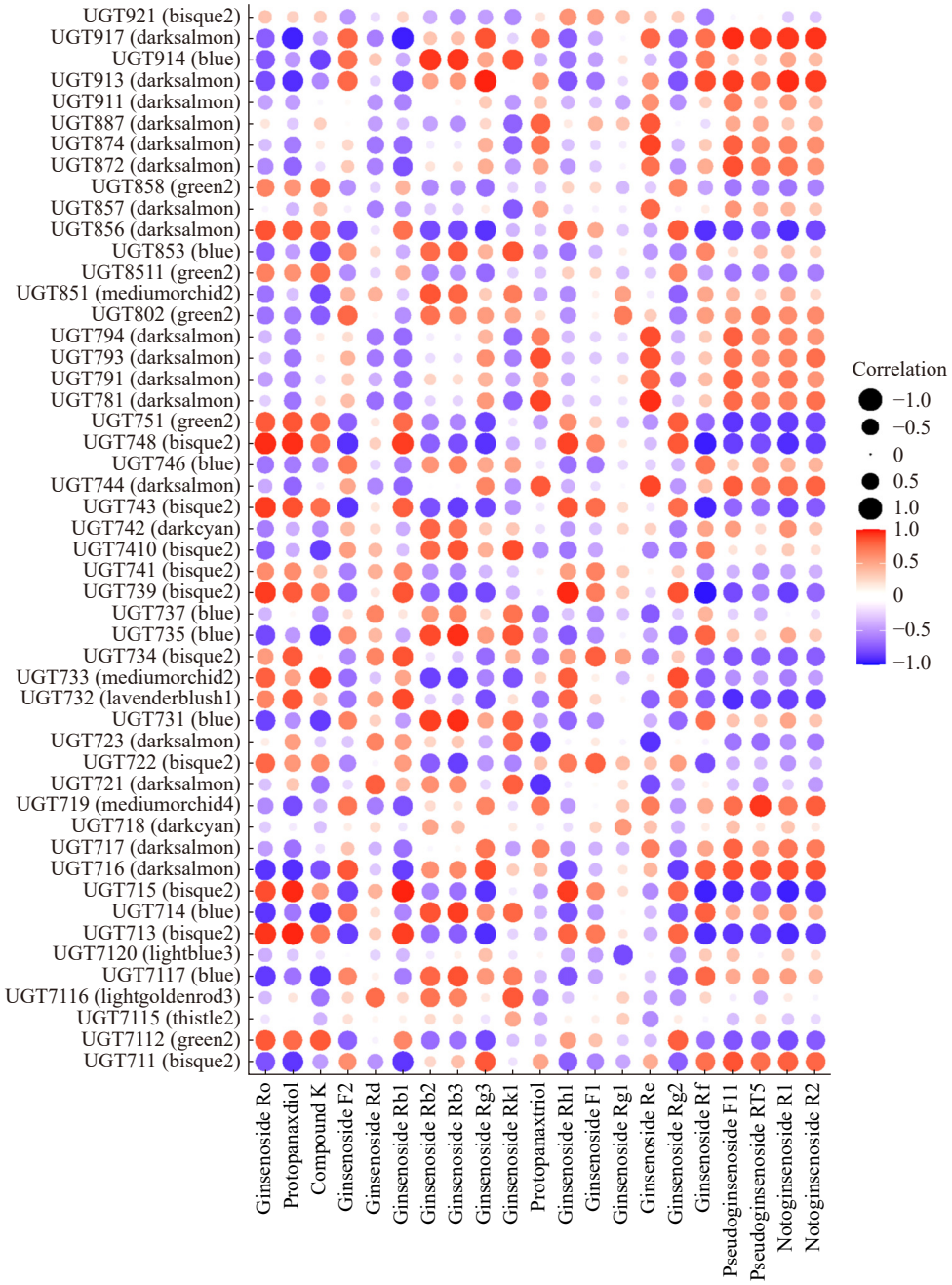

Fig. 7 Pearson correlation bubble chart of annotation UGTs and ginsenoside contents in *P. quinquefolium*. The list in brackets indicates the module name of each UGT

highly expressed in the above-ground parts. These results suggested that some ginsenosides are synthesized in leaves, then transported into roots for storage or physiological function during foliation. Previous research on ginsenoside biosynthesis in *P. ginseng* also supported this hypothesis [45]. These findings provide insights into the expression patterns of the ginsenoside biosynthetic genes in *P. quinquefolium* and will promote the research on other medicinal plants of *Panax* species.

