scientific reports

Check for updates

OPEN

Comparative transcriptomic analysis provides insights into the regulation of root-specific saponin production in *Panax japonicus*

Xiangyin Luo[1], Zhengxiu Ye[1], Xianming Shi[1], Ziyao Hu[2], Junhao Shen[1], Lei You[1], Ping Huang[1], Guodong Wang[2], Lanlan Zheng[1], Chen Li[1] & Yonghong Zhang[1]✉

Panax japonicus*, a traditional medicinal plant from the Araliaceae family, produces bioactive triterpenes with health benefits. In Traditional Chinese Medicine, its roots have been used, but the chemical basis of its medicinal use is unclear, particularly regarding the metabolism and regulation of triterpene saponin biosynthesis. This study employed an integrative approach using Ultra Performance Liquid Chromatography (UPLC) and transcriptome analysis. Our UPLC analysis showed that ginsenoside Ro and chikusetsusaponin IVa were mainly detected in *P. japonicus* root. Subsequently, a comparative transcriptome analysis of four *P. japonicus* tissues (roots, leaves, flowers and fruits) was conducted using Illumina sequencing. As a result, 90,985 unigenes were functionally annotated from a total of 211,650 assembled non-redundant transcripts. Among these, 42,829 unigenes were annotated in NR database. Tissue-specific gene analysis revealed that roots had the highest number of specifically expressed unigenes (11,832). The majority of these unigenes were associated with metabolic processes. Additionally, tissue expression patterns analysis for three common transcription factor families indicated that WRKY family genes showed a significantly root-specific expression pattern, potentially playing a role in triterpene saponin biosynthesis. Notably, we investigated the expression profiles of genes related to the biosynthesis of triterpene saponins and found that four genes, *ACCT*, *HMGS*, *HMGR* and *SS*, encoding key enzymes in triterpene saponins biosynthesis pathway, were primarily expressed in the root. Overall, our study provides a set of *P. japonicus* tissue transcriptome data, which will aid in the discovery of triterpene saponin biosynthetic genes and offers valuable genetic information for this medical plant.

Keywords *Panax japonicus*, RNA-seq, Comparative transcriptome, Triterpene saponins biosynthesis, Transcriptional factor

Abbreviations

AP2/EREBP	APETALA2/ethylene-responsive element binding protein
bHLH	Basic helix-loop-helix
DEGs	Differential expressed genes
FPKM	Fragments per kilobase of transcript per million fragments mapped
GO	Gene Ontology
UPLC	Ultra Performance Liquid Chromatography
IPP	Isopentenyl diphosphate
MVA	Mevalonic acid
MEP	Methylerythritol phosphate
OSGs	Organ-specific genes
TFs	Transcription factors
TCM	Traditional Chinese Medicine

[1]Laboratory of Medicinal Plant, Institute of Basic Medical Sciences, School of Basic Medicine, Biomedical Research Institute, Hubei Key Laboratory of Wudang Local Chinese Medicine Research, Hubei University of Medicine, Shiyan 442000, China. [2]College of Life Sciences, Engineering Research Center of High Value Utilization of Western China Fruit Resources, Ministry of Education, Shaanxi Normal University, Xi'an 710119, Shaanxi, China. ✉email: zhangyh@hbmu.edu.cn

TS Triterpene saponin

Panax japonicus C.A. Mey, also called as "Zhu Zi Shen", is a plant belonging to the Panax genus in the Araliaceae family. Its rhizome, characterized with bead-like appearance is used as a traditional Chinese medicinal herb, which exhibits a wide range of pharmacological effects, including anti-fatigue, immune-regulation, anti-oxidative stress, anti-inflammatory and anti-cancer properties[1–3]. Recent pharmacological research has substantiated that the principal bioactive compounds discovered in *P. japonicus* are ginsenosides and chikusetsusaponins, which both belong to triterpene saponin (TS)[4].

Triterpene saponins are typically biosynthesized in cytoplasm through mevalonic acid (MVA) pathway or in plastid via methylerythritol phosphate (MEP) pathway[5,6]. Both pathways produce the necessary isoprene units-isopentenyl diphosphate (IPP). After the formation of isopentenyl diphosphate (IPP), short-chain prenyltransferases, involving GPPS, GGPPS and FPS, catalyze the sequential addition of isoprene units, extending the carbon chain by five carbons with each addition to form squalene. Next, various triterpene skeletons are formed through the cyclization of 2,3-oxidosqualene by diverse oxidosqualene cyclases, including dammarane-type and oleanane-type terpenoids[7]. Subsequent catalytic steps include hydroxylation via CYP450s and glycosylation via UGTs respectively, result in the formation of diverse saponins[8–10].

With the fast development of "omics" platform, molecular mechanism exploration underlying saponin biosynthesis using transcriptomic sequencing in *P. japonicus* has gained attentions in recent years. However, the overall transcriptomic study regarding different tissues in *Panax japonicus* is still incomplete. Our UPLC analysis suggested that ginsenoside Ro and chikusetsusaponin IVa were primarily accumulated in *P. japonicus* root. Thus, a comparative transcriptome study of four *P. japonicus* tissues (root, leaf, fruit and flower) were conducted to identify putative genes involved in triterpene saponins biosynthesis in *P. japonicus*. Organ specific expressed gene analysis revealed that the root possessed largest number of OSGs and was enriched in various metabolic pathways. Notably, analysis of transcriptional factor families revealed that a large number of WRKY transcriptional factors exhibited significantly enriched expression in *P. japonicus* root. Our study provides new insights for future functional genomics research focusing on triterpene saponins biosynthesis and regulation of *P. japonicus*.

Materials and methods

Plant material

The plant variety of *P. japonicus* used in this study was *Panax japonicus* C. A. Mey. var. major (Burk.) C. Y. Wu et K. M. Feng (Zhu zi shen in Chinese). The plants were cultivated in a designated plantation field for physiological experiments at Longwangzhai Rare Chinese Herbal Medicine Cultivation Farmers' Cooperative in Shiyan, Hubei province, China. Fresh samples of the four organs (root, leaf, flower and fruit) were collected from three-year-old plants. Samples from the same organ of three individual plants were combined to form one biological replicate, with each replicate created by randomly selecting and mixing three individuals to minimize genetic background variability. This process was repeated twice, resulting in three biological replicates, all of which were submitted for RNA sequencing. All samples were thoroughly rinsed and cleansed twice using distilled water. Subsequently, they were immediately frozen in liquid nitrogen and stored at -80℃ for RNA extraction.

UPLC analysis of ginsenosides Ro and chikusetsusaponins IVa

The UPLC analysis procedure was performed as previously reported with minor modifications[11]. Briefly, fresh sample was harvested, collected and incubated at 105 °C for no less than 15 min in an oven, followed by 60 °C until the samples were completely dried. Dry samples were ground into powder and about 0.1 g dry powder was extracted with 10 mL 80% methanol under ultrasonic water bath at 40℃ for 40 min. Then the extract was filtered through a syringe filter (0.22 μm). UPLC was conducted using Waters ACQUITY UPLC system with a PDA eλ detector (Waters Corporation, Milford, MA, US) together with a Waters BEH C18 column (1.7 μm, 2.1 mm × 100 mm) using a gradient elution program with flow velocity of 0.2 mL/min (A, acetonitrile. B, 0.1% phosphoric acid water solution): 0 min-2 min, acetonitrile concentration increased from 40 to 50%; 2 min-6 min, acetonitrile concentration increased from 50 to 55%; 6 min-6.1 min, acetonitrile concentration increased from 55 to 100%; 6.1 min-9 min, acetonitrile concentration kept at 100%; 9 min-9.1 min, acetonitrile concentration decreased from 100 to 40%; 9.1 min-11 min, acetonitrile concentration kept at 40%. The detection wavelength was 203 nm.

