Scientific Reports

https://doi.org/10.1038/s41598-026-39037-x

Article in Press

Integrative transcriptomic and metabolomic analysis of *Drynaria roosii* reveals genes involved in the biosynthesis of medicinal compounds

Received: 28 August 2025

Accepted: 2 February 2026

Published online: 14 February 2026

Cite this article as: Zhang X., Chen X., Wang Y. *et al.* Integrative transcriptomic and metabolomic analysis of *Drynaria roosii* reveals genes involved in the biosynthesis of medicinal compounds. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-39037-x

Xiangyu Zhang, Xiaofang Chen, Yong Wang, Min Liu, Caiyun Wang & Tao Zhou

Integrative Transcriptomic and Metabolomic Analysis of *Drynaria roosii* Reveals Genes Involved in the Biosynthesis of Medicinal Compounds

Xiangyu Zhang[1,2,*] , Xiaofang Chen[3,*], Yong Wang[1,2], Min Liu[1,2], Caiyun Wang[1,2], Tao Zhou[2,4]

[1] *Bijie Institute of Traditional Chinese Medicine, Bijie, Guizhou 551700, China;*
[2] *Guizhou Key Laboratory for Germplasm Innovation and Resource-Efficient Utilization of Dao-di Herbs, Guiyang 550025, China*
[3] *Bijie Medical College, Bijie Guizhou 551700, China*
[4] *Resource Institute for Chinese & Ethnic Materia Medica, Guizhou University of Traditional Chinese Medicine, Guiyang 550025, China*
*****Correspondence: 19808661986@163.com (Z.X.); nuannuan_1225@126.com(C.X.)

Abstract

The rhizome of *Drynaria roosii* (Drynariae Rhizoma) holds significant medicinal and economic value. It is traditionally used to promote blood circulation, remove blood stasis, and strengthen the kidneys and bones. However, the distribution and biosynthetic pathways of medicinal compounds in different tissues of *D. roosii* remain unclear. In this study, non-targeted metabolomics and transcriptomics analyses were conducted on leaves, stems, and tubers of *D. roosii*, and a high-quality reference transcriptome was obtained using Pacific BioSciences (PacBio) single-molecule real-time (SMRT) sequencing. A total of 1,151 metabolites were identified, including 203 flavonoid-related compounds. Among them, 31 flavonoids-such as quercetin 7-glucoside, tamarixetin, and naringenin 7-rutinoside—were found to be relatively abundant in the tuber. PacBio SMRT sequencing yielded 151,192 consensus reads. A total of 5,581 intron retention (IR) events were identified through alternative splicing analysis, and 56,773 non-redundant transcripts were obtained after transcript redundancy removal. Comparative transcriptome analysis revealed that metabolic pathways such as steroid biosynthesis (ko00100) and phenylpropanoid biosynthesis were enriched in the tuber and leaf. Correlation network analysis identified key genes, including *Glycosyltransferase*, *4CL*, *DELLA* and others, to be significantly associated with the biosynthesis of naringin 6"-rhamnoside and naringenin 7-rutinoside. This study provides a foundation for the resource utilization, medicinal compound biosynthesis, and molecular breeding of *D. roosii.*

Keywords: *Drynaria roosii*; PacBio SMRT; transcriptome; LC-MS

Introduction

Drynaria roosii (Fam: Polypodiaceae) is a perennial traditional Chinese medicinal herb, and its rhizome—commonly known as “Gu-Sui-Bu” (Drynariae Rhizoma) —is