Glycosylation is the last reaction essential for the formation of various ginsenosides. One or more monomers are transferred to the triterpene aglycone UDP-glycosyltransferase to synthesize various saponins [46, 47]. UGTPg1 (PgUGT71A27) can transfer the UDP-glucose to the C-20 hydroxyl group of dammarenediol-II to generate 20*S*-*O*-*β*-(D-glucosyl)-dammarenediol-II in engineered yeast [48]. Besides, PgUGT74AE2 catalyzes the steps of C-3 hydroxyl glycosylation to form ginsenosides Rh2 and F2 from protopanaxadiol and compound K, respectively [49]. However, there is little research on UGTs in *P. quinquefolium* [50-52]. Thus, the function of UGTs in ginsenoside modification and biosynthesis in *P. quinquefolium* are still unknown.

WGCNA is used to identify gene modules related to specific phenotypic traits expressed in certain samples, so as to identify vital central genes involved in particular processes [53]. Herein, 48 000 genes with FPKM > 3 were aggregated into 31 modules based on their expression patterns. The grouping results were used to examine the relationship among modules, ginsenoside content and tissue specificity. Several modules, such as mediumblue, darkorchid2, and lightgoldenrod3 ($R > 0.9$, $P < 0.05$), were closely related to ginsenoside biosynthesis. These modules can help discover candidate genes associated with ginsenoside accumulation, especially for those involved in the ginsenoside biosynthetic pathway. However, further study is needed to explore the biological effects of these genes. The WGCNA results were screened, and the filtered UGTs were correlated to various ginsenosides. For instance, UGT917, UGT913, UGT716, and UGT711 were positively correlated to pseudoginsenosides F11 and RT5, and notoginsenosides R1 and R2. However, further functional study of these UGTs can provide insights for analyzing unknown catalytic steps. This study provides a basis for ginsenoside biosynthesis in medicinal plants for future studies.

References

[1] Liu Y, Wang X, Wang L, *et al*. A nucleotide signature for the identification of american ginseng and its products [J]. *Front Plant Sci*, 2016, **7**: 319.

[2] Fernandez-Moriano C, Gonzalez-Burgos E, Iglesias I, *et al*. Evaluation of the adaptogenic potential exerted by ginsenosides Rb1 and Rg1 against oxidative stress-mediated neurotoxicity in an *in vitro* neuronal model [J]. *PLoS One*, 2017, **12**(8): e0182933.

[3] Szczuka D, Nowak A, Zaklos-Szyda M, *et al*. American ginseng (*Panax quinquefolium* L.) as a source of bioactive phytochemicals with pro-health properties [J]. *Nutrients*, 2019, **11**(5): 1041.

[4] Sen S, Chen S, Wu Y, *et al*. Preventive effects of North American ginseng (*Panax quinquefolius*) on diabetic retinopathy and cardiomyopathy [J]. *Phytother Res*, 2013, **27**(2): 290-298.

[5] Liu W, Zheng Y, Han L, *et al*. Saponins (ginsenosides) from stems and leaves of *Panax quinquefolium* prevented high-fat diet-induced obesity in mice [J]. *Phytomedicine*, 2008, **15**(12): 1140-1145.

[6] Chen Z, Lu T, Yue X, *et al*. Neuroprotective effect of ginsenoside Rb1 on glutamate-induced neurotoxicity: with emphasis on autophagy [J]. *Neurosci Lett*, 2010, **482**(3): 264-268.

[7] Niu XN, Luo W, Lv CN, *et al*. Research progress on naturally-occurring and semi-synthetic ocotillol-type ginsenosides in the genus *Panax L.* (Araliaceae) [J]. *Chin J Nat Med*, 2021, **19**(9): 648-655.

[8] Ratan ZA, Haidere MF, Hong YH, *et al*. Pharmacological potential of ginseng and its major component ginsenosides [J]. *J Ginseng Res*, 2021, **45**(2): 199-210.

[9] Yuan CS, Wang CZ, Wicks SM, *et al*. Chemical and pharmacological studies of saponins with a focus on American ginseng [J]. *J Ginseng Res*, 2010, **34**(3): 160-167.

[10] Wang Y, Choi HK, Brinckmann JA, *et al*. Chemical analysis of *Panax quinquefolius* (North American ginseng): A review [J]. *J Chromatogr A*, 2015, **1426**: 1-15.

[11] Lin H, Zhu H, Tan J, *et al*. Non-targeted metabolomic analysis of methanolic extracts of wild-simulated and field-grown American ginseng [J]. *Molecules*, 2019, **24**(6): 1053.

[12] Huang X, Liu Y, Zhang Y, *et al*. Multicomponent assessment and ginsenoside conversions of *Panax quinquefolium* L. roots before and after steaming by HPLC-MS^n [J]. *J Ginseng Res*, 2019, **43**(1): 27-37.

[13] Qi LW, Wang HY, Zhang H, *et al*. Diagnostic ion filtering to characterize ginseng saponins by rapid liquid chromatography with time-of-flight mass spectrometry [J]. *J Chromatogr A*, 2012, **1230**: 93-99.