RNA Isolation and RNA-Seq

Total RNA extraction and purification were performed using the RNAsimple total RNA kit (TIANGEN, China). RNA purity and integrity were determined using a NanoPhotometer spectrophotometer (IMPLEN) and Qubit RNA Assay Kit in Qubit 2.0 Fluorometer (Life Technologies). RNA integrity number exceeded 7.4 was considered as high-quality for transcriptome sequencing. In total, twelve sequencing libraries were generated using an Illumina HiSeq Library Preparation Kit and subsequently sequenced on the Illumina HiSeq 2500 platform by Wuhan Frasergen informative Co., Ltd (http://www.frasergen.com). Trimmomatic software was used to eliminate low-quality reads and obtain clean data from the raw reads[12].

Transcriptomic assembly, BUSCO assessment and functional annotation

De novo assembly was conducted using Trinity assembler version 2.6.6 to merge the clean reads into contigs[13]. BUSCO's completeness analysis was used to check the assembly correctness. The longest transcripts were assigned as unigenes to avoid different isoforms being miscalculated as multiple copies. Then BUSCO (v5.2.2) was used for conservative gene detection with embryophyta_odb10 as the lineage dataset[14]. The alignment against Swiss-Port and Pfam databases was employed to predict open reading frame (ORF). TransDecoder

was utilized to identify the candidate coding regions and generate their corresponding amino acid sequences. Functional annotation was performed using Blast2GO tool ($E \leq 10^{-2}$) against various databases, including NCBI non-redundant protein sequences (Nr) database, Gene Ontology (GO) database, Kyoto Encyclopedia of Genes and Genomes (KEGG) database[15–17], euKaryotic Ortholog Groups/Clusters of Orthologous Groups (KOG/COG) database and Swiss-Prot database[18].

Differential gene expression analysis

The estimation of gene expression levels was firstly performed using Expectation–Maximization (RSEM) software[19]. The expression levels of transcripts were then quantified as FPKM using SARTools[20]. The P-values were adjusted employing the Benjamini–Hochberg method to control the False Discovery Rate (FDR). Finally, differential expression genes (DEGs) were identified using a threshold of $\log_2$FoldChange ≥ 1 and FDR < 0.05.

Triterpenoid saponin biosynthesis related genes identification

In order to obtain the amino acid sequences of putative enzymes involved in triterpenoid saponin biosynthesis, the KEGG database was accessed[15–17]. Representative sequences were used as queries to retrieve the corresponding transcripts in *P. japonicus* through BLASTP analysis.

Transcription factor analysis

Candidate transcription factors (TFs) genes were screened and identified using PlantTFcat and iTAK analysis tool[21,22]. The results were then verified in plant transcription factor database (PlantTFdb)[23].

Result

UPLC analysis of ginsenosides ro and chikusetsusaponins IVa in different *Panax* japonicus organs

Fresh samples of four organs (root, leaf, flower and fruit) were collected from three-year-old *P. japonicus* plants (Fig. 1a-e). UPLC was used to analyze the content of two major saponins, Ginsenosides Ro and Chikusetsusaponins IVa, in different organs. The results showed that Ginsenosides Ro and Chikusetsusaponins IVa were primarily observed in the root tissue (Fig. 1f), aligning with the root being the major medicinal part of *P. japonicus*.

RNA-seq and de novo transcripts assembly

To explore potential genes involved in saponin biosynthesis, four tissues (root, leaf, flower, and fruit) transcriptomes of *P. japonicus* were generated by Illumina paired-end sequencing with three biological replicates. Clean reads pairs were obtained through filtering sequencing adaptor and removing low-quality raw data and the clean reads number and quality of all 12 samples were listed in Table S1. Then transcript assembly was performed using Trinity software. Finally, 211650 non-redundant transcripts were obtained and the length distribution was shown in Fig. 2a. The mean length and N50 of assembled transcripts were 1185 bp and 1895 bp respectively (Table S2). Additionally, BUSCOs completeness analysis was used to check the assembly correctness. The results showed that the BUSCOs completeness was 94.8% (including 93.7% single-copy BUSCOs and 1.1% duplicated BUSCOs), suggesting a high quality of transcripts assembly (Fig. 2b).

Functional unigene annotation

To get a representative transcript regardless of splicing variants for one single gene, the longest transcript was designated as the unigene for each gene. As a result, 90,985 unigenes were obtained from 211,650 transcripts. Five databases were used to annotate unigenes, namely NR, GO, KOG, KEGG and Swiss-prot. As a result, there were 42,829 (47.07%), 36,629 (40.26%), 22,968 (25.24%), 13,493 (14.83%) and 34,925 (38.39%) unigenes annotated in NR, GO, KOG, KEGG and Swiss-prot databases, respectively (Fig. 3a and Table S3). Totally, there were 44,023 (48.38%) unigenes annotated in all databases. In order to gain a comprehensive understanding of homologous genes in *P. japonicus* across various plant species, the annotated unigenes were queried in NCBI NR database. As shown, the largest percentage of unigenes were homologous to those of *Daucus carota* (11.2%), followed by *Quercus suber* (4.9%), *Nyssa sinensis* (3.9%) and *Camellia sinensis* (2.2%), with *Panax ginseng* showed 0.4% sequence homology with *P. japonicus* (Fig. 3b; Table S4).

Then Gene Ontology (GO) classification was performed to uncover the functional roles of unigenes. In the category of biological processes, the mostly enriched GO terms were cellular process and metabolic process. The most enriched GO terms in the molecular function category were binding and catalytic activity. In the cellular components category, the majority of unigenes showed a significant enrichment in cellular anatomical entities (Fig. 3c; Table S5). Additionally, 15,257 (16.77%) unigenes were classified to 25 function classes in the KOG database (Fig. 3d; Table S6). The most percent of unigenes were enriched in the class of "Translation, ribosomal structure and biogenesis" (2570), followed by "Posttranslational modification, protein turnover, chaperones" (2015), "General function prediction only" (1439) and "Energy production and conversion" (1238).