the medicinal part [1]. Rhizome shortly creeping, 1-2 cm in diam.; scales peltate, 7-12 × 0.8-1.5 mm, margin dentate; fronds dimorphic, glabrous. It can be harvested throughout the year, after which the rhizomes are cleaned of soil, dried, and optionally singed to remove fine hairs (scales). Widely utilized in traditional formulations, Drynariae Rhizoma is a component of nearly 183 Chinese patented medicines, driving an estimated yearly demand of up to 6.5 million kilograms across China [2]. The phylogenetic position of *D. roosii* was closely clustered with *Adiantum capillus-veneris*, *Cheilanthes lindheimeri*, and *Pteridium aquilium* subsp [3]. The rhizome is used medicinally and plays a significant role in promoting healing and relieving pain, as well as in tonifying the kidneys, strengthening bones, and treating osteoporosis[4,5]. The water extract of *D. roosii* significantly decreased the receptor activator of nuclear k-B ligand-induced osteoclast differentiation in bone marrow-derived macrophages [2]. A total of 1,435 components were identified across 13 categories, with flavonoids, amino acids, and their derivative, lipids, identified as the main components based on a widely targeted metabolomics analysis.[2]. Drynariae Rhizoma is considered rich in phenolic compounds and flavonoids, such as naringin, neoeriocitrin, and 5,7-dihydroxychromone-7-O-neohesperidoside [6]. Naringin is a natural flavanone glycoside found in *D. fortunei*, *Citrus aurantium* and other citrus fruits [7]. Flavonoids are the main components of Drynariae Rhizoma [8]. Total flavonoids of Drynariae rhizoma exhibit promising pharmacological activities [9,10]. Total flavonoids of Drynariae rhizoma can stimulate bone formation and prevent bone resorption in osteoporosis patients [11]. Integrating the results of qualitative analysis, quercetin, luteolin and kaempferol are the potential pharmacodynamic substances of total flavonoids of Drynariae rhizoma against glucocorticoid-induced osteoporosis [12]. Drynariae Rhizoma samples derived from other botanical origins and Type II stir-fried samples cannot substitute for *D. roosii* rhizome [2]. Greenhouse-cultivated Drynariae Rhizoma exhibits a superior profile of main active constituents compared to the wild type, with significantly elevated contents of flavonols, flavone C-glycosides, sugars, and sugar-related metabolites [13]. However, the distribution and biosynthetic regulation of flavonoids and phenolic compounds in the whole *D. fortunei* plant remain unclear.

Third-generation Pacific BioSciences (PacBio) single-molecule real-time (SMRT) sequencing is advantageous for the identification of gene isoforms and facilitates reliable discoveries of novel genes and novel isoforms of annotated genes, due to its ability to sequence full-length transcripts or fragments with significant lengths [14]. 37,111 isoforms were detected by PacBio SMRT sequencing with an N50 length of 2,317 nt in *Platycladus orientalis* [15]. 71,740 long-noncoding RNA (LncRNAs), 57,548 simple sequence repeats (SSRs), and 3,486 hypothetical transcript factors (TFs) were identified in *Metacrinus rotundus* [16]. Based on the weighted gene co-expression network analysis, *PAL*, *4CL* and *C4H*, and *C3H* and *HCT* acted as the major hub genes involved in naringin and neoeriocitrin synthesis in *Drynaria roosii* [17], respectively. In this study, untargeted LC-MS analysis on different parts of *D. roosii* were performed to identify medicinal compounds highly expressed in the rhizome. PacBio SMRT sequencing were performed on mixed cDNA from various parts to generate a full-length transcriptome, providing a good transcriptomic reference. By integrating transcriptomic and metabolomic data from different tissues, key genes involved in the biosynthesis of highly expressed metabolites can be identified, providing valuable insights into the regulatory pathways of medicinal compound synthesis.

2Materials and Methods

Plant Materials

Plant materials of *D. roosii* were obtained from healthy individuals cultivated under controlled greenhouse conditions at Bijie Institute of Traditional Chinese Medicine. To ensure consistency and minimize environmental variation, plants were maintained under uniform light, temperature, and watering regimes throughout the growth period. Three types of tissues, including tubers, stems, and leaves, were selected for sampling to represent both underground and aboveground organs. For each tissue type, five independent biological replicates were collected from different plants at the same developmental stage. Immediately after excision, all samples were carefully rinsed with sterile distilled water to remove surface contaminants, blotted dry with clean filter paper, and rapidly frozen in liquid nitrogen to preserve RNA integrity and metabolite stability. The frozen materials were then transferred to pre-cooled cryovials, labeled accordingly, and stored at −80°C in an ultra-low temperature freezer until further processing. These preserved samples were subsequently used for transcriptomic and metabolomic analyses.