[14] Wei G, Yang F, Wei F, *et al*. Metabolomes and transcriptomes revealed the saponin distribution in root tissues of *Panax quinquefolius* and *Panax notoginseng* [J]. *J Ginseng Res*, 2020, **44**(6): 757-769.

[15] Qu CL, Bai YP, Jin XQ, *et al*. Study on ginsenosides in different parts and ages of *Panax quinquefolius* L. [J]. *Food Chemistry*, 2009, **115**(1): 340-346.

[16] Jiang Z, Tu L, Yang W, *et al*. The chromosome-level reference genome assembly for *Panax notoginseng* and insights into ginsenoside biosynthesis [J]. *Plant Commun*, 2020, **2**(1): 100113.

[17] Kim NH, Jayakodi M, Lee SC, *et al*. Genome and evolution of the shade-requiring medicinal herb *Panax ginseng* [J]. *Plant Biotechnol J*, 2018, **16**(11): 1904-1917.

[18] Boopathi V, Subramaniyam S, Mathiyalagan R, *et al*. Till 2018: a survey of biomolecular sequences in genus *Panax* [J]. *J Ginseng Res*, 2020, **44**(1): 33-43.

[19] Kim YJ, Zhang D, Yang DC. Biosynthesis and biotechnological production of ginsenosides [J]. *Biotechnol Adv*, 2015, **33** (6 Pt 1): 717-735.

[20] Sun C, Li Y, Wu Q, *et al*. De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis [J]. *BMC Genomics*, 2010, **11**: 262.

[21] Wu D, Austin RS, Zhou S, *et al*. The root transcriptome for North American ginseng assembled and profiled across seasonal development [J]. *BMC Genomics*, 2013, **14**: 564.

[22] Qi J, Sun P, Liao D, *et al*. Transcriptomic analysis of American ginseng seeds during the dormancy release process by RNA-Seq [J]. *PLoS One*, 2015, **10**(3): e0118558.

[23] Wang J, Li J, Li J, *et al*. Transcriptome profiling shows gene regulation patterns in ginsenoside pathway in response to methyl jasmonate in *Panax quinquefolium* adventitious root [J]. *Sci Rep*, 2016, **6**: 37263.

[24] Huddleston J, Ranade S, Malig M, *et al*. Reconstructing complex regions of genomes using long-read sequencing technology [J]. *Genome Res*, 2014, **24**(4): 688-696.

[25] Xu Z, Peters RJ, Weirather J, *et al*. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis [J]. *Plant J*, 2015, **82**(6): 951-961.

[26] Zhang D, Li W, Chen ZJ, *et al*. SMRT- and illumina-based RNA-seq analyses unveil the ginsinoside biosynthesis and transcriptomic complexity in *Panax notoginseng* [J]. *Sci Rep*, 2020, **10**(1): 15310.

[27] Jo IH, Lee J, Hong CE, *et al*. Isoform sequencing provides a more comprehensive view of the *Panax ginseng* transcriptome [J]. *Genes (Basel)*, 2017, **8**(9): 228.

[28] Li Y, Dai C, Hu C, *et al*. Global identification of alternative splicing *via* comparative analysis of SMRT- and illumina-based RNA-seq in strawberry [J]. *Plant J*, 2017, **90**(1): 164-176.

[29] Chen W, Gong L, Guo Z, *et al*. A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics [J]. *Mol Plant*, 2013, **6**(6): 1769-1780.