Organ-specific genes identification and differential expression genes analysis

To determine the gene expression levels, the clean reads from four organs were initially mapped using RSEM and bowtie 2. Subsequently, unigenes expression levels were quantified using fragments per kilobase of transcript per million fragments mapped (FPKM) values. As a result, 83,038 unigenes were detected to be expressed in four organs in total (FPKM > 0.01) (Table S7). The expression level distributions are shown in Fig. 4a, showing that each organ displays a stable distribution pattern. The correlation analysis between samples showed that flower is correlated with fruit to the highest extent (0.89) and leaf is correlated with flower to the lowest extent (0.1) (Fig. 4b). Differential expressed genes analysis between every two organs showed that the largest number

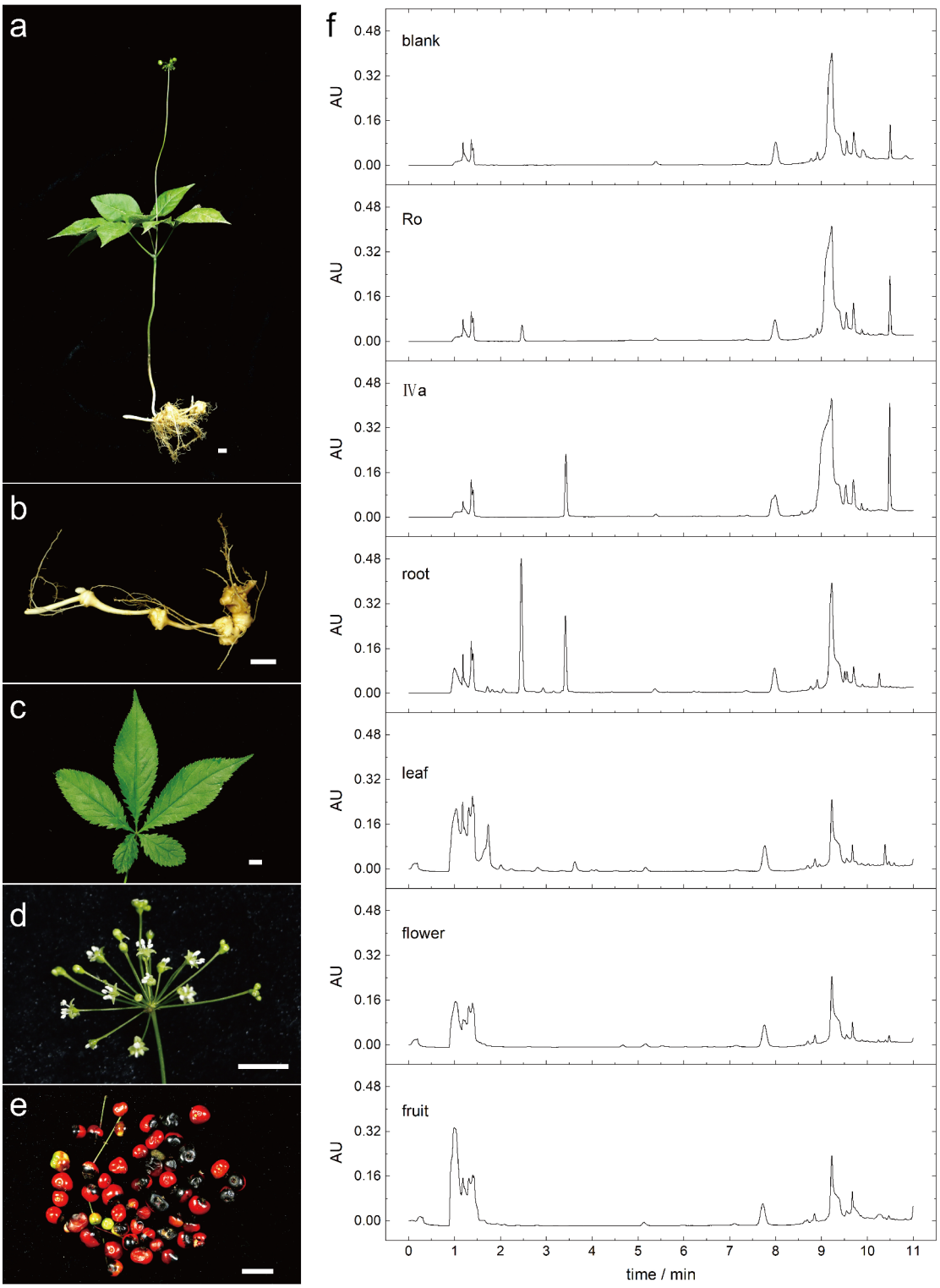

Fig. 1. Photographs of *Panax japonicus* plants and UPLC analysis. (**a**) Whole plant of three-year-old *Panax japonicus* plants. (**b**–**e**) Root, leaf, flower and fruit of *Panax japonicus*. bar = 1 cm. (**f**) UPLC-PDA chromatograms of Ginsenoside Ro and Chikusetsusaponin IVa across four organs. The x-axis represents retention time, and absorbance is shown in AU (absorbance units). Blank refers to 80% methanol used as a control.

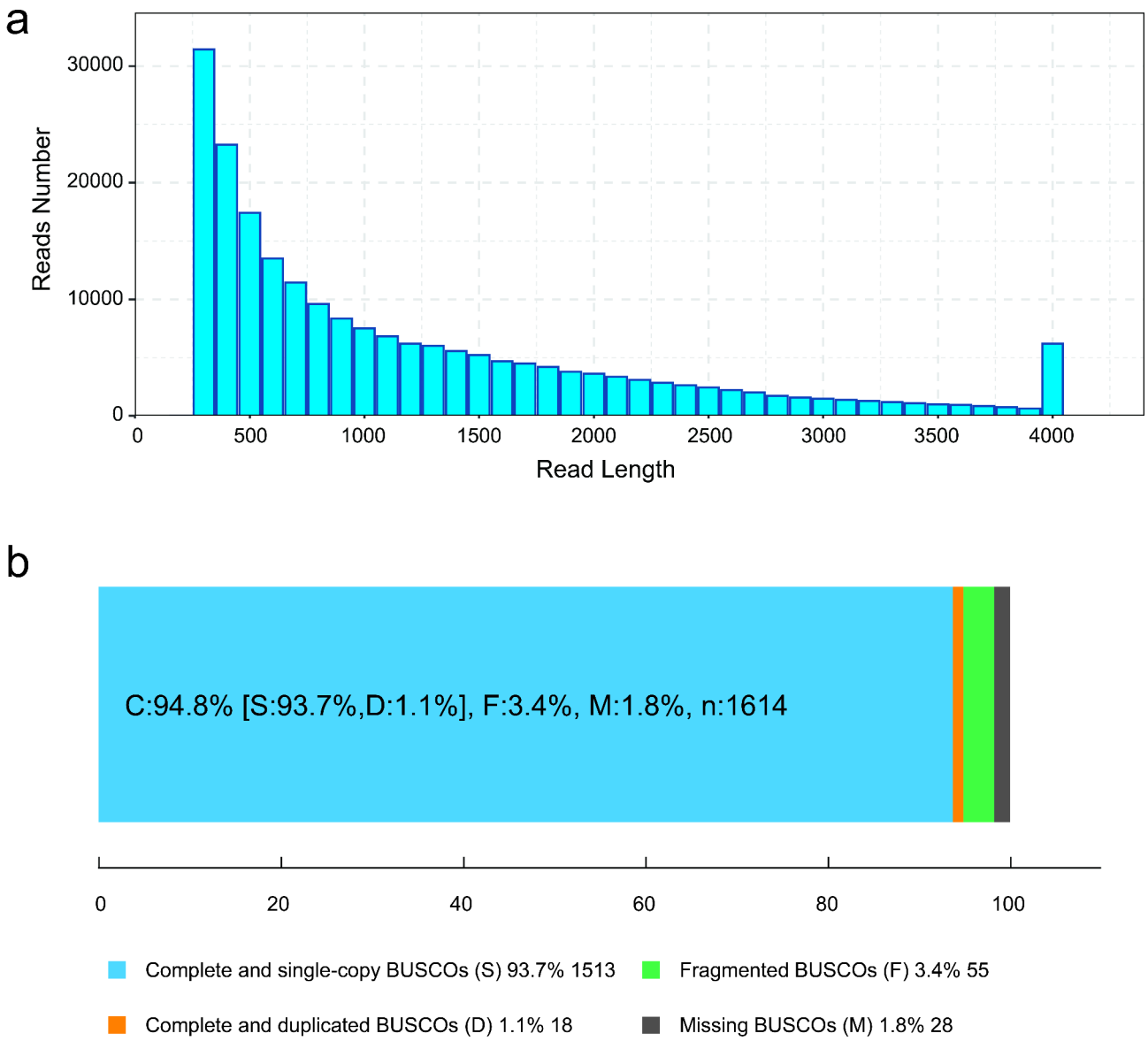

Fig. 2. Transcripts length distribution and BUSCO analysis. (**a**) The length distribution of assembled non-redundant transcripts. (**b**) BUSCO's assembly completeness analysis.