Untargeted metabolomic profiling using LC-MS/MS

Samples were placed into centrifuge tubes with grinding beads. Metabolites were extracted using 400 μL of extraction solvent (methanol:water = 4:1, v/v) containing 0.02 mg/mL of internal standard (L-2-chlorophenylalanine). After grinding under frozen conditions, the samples underwent ultrasonic extraction at low temperature. The extracts were left to stand at −20 °C, and the supernatants were collected. Equal volumes of all extracted metabolites were pooled to prepare quality control (QC) samples. Metabolomic analysis was performed using a UHPLC-Q Exactive HF-X system (Thermo Fisher Scientific) equipped with ultra-high-performance liquid chromatography coupled to high-resolution tandem mass spectrometry (LC-MS/MS). Raw LC-MS data were imported into Progenesis QI software (Waters Corporation, Milford, USA) for baseline filtering, peak detection, integration, retention time correction, and peak alignment. Variables with a relative standard deviation (RSD) >30% in QC samples were removed, and the data were log10-transformed for further analysis. Significantly different metabolites were identified based on the variable importance in projection (VIP) values from the OPLS-DA model and *p-values* from Student's t-test. Metabolites with VIP > 1 and $p < 0.05$ were considered significantly differential.

PacBio Library Construction and Single-Molecule Sequencing

Total RNA was extracted from tissue samples. The concentration and purity of the extracted RNA were assessed using a NanoDrop 2000 spectrophotometer. RNA integrity was evaluated by agarose gel electrophoresis, and RNA integrity number (RIN) values were determined using the Agilent 2100 Bioanalyzer. Samples with OD260/280 values between 1.8 and 2.2 were considered qualified for subsequent library construction. Full-length cDNA of mRNA was synthesized using the SMARTer™ PCR cDNA Synthesis Kit, followed by PCR amplification. Amplified cDNA was purified using PB magnetic beads. The cDNA was end-repaired and ligated to SMRTbell adapters. After exonuclease treatment, the libraries were purified again using PB magnetic beads to obtain sequencing-ready libraries. Libraries passing quality control were sequenced on the PacBio platform (Pacific Biosciences, Menlo Park, CA, USA). The raw BAM files have been uploaded to the NCBI SRA database (BioProject accession number: PRJNA1291835). The raw sequencing reads were

processed through the SMRT Link v8.0 pipeline to generate circular consensus sequence (CCS) reads, during which low-quality reads and chimeric sequences were removed. Quality control steps included CCS generation, classification, and clustering. Redundant transcripts were removed using MMseqs2 [18], and transcriptome completeness was evaluated using BUSCO (Benchmarking Universal Single-Copy Orthologs) [19]. All transcript sequences were annotated using Diamond [20] and aligned against public databases, including the NCBI non-redundant protein (NR), Swiss-Prot, KOG, Pfam, KEGG [21], and eggNOG databases [22].

Illumina RNA-Seq Library Construction and Sequencing

Total RNA was extracted from the tissue using TRIzol® Reagent . Then RNA quality was determined by 5300 Bioanalyser (Agilent) and quantified using the ND-2000 (NanoDrop Technologies). Only high-quality RNA sample[OD260/280=1.8~2.2,OD260/230≥2.0,RQN≥6.5, 28S:18S≥1.0, >1μg] was used to construct sequencing library. The RNA-seq transcriptome library was prepared following Illumina® Stranded mRNA Prep, Ligation (San Diego, CA). The sequencing library was performed on DNBSEQ-T7 platform(PE150) using DNBSEQ-T7RS Reagent Kit (FCL PE150) version 3.0. The raw reads have been uploaded to the NCBI SRA database (BioProject accession number: PRJNA1291623). Raw sequencing data were quality controlled using fastp [23] with default parameters. The non-redundant PacBio SMRT transcriptome was used as the reference. Clean reads were aligned to the reference using Bowtie2 v2.4.3 [24], and transcript quantification was performed using RSEM v1.2.12 [25], with gene expression levels estimated as FPKM (fragments per kilobase of transcript per million mapped reads). Differentially expressed genes (DEGs) were identified using DESeq2 [26] based on the read counts. The criteria for DEG selection were |fold change| > 2 and *adjusted p-value* < 0.05. Functional enrichment analysis, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, were conducted using the clusterProfiler v4.8.1 R package [27]. The WGCNA R package was utilized to construct a weighted gene co-expression network [28]. Genes with an expression level greater than 1 in any sample were filtered, and the top 5,000 genes were then selected based on highest median absolute deviation (MAD). The correlations between metabolites and genes were calculated using the `corAndPvalue` function from the WGCNA package. Significant correlations were selected based on the criteria $p < 0.01$ and $|r| > 0.95$. Network visualization was performed using ggnetwork v0.5.13 package, with the layout set to "fruchtermanreingold".