[30] Altschul SF, Madden TL, Schaffer AA, *et al*. Gapped BLAST

and PSI-BLAST: a new generation of protein database search programs [J]. *Nucleic Acids Res*, 1997, **25**(17): 3389-3402.
[31] Wang L, Feng Z, Wang X, *et al*. DEGseq: an *R* package for identifying differentially expressed genes from RNA-seq data [J]. *Bioinformatics*, 2010, **26**(1): 136-138.
[32] Langfelder P, Horvath S. WGCNA: an *R* package for weighted correlation network analysis [J]. *BMC Bioinformatics*, 2008, **9**: 559.
[33] Li Y, Baldauf S, Lim EK, *et al*. Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana* [J]. *J Biol Chem*, 2001, **276**(6): 4338-4343.
[34] Li Y, Li P, Wang Y, *et al*. Genome-wide identification and phylogenetic analysis of Family-1 UDP glycosyltransferases in maize (Zea mays) [J]. *Planta*, 2014, **239**(6): 1265-1279.
[35] Zhang K, Wang X, Ding L, *et al*. Determination of seven major ginsenosides in different parts of *Panax quinquefolius* L. (American ginseng) with different ages [J]. *Chem Res Chin U*, 2008, **24**(6): 707-711.
[36] Zhang X, Ma X, Si B, *et al*. Simultaneous determination of five active hydrolysis ingredients from *Panax quinquefolium* L. by HPLC-ELSD [J]. *Biomed Chromatogr*, 2011, **25**(6): 646-651.
[37] Xia YG, Song Y, Liang J, *et al*. Quality analysis of American ginseng cultivated in Heilongjiang using UPLC-ESI(−)-MRM-MS with chemometric methods [J]. *Molecules*, 2018, **23**(9): 2396.
[38] Xu R, Zhang J, You J, *et al*. Full-length transcriptome sequencing and modular organization analysis of oleanolic acid- and dammarane-type saponin related gene expression patterns in *Panax japonicus* [J]. *Genomics*, 2020, **112**(6): 4137-4147.
[39] Sun MY, Li JY, Li D, *et al*. Full-length transcriptome sequencing and modular organization analysis of the naringin/neoeriocitrin-related gene expression pattern in *Drynaria roosii* [J]. *Plant Cell Physiol*, 2018, **59**(7): 1398-1414.
[40] Kim DH. Chemical diversity of *Panax ginseng*, *Panax quinquifolium*, and *Panax notoginseng* [J]. *J Ginseng Res*, 2012, **36**(1): 1-15.
[41] Xu J, Chu Y, Liao B, *et al*. *Panax ginseng* genome examination for ginsenoside biosynthesis [J]. *Gigascience*, 2017, **6**(11): 1-15.
[42] Kim YK, Yoo DS, Xu H, *et al*. Ginsenoside content of berries and roots of three typical Korean ginseng (*Panax ginseng*) cultivars [J]. *Nat Prod Commun*, 2009, **4**(7): 903-906.
[43] Searels JM, Keen KD, Horton JL, *et al*. Comparing ginsenoside production in leaves and roots of wild American ginseng (*Panax quinquefolius*) [J]. *Am J Plant Sci*, 2013, **4**(6): 1252-1259.
[44] Xue L, He Z, Bi X, *et al*. Transcriptomic profiling reveals MEP pathway contributing to ginsenoside biosynthesis in *Panax ginseng* [J]. *BMC Genomics*, 2019, **20**(1): 383.
[45] Kim YJ, Jeon JN, Jang MG, *et al*. Ginsenoside profiles and related gene expression during foliation in *Panax ginseng* Meyer [J]. *J Ginseng Res*, 2014, **38**(1): 66-72.
[46] Rahimi S, Kim J, Mijakovic I, *et al*. Triterpenoid-biosynthetic UDP-glycosyltransferases from plants [J]. *Biotechnol Adv*, 2019, **37**(7): 107394.
[47] De-Bruyn F, Maertens J, Beauprez J, *et al*. Biotechnological advances in UDP-sugar based glycosylation of small molecules [J]. *Biotechnol Adv*, 2015, **33**(2): 288-302.
[48] Yan X, Fan Y, Wei W, *et al*. Production of bioactive ginsenoside compound K in metabolically engineered yeast [J]. *Cell Res*, 2014, **24**(6): 770-773.
[49] Jung SC, Kim W, Park SC, *et al*. Two ginseng UDP-glycosyltransferases synthesize ginsenoside Rg3 and Rd [J]. *Plant Cell Physiol*, 2014, **55**(12): 2177-2188.
[50] Lu C, Zhao SJ, Feng PC, *et al*. Functional analysis of the promoter of a UDP-glycosyltransferase gene from *Panax quinquefolius* [J]. *Plant Cell Tiss Org*, 2018, **135**(3): 381-393.
[51] Lu C, Zhao S, Wei G, *et al*. Functional regulation of ginsenoside biosynthesis by RNA interferences of a UDP-glycosyltransferase gene in *Panax ginseng* and *Panax quinquefolius* [J]. *Plant Physiol Biochem*, 2017, **111**: 67-76.
[52] Lu C, Zhao SJ, Wang XS. Functional regulation of a UDP-glucosyltransferase gene (Pq3-*O*-UGT1) by RNA interference and overexpression in *Panax quinquefolius* [J]. *Plant Cell Tiss Org*, 2017, **129**(3): 445-456.
[53] Hollender CA, Kang C, Darwish O, *et al*. Floral transcriptomes in woodland strawberry uncover developing receptacle and anther gene networks [J]. *Plant Physiol*, 2014, **165**(3): 1062-1075.

Cite this article as: DI Peng, YAN Yan, WANG Ping, YAN Min, WANG Ying-Ping, HUANG Lu-Qi. Integrative SMRT sequencing and ginsenoside profiling analysis provide insights into the biosynthesis of ginsenoside in *Panax quinquefolium* [J]. *Chin J Nat Med*, 2022, **20**(8): 614-626.