(13,743) of DEGs was in the leaf-vs-root comparison, among which 7,723 were up-regulated genes and 6,020 were down-regulated genes. Notably, root always displayed more up-regulated genes than down-regulated genes when compared with other organs (Fig. 4c). All expressed genes venn diagram in four organs were displayed and a total number of 24,450 genes that exhibited expression exclusively in a single organ were regarded as organ-specific genes (OSGs) in comparison with 40,674 common genes that were expressed ubiquitous in all four organs (Fig. 4d), among which the root possessed largest number of OSGs (11,832), followed by flower (7,165), leaf (3,144) and fruit (2,309). Hierarchical clustering analysis was conducted to reveal global gene expression patterns of four organs (Fig. 4e). As a result, the root exhibits considerably distinct expression patterns compared with other organs.

Root-specifically expressed genes identification and analysis

P. japonicus root plays a significant role as a primary source of crude drugs in Traditional Chinese Medicine (TCM) practices. Therefore, an investigation was conducted to examine genes specifically expressed in root, aiming to unravel the dynamic patterns of gene expression in this organ. As mentioned above, there were totally 11,832 unigenes expressed specifically in root. To explore the primary functional molecular roles of these organ-specific genes (OSGs) in root metabolite biosynthesis, Gene Ontology (GO) enrichment analysis was conducted. As a result, the largest number of root OSGs were enriched in binding, catalytic activity, structural molecule activity and transporter activity in the GO term of molecular function. Within the cellular component category, the most significant terms for root organ-specific genes (OSGs) were cellular anatomical entity and protein-

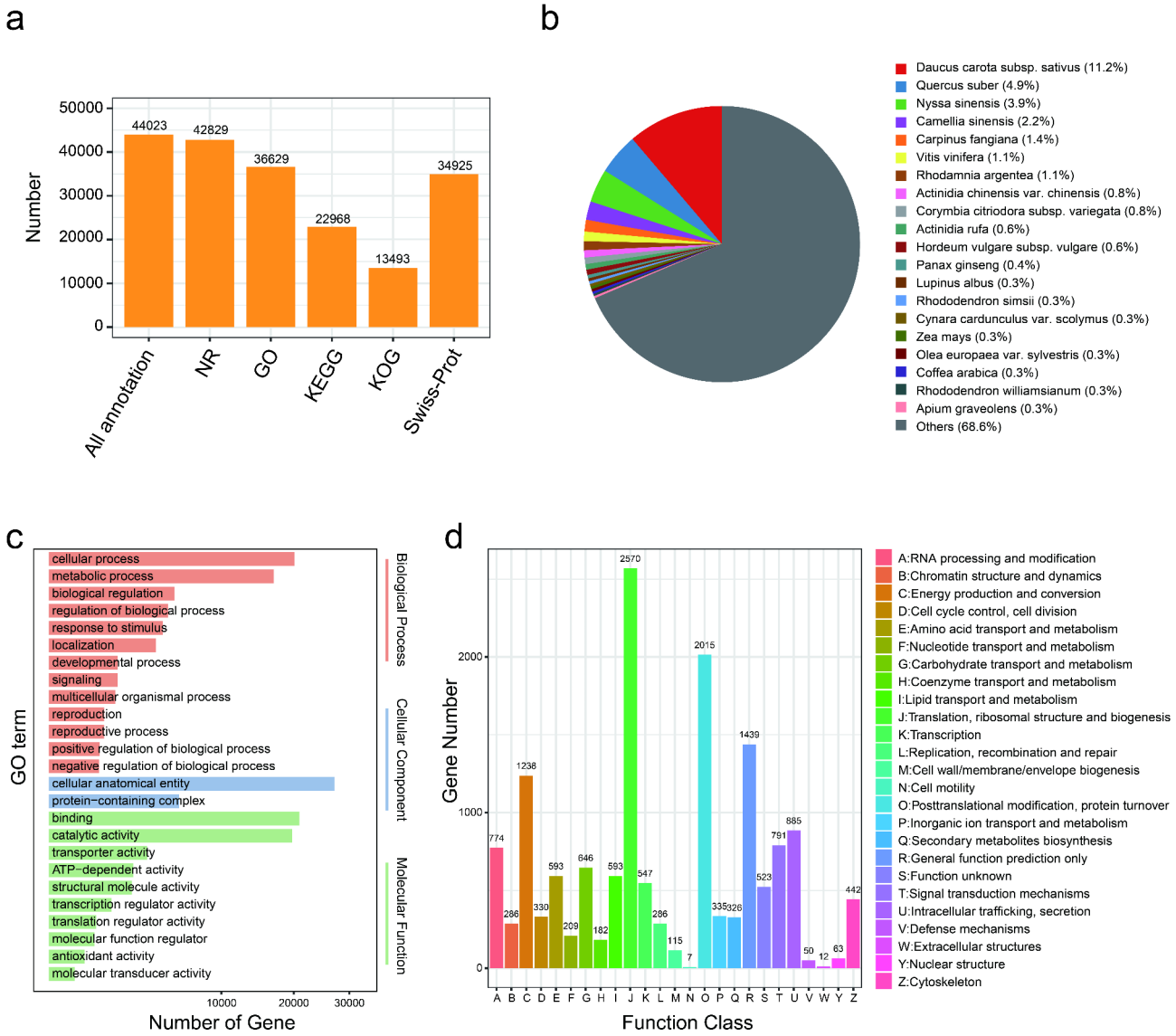

Fig. 3. Unigenes length distribution, functional annotation and GO/KOG analysis. (**a**) Unigene numbers annotated referring to various databases were displayed based on sequence BLAST analysis. (**b**) The pie chart showing the percentage of significant hit species based on unigenes sequence blast. (**c**) Gene Ontology Classification and (**d**) KOG analysis of all identified unigenes.

containing complex. Furthermore, a large number of root organ-specific genes (OSGs) was observed in cellular process, metabolic process, localization, response to stimulus and biological regulation in terms of biological process category (Fig. 5a; Table S8). This observation suggests that the root serves as the primary organ for biosynthesis of metabolites. KEGG analysis of root OSGs were also performed and the result showed that many metabolic pathways, such as "Alanine, aspartate and glutamate metabolism" and "Arginine biosynthesis" were prominently enriched (Fig. 5b; Table S9).

Identification of candidate transcription factors genes

Transcription factors (TFs) play a crucial role in regulating a wide range of developmental processes and metabolic pathways associated with natural product synthesis in plants[24,25]. In this study, 4,916 unigenes were identified as candidate TFs and classified to 58 TF families, among which MYB (723), basic helix-loop-helix (bHLH) (343), AP2-EREBP (315), C3H type zinc finger (249), NAC (234) and WRKY (201) families were the top six members (Fig. 6a; Table S10). Then, the expression patterns of three common TF families (MYB, bHLH and WRKY) were checked in four tissues (Table S10). Notably, WRKY members exhibited significant root enriched expression compared with MYB and bHLH TF families (Fig. 6b; Figure S1), revealing their potential functions in regulating metabolites biosynthesis in *P. japonicus* root. Moreover, 87 TFs encoding genes were identified among 11,832 root OSGs (Table S10), these would be root specific TFs in *P. japonicus*.