Quantitative real-time PCR analysis (qRT-PCR)

Seven genes were selected, and *Actin* gene was use as the reference gene. Primer Premier 5 was used to design primers, and primer sequences are listed in Table S1. The SYBR Green QPCR Mix was used to conduct qRT-PCR. Three biological replicates were analyzed. Reactions were carried out at 94 °C for 30 s, followed by 450 cycles at 94 °C for 5 s and 60 °C for 30 s. The mRNAs expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method [29].

Results

Untargeted Metabolites in Different Tissue Parts

The raw data underwent a series of preprocessing steps, including filtering, imputation, normalization, and logarithmic transformation. A total of 917 metabolites were identified in positive ion mode and 488 in negative ion mode. Metabolite annotation based on the HMDB database (Fig. 1a) revealed that they mainly belonged to Lipids and lipid-like molecules (319), Phenylpropanoids and polyketides (267), and Organic acids and derivatives (152). KEGG database annotation (Fig. 1b) showed that 261 metabolites were involved in Global and overview maps, 100 in Biosynthesis of other secondary metabolites, and 66 in Amino acid metabolism. Principal Component Analysis (PCA) of untargeted metabolites across samples (Fig. 1c) indicated clear differences among tissues, with close clustering within biological replicates, suggesting good reproducibility. Based on defined thresholds for differential metabolite selection, analysis of differential expression between Stem_vs_Leaf (Fig. 1d) identified 115 upregulated and 42 downregulated metabolites. Between Tuber_vs_Leaf (Fig. 1e), 468 metabolites were upregulated and 631 downregulated. Between Tuber_vs_Stem (Fig. 1f), 293 were upregulated and 516 downregulated. A heatmap of flavonoid-related metabolites showed consistent levels within the same tissue, with 31 flavonoid compounds-including Quercetin 7-glucoside, Tamarixetin, and Naringenin 7-rutinoside-found at relatively high levels in the tuber (Fig. 1g). Cluster analysis of annotated differential metabolites across the three tissue comparisons (Fig. 1h–q) revealed that metabolites in Cluster 1 (including (1R,2R,4R)-1,8-Epoxy-p-menthane-2,4-diol,2-(3,4-Dihydroxyphenyl)-5,6-dihydroxy-7-methoxy-4H-1-benzopyran-4-one, 3-Amino-2,3-dihydrobenzoic acid, Tobramycin, Undecanedioic acid) and Cluster 7 (including 6''-Acetylapiin, Niacinamide) were enriched in leaf, while those in Cluster 4 were more abundant in stem.