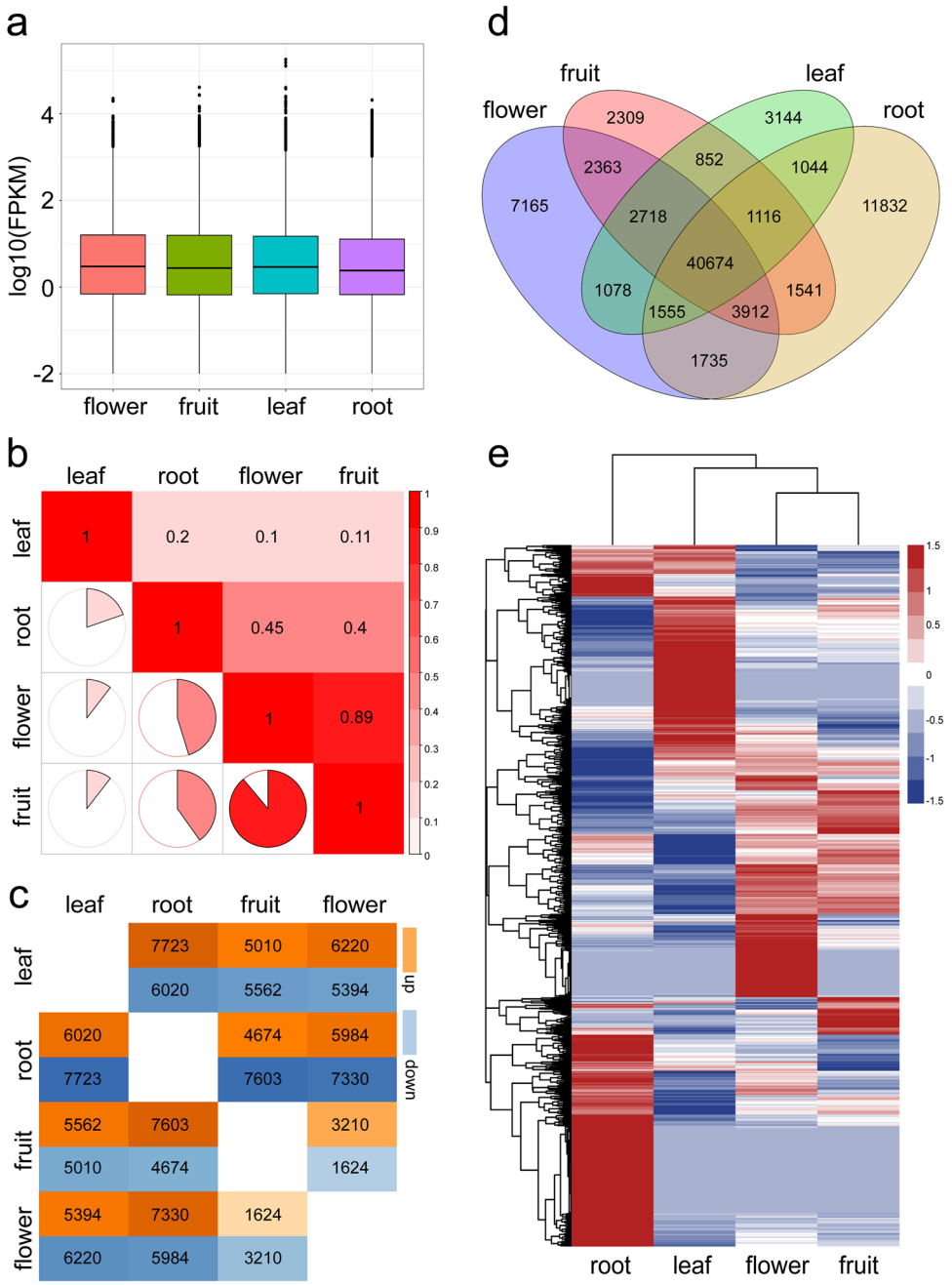

Fig. 4. The overview of expressed unigenes, OSGs, and DEGs. (**a**) The distributions of unigenes expression levels in four organs. (**b**) The square values of Pearson's correlation coefficients (R^2) between each set of organs. (**c**) The number of differential expression genes (up and down) between every two organs. (**d**) Venn diagram showing the number of uniquely or commonly expressed unigenes among organs. (**e**) All unigenes clustering analysis and expression heat map of all organs (root, leaf, flower and fruit). To analyze gene expression, all gene FPKM values were averaged and z-transformed, forming scaled expression values ranging from -1.5 to + 1.5.

Identification of transcripts potentially involved in saponin biosynthesis in *P. japonicus*

Triterpenoid saponin biosynthetic pathway has been reported in many plants[26–28], which mainly involves Mevalonate (MVA) pathway in cytoplasm and Methyl Erythritol Phosphate (MEP) pathway in plastid. Considering that Saponins accumulate in an organ- and/or tissue-specific manner[29], the mechanisms underlying the coordination between the MVA and MEP pathways, as well as the activities of relevant genes, are not yet fully understood. To get knowledge of the enzymes involved in saponin biosynthesis in *P. japonicus*, 35 MVA pathway related genes, 32 MEP pathway related genes and 34 saponin biosynthesis downstream genes were identified from *P. japonicus* transcriptome dataset (Fig. 7; Table S11). Many of these identified genes were preferentially expressed in the root, suggesting that they might play key roles in saponin biosynthesis of *P. japonicus* root. Notably, 13 MVA pathway related genes displayed root specific expression, while only 3 MEP pathway related