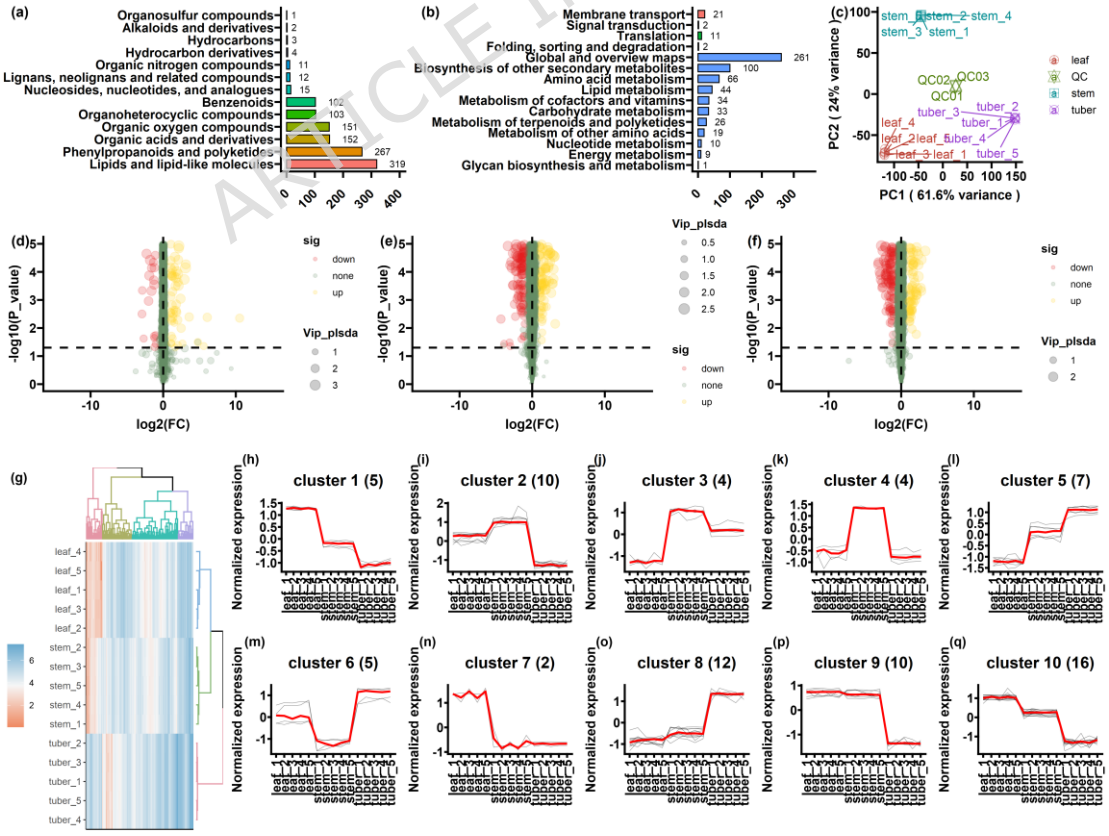

Figure 1. Classification and clustering analysis of untargeted metabolites in different tissues of *D. roosii*.
a: Metabolite classification based on the HMDB database; b: Metabolite classification based on the KEGG database; c: PCA analysis of untargeted metabolites; d: Volcano plot of differential metabolites between Stem and Leaf ; e: Volcano plot of differential metabolites between Tuber and Leaf); f: Volcano plot of differential metabolites between Tuber and Stem; g: Heatmap of relative flavonoid content in different tissues; h–q: Cluster analysis of differential metabolites. Note: In subfigure c, different groups are represented using distinct colors and shapes. In subfigures g–p, clustering was performed using the mfuzz method.

Acquisition of a High-Quality PacBio SMRT Reference Transcriptome

Through PacBio SMRT sequencing, a total of 6,955,473 HiFi reads and 6,942,837 FLNC reads were obtained, resulting in 151,192 consensus reads (Fig. 2a). SSR analysis of the consensus reads revealed that the p2 type exhibited a greater diversity of categories (Fig. 2b). Alternative splicing analysis (Fig. 2c) identified 5,581 IR-type events, while the MXE type had the lowest number, with only 22 events. After removing redundancy, a total of 56,773 transcripts were obtained (Fig. 2d). To assess the completeness of the transcriptome, BUSCO analysis was performed, and more than 66% of the BUSCOs were found to be complete and single-copy. After transcription factor identification, the most abundant types were C3H (229), bHLH (181), and bZIP (140) (Fig. 2e). Transcriptional regulator (TR) identification revealed that SET (86), PHD (68), and mTERF (66) were the most common (Fig. 2f).