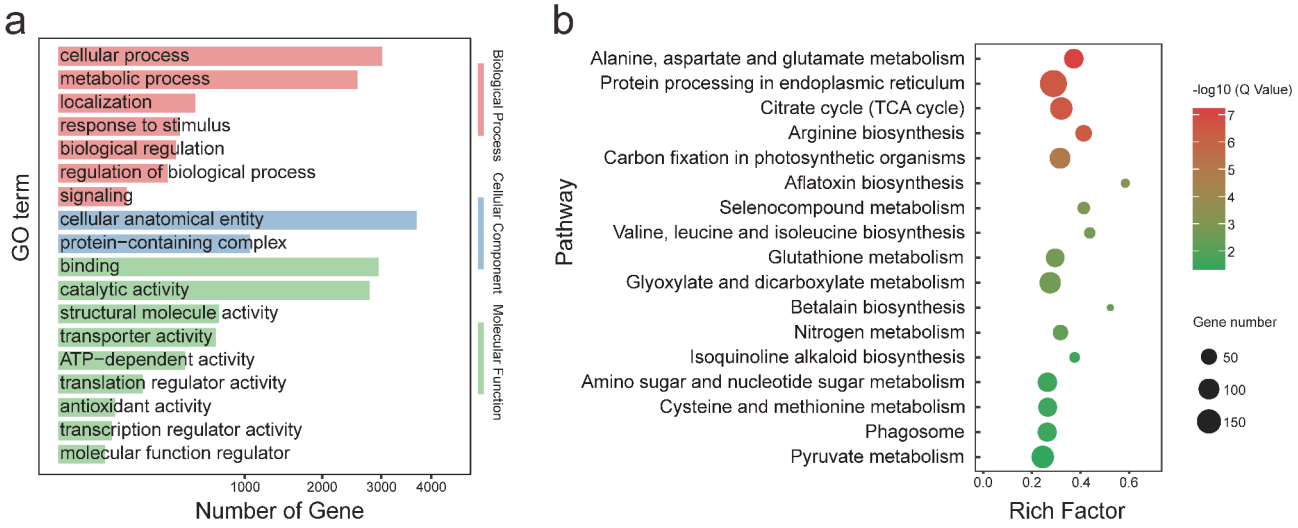

Fig. 5. Gene ontology classification (**a**) and KEGG enrichment analysis (**b**) of root-specifically expressed genes. (**a**) Three GO terms, molecular function (green color), cellular component (blue color) and biological process (red color) were indicated in the GO classification. (**b**) The bubble plot visually presents the statistics of KEGG enrichment. The x-axis displays the rich factor, which represents the ratio of differential expression genes to the total number of genes annotated to the specific entry. The size of each dot corresponds to the number of genes enriched in the pathway. The color of the dot indicates the p-value, reflecting the significance of the enrichment.

genes have root enriched expression (Fig. 7; Table S11), suggesting that MVA pathway was more active in root tissue. This is consistent with previous knowledge that MVA pathway played a more important role in saponin biosynthesis of Panax genus[30,31]. Together, these results provided new insights into saponin biosynthesis related gene mining in *P. japonicus*.

Discussion

Metabolic pathways exploration involved in natural product synthesis serves a dual purpose. Firstly, it deepens our understanding of how these valuable compounds are created. Secondly, it provides valuable resources for facilitating the large-scale production of natural products and fostering the generation of novel chemicals in synthetic biology area. This knowledge and application have the potential to revolutionize future advancements in synthetic biology. A key objective of this study is to gain molecular insights into the metabolic pathways responsible for the synthesis of saponins, which are the primary bioactive compounds found in *P. japonicus*.

Transcriptomic analysis is a convenient technology in identifying potential genes related to certain biological process and is widely used in medicinal plant research[32,33]. Rai et al. performed a de novo transcriptome assembly of *P. japonicus* using five tissues (flower, leaf, secondary root, young rhizome and old rhizome)[34]. The mean length and N50 of assembled contigs were 668 bp and 957 bp, respectively in their study. In our study, the mean length and N50 of assembled transcripts were 1185 bp and 1895 bp (Table S2), suggesting a significantly better quality of de novo assembly.

Since the MVA and MEP pathways are essential for saponin biosynthesis by providing key precursors, many transcriptome-based studies in less-explored plants focus on dissecting these pathways[34,35]. Therefore, identifying these two pathway genes and exploring their potential links to root-accumulating saponin regulation is a key objective of this study. Through comparative transcriptome analysis, a number of potential gene members encoding key enzymes involved in the MVA pathway and the MEP pathway in *P. japonicus* have been identified (Fig. 7). The expression patterns of these identified genes were analyzed across four different organs. A significant portion of these genes exhibited primary expression in the root, indicating that roots are a major site for saponins accumulation. It is noteworthy that thirteen MVA pathway genes exhibited preferential expression in roots, while only three MEP pathway genes were root-enriched (Fig. 7; Table S11). This observation demonstrated that MVA pathway plays a major role in root-accumulating saponin regulation, which is consistent with previous reports in other panax plants[31]. In particular, our comparative transcriptome analysis across different organs is effective for identifying organ-specific regulators in the regulation of root-accumulating saponins. Future genetic validation will be necessary. In addition to expression pattern-centered research, assessing the impact of abiotic stresses on gene activity may enhance our understanding of how these two pathways coordinate. However, it is important to note that not all these genes are exclusively expressed in the root, with some saponin biosynthetic genes expressing in aboveground tissues (Fig. 7). The inconsistency between gene expression level and its corresponding metabolite is a common phenomenon, which might be attributed to post-transcriptional regulations. Alternatively, it is possible that intermediate and derivative metabolites are synthesized in both the root and other parts of the plant, but their accumulation is primarily observed in the root due to transportation between tissues, which involves cross-talk in the coordination and regulation at the tissue level.

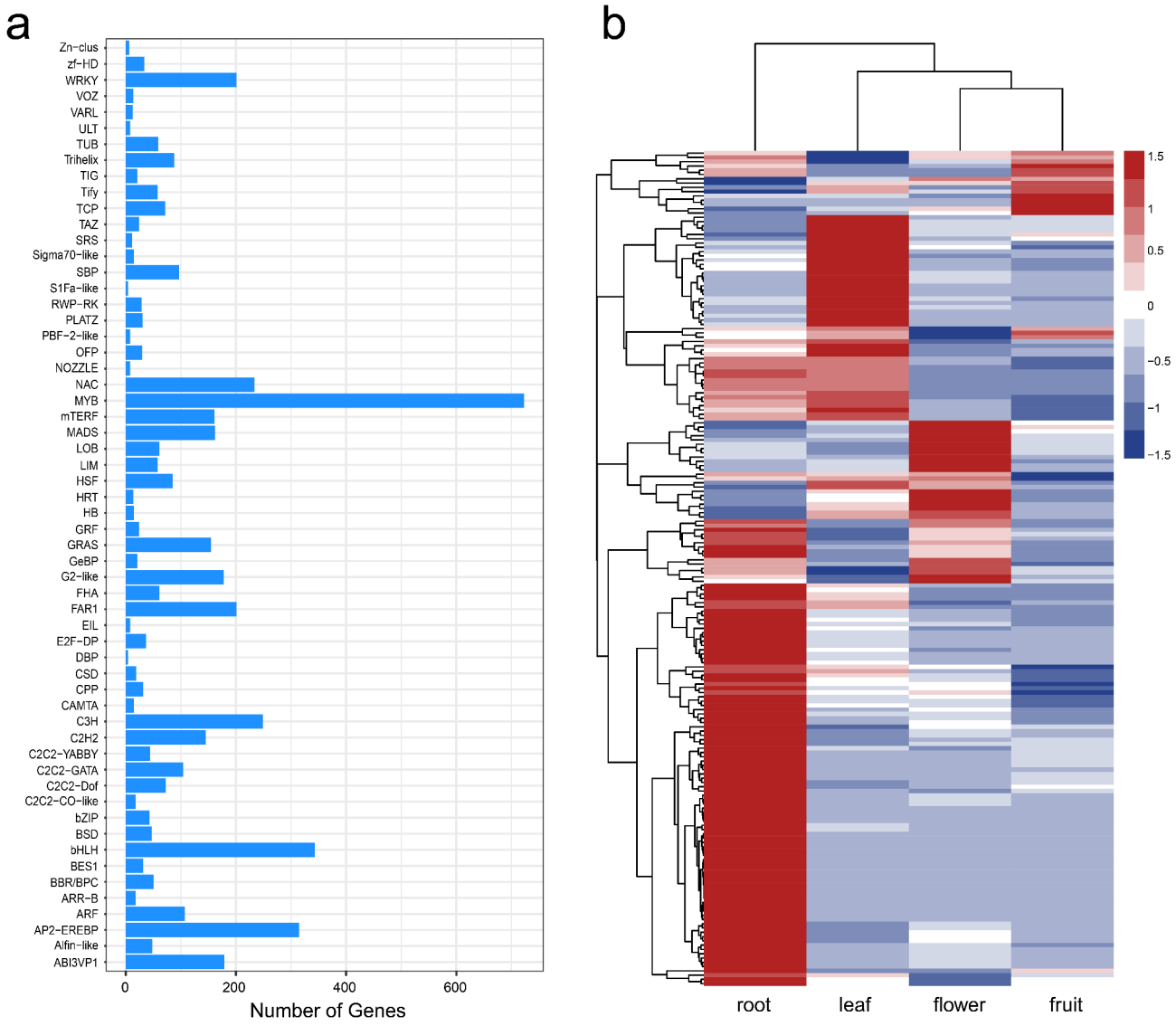

Fig. 6. Classification and clustering analysis of identified transcription factor families. (**a**) Statistics of genes numbers encoding different types of TFs. (**b**) Overview of expression and clustering analysis of WRKY TF family in four organs. Gene expression level was transformed as mentioned above.

In summary, this study has generated an integrative collection of high-quality transcriptomes for *P. japonicus* different tissues. By analyzing the gene expression profiles of distinct organs in *P. japonicus*, we gained valuable insights into organ-specific genes, revealing that each organ expresses a unique set of genes. We specifically focused on genes expressed exclusively in the root and our UPLC data (Fig. 1) support the notion that the root serves as the primary source of bioactive compounds. These findings suggest that root-specifically expressed genes play a crucial role in activating specialized metabolic pathways within the roots. Collectively, these results provide a foundation for uncovering the molecular mechanisms responsible for the tissue-specific accumulation of biologically active products in *P. japonicus*.

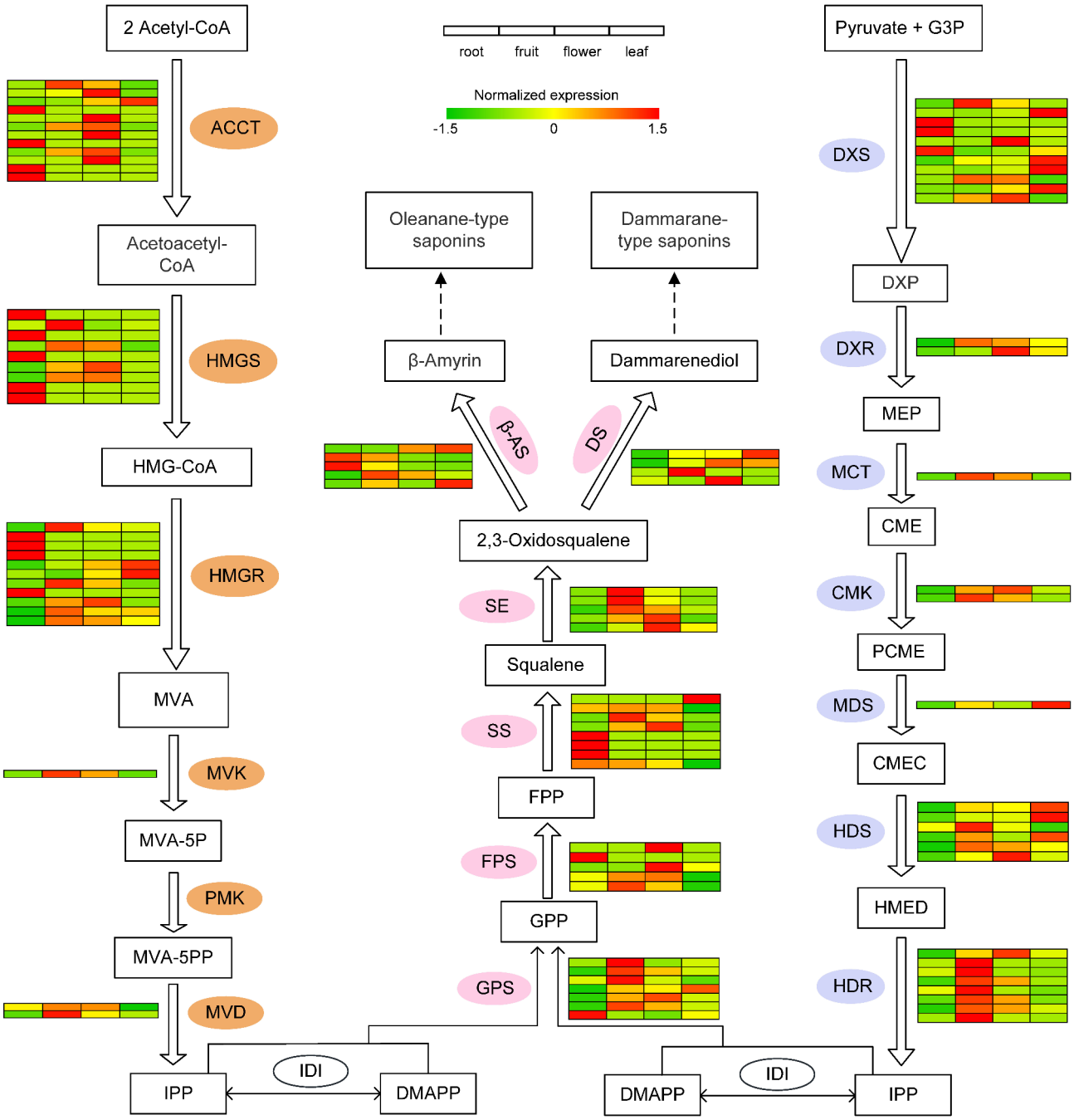

Fig. 7. Heatmap diagrams of putative genes involved in triterpene saponins biosynthesis. Enzymes with orange color represents MVA pathway, purple color represents MEP pathway and pink color represents common saponin biosynthesis pathway. AACT, Acetyl-CoA acetyltransferase; HMGS, hydroxymethylglutaryl-CoA synthase; HMGR, hydroxymethylglutaryl-CoA reductase; MVK, mevalonate kinase; PMK, phosphor mevalonate kinase; MVD, mevalonate diphosphate decarboxylase; IPP, isopentenyl pyrophosphate; DMAPP, allylic isomer dimethylallyl pyrophosphate; IPI, isopentenyl diphosphate isomerase; GPPS, geranylgeranyl pyrophosphate synthase; FPS, farnesyl diphosphate synthase; SS, squalene synthase; SE, squalene epoxidase; DS, dammarendiol synthase; AS, beta-amyrin synthase.

Data availability

The raw sequence data of RNA-seq have been deposited in the Sequence Read Archive (SRA) database, under accession number PRJNA1115811.

Received: 29 June 2024; Accepted: 4 November 2024
Published online: 11 November 2024

References

1. Wan, J.-Z. et al. Saponins of panax japonicus confer neuroprotection against brain aging through mitochondrial related oxidative stress and autophagy in rats. *Curr. Pharm. Biotechnol.* **21**, 667–680 (2020).
2. Wang, X.-J. et al. Panax japonicus and chikusetsusaponins: A review of diverse biological activities and pharmacology mechanism. *Chin. Herb. Med.* **13**, 64–77 (2021).
3. Guo, X. et al. Anti-inflammatory and osteoprotective effects of Chikusetsusaponin IVa on rheumatoid arthritis via the JAK/STAT signaling pathway. *Phytomedicine* **93**, 153801 (2021).
4. Ling, Y., Ouyang, Y., Xue, Y., Yu, J. & Liu, W. In-depth excavation and screening of triterpene saponins from the rhizome of Panax japonicus using high performance liquid chromatography coupled to electrospray ionization and quadrupole time-of-flight mass spectrometry. *J. Chromatogr. Sci.*, bmad014 (2023).
5. Lin, M. et al. Chromosome-level genome of Entada phaseoloides provides insights into genome evolution and biosynthesis of triterpenoid saponins. *Mol. Ecol. Resour.* **22**, 3049–3067 (2022).
6. Dinday, S. & Ghosh, S. Recent advances in triterpenoid pathway elucidation and engineering. *Biotechnol. Adv*, 108214 (2023).
7. da Silva Magedans, Y. V., Phillips, M. A. & Fett-Neto, A. G. Production of plant bioactive triterpenoid saponins: From metabolites to genes and back. *Phytochem. Rev.* **20**, 461–482 (2021).
8. Rahimi, S. et al. Triterpenoid-biosynthetic UDP-glycosyltransferases from plants. *Biotechnol. Adv.* **37**, 107394 (2019).
9. Malhotra, K. & Franke, J. Cytochrome P450 monooxygenase-mediated tailoring of triterpenoids and steroids in plants. *Beilstein J. Org. Chem.* **18**, 1289–1310 (2022).
10. Zhang, Y. et al. Gene excavation and expression analysis of CYP and UGT related to the post modifying stage of gypenoside biosynthesis in Gynostemma pentaphyllum (Thunb.) Makino by comprehensive analysis of RNA and proteome sequencing. *PloS one* **16**, e0260027 (2021).
11. Xie, G. et al. Ultra-performance LC/TOF MS analysis of medicinal Panax herbs for metabolomic research. *J. Sep. Sci.* **31**, 1015–1026 (2008).
12. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
13. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protocols* **8**, 1494–1512 (2013).
14. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
15. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30 (2000).
16. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
17. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
18. Zheng, L. *et al.* Global transcriptome analysis reveals dynamic gene expression profiling and provides insights into biosynthesis of resveratrol and anthraquinones in a medicinal plant Polygonum cuspidatum. *Ind. Crops Prod.* **171**, https://doi.org/10.1016/j.indcrop.2021.113919 (2021).
19. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* **12**, 1–16 (2011).
20. Varet, H., Brillet-Guéguen, L., Coppée, J.-Y. & Dillies, M.-A. SARTools: A DESeq2-and EdgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data. *PloS one* **11**, e0157022 (2016).
21. Dai, X., Sinharoy, S., Udvardi, M. & Zhao, P. X. PlantTFcat: an online plant transcription factor and transcriptional regulator categorization and analysis tool. *BMC Bioinf.* **14**, 1–6 (2013).
22. Zheng, Y. et al. iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* **9**, 1667–1670 (2016).
23. Jin, J. *et al.* PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res*, gkw982 (2016).
24. Zhou, M. & Memelink, J. Jasmonate-responsive transcription factors regulating plant secondary metabolism. *Biotechnol. Adv.* **34**, 441–449 (2016).
25. Iwase, A., Matsui, K. & Ohme-Takagi, M. Manipulation of plant metabolic pathways by transcription factors. *Plant Biotechnol.-Nar* **26**, 29–38 (2009).
26. Liao, W., Mei, Z., Miao, L., Liu, P. & Gao, R. Comparative transcriptome analysis of root, stem, and leaf tissues of Entada phaseoloides reveals potential genes involved in triterpenoid saponin biosynthesis. *BMC Genom.* **21**, 1–12 (2020).
27. Chen, C. et al. Comparative transcriptome and phytochemical analysis provides insight into triterpene saponin biosynthesis in seeds and flowers of the tea plant (Camellia sinensis). *Metabolites* **12**, 204 (2022).
28. Timilsena, Y. P., Phosanam, A. & Stockmann, R. Perspectives on saponins: Food functionality and applications. *Int. J. Mol. Sci.* **24**, 13538 (2023).
29. Moses, T., Papadopoulou, K. K. & Osbourn, A. Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. *Crit. Rev. Biochem. Mol. Boil.* **49**, 439–462 (2014).
30. Zhao, S. et al. Both the mevalonate and the non-mevalonate pathways are involved in ginsenoside biosynthesis. *Plant Cell Rep.* **33**, 393–400. https://doi.org/10.1007/s00299-013-1538-7 (2014).
31. Hou, M., Wang, R., Zhao, S. & Wang, Z. Ginsenosides in Panax genus and their biosynthesis. *Acta Pharmaceutica Sin. B* **11**, 1813–1834 (2021).
32. Tripathi, S., Jadaun, J. S., Chandra, M. & Sangwan, N. S. Medicinal plant transcriptomes: the new gateways for accelerated understanding of plant secondary metabolism. *Plant Genet. Resour.* **14**, 256–269 (2016).
33. Guo, J., Huang, Z., Sun, J., Cui, X. & Liu, Y. Research progress and future development trends in medicinal plant transcriptomics. *Front. Plant Sci.* **12**, 691838 (2021).
34. Rai, A. et al. RNA-seq transcriptome analysis of Panax japonicus, and its comparison with other Panax species to identify potential genes involved in the saponins biosynthesis. *Front. Plant Sci.* **7**, 187083 (2016).
35. Xu, R. et al. Full-length transcriptome sequencing and modular organization analysis of oleanolic acid-and dammarane-type saponins related gene expression patterns in Panax japonicus. *Genomics* **112**, 4137–4147 (2020).

Acknowledgements

We acknowledge Lili Gui, Huifan Yu (Hubei Key Laboratory of Wudang Local Chinese Medicine Research, Hubei University of Medicine) for the help of UPLC analysis.

Author contributions

Y.Z. designed and supervised the study. X.L., Z.Y., X.S., Z.H., J.S., L.Y., P.H., L.Z., and C.L. contributed to the execution of the experiments, Investigation, data management. X.L., G.W., and Y.Z. wrote the manuscript. All

authors reviewed and approved the manuscript.

Funding
This work was funded by the National Natural Science Foundation of China (32000609, 32402391), the Science Research and Development Project of Hubei Province (2022BCE035), the Natural Science Foundation of Hubei Provincial Department of Education (Q20222106), Natural Science Basic Research Program of Shaanxi (2023-JC-ZD-09), Shiyan Innovation and Development Joint Fund (2024AFD105, 2024AFD106), Principle Investigator Program at Hubei University of Medicine (HBMUPI202104), the Advantages Discipline Group (Biology and Medicine) Project in Higher Education of Hubei Province (2022BMXKQT4), the Science Research Program of Hubei Provincial Department of Education (T2023016).

Competing interests
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Statement of permission for collecting *Panax* japonicus
Experimental research and field studies on plants (either cultivated or wild), including the collection of plant material have complied with relevant institutional, national, and international guidelines and legislation. The experimental studies and collection of cultivated *Panax japonicus* plants were conducted in compliance with the collaboration agreements between the Biomedical Research Institute, Hubei University of Medicine (HBMU), and the Longwangzhai Rare Chinese Herbal Medicine Cultivation Farmers' Cooperative (LRCC). Permission for the collection of *P. japonicus* was obtained through this collaboration. Voucher specimens of the plants used in this study have been deposited at the Biomedical Research Institute, HBMU, under the code number VofP-ZZS-2022. The identification of the plant material was confirmed by Prof. Yonghong Zhang at HBMU.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-78720-9.

Correspondence and requests for materials should be addressed to Y.Z.