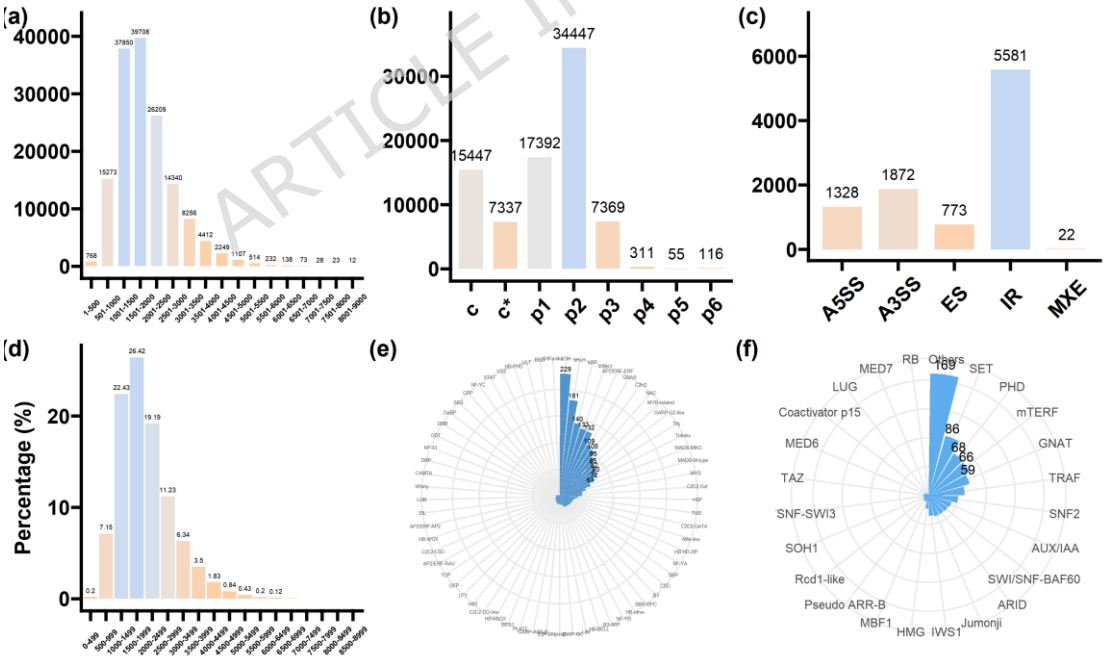

Figure 2. Transcriptome of *D.roosii* based on PacBio SMRT sequencing.
a: Length distribution of consensus reads; b: Distribution of SSR types; c: Distribution of alternative splicing types; d: Length distribution of non-redundant sequences; e: Classification of transcription factors in non-redundant sequences; f: Classification of transcriptional regulators in non-redundant sequences. Note: in c,

A5S: alternative 5′ splice site; A3S: alternative 3′splice site; ES: exon skipping; IR: Intron retention; MXE (mutually exclusive exons).

Comparative Analysis of Transcriptomes Across Tissues

A total of 15 transcriptome libraries were constructed and quantitatively aligned against the non-redundant full-length transcriptome generated by PacBio SMRT sequencing. The mapping rates ranged from 84.44% to 85.66%, indicating high alignment efficiency and suggesting that the PacBio SMRT non-redundant transcriptome is suitable as a reference for downstream analyses. Gene expression quantification was followed by clustering and PCA using all samples. The hierarchical clustering heatmap (Fig. 3a) revealed high correlation among biological replicates and distinct differences between tissues. The first two principal components (PC1 and PC2) accounted for 72% and 16.3% of the variance (Fig. 3b), respectively, indicating a strong cumulative contribution. Compared with stem, a total of 9,821 differentially expressed genes (DEGs) were downregulated and 6,524 (39.91%) were upregulated in tuber (Fig. 3c). In the comparison between tuber and leaf, 11,675 DEGs were downregulated and 10,033 (46.22%) were upregulated in tuber. GO enrichment analysis of DEGs in the tuber_vs_stem comparison (Fig. 3d) revealed significant enrichment in biological processes such as microtubule-based process (GO:0007017), calcium-mediated signaling (GO:0019722), and second-messenger-mediated signaling (GO:0019932). KEGG pathway analysis showed enrichment in Cytochrome P450 (ko00199) and Flavonoid biosynthesis (ko00941). In the tuber_vs_leaf comparison, GO enrichment highlighted chlorophyll metabolic process (GO:0015994) and tetrapyrrole metabolic process (GO:0033013), while KEGG pathways such as Photosynthesis proteins (ko00194), Steroid biosynthesis (ko00100), and Phenylpropanoid biosynthesis (ko00940) were significantly enriched. Clustering analysis of all DEGs (Fig. 3e-n) further revealed that certain genes maintained tissue-specific expression patterns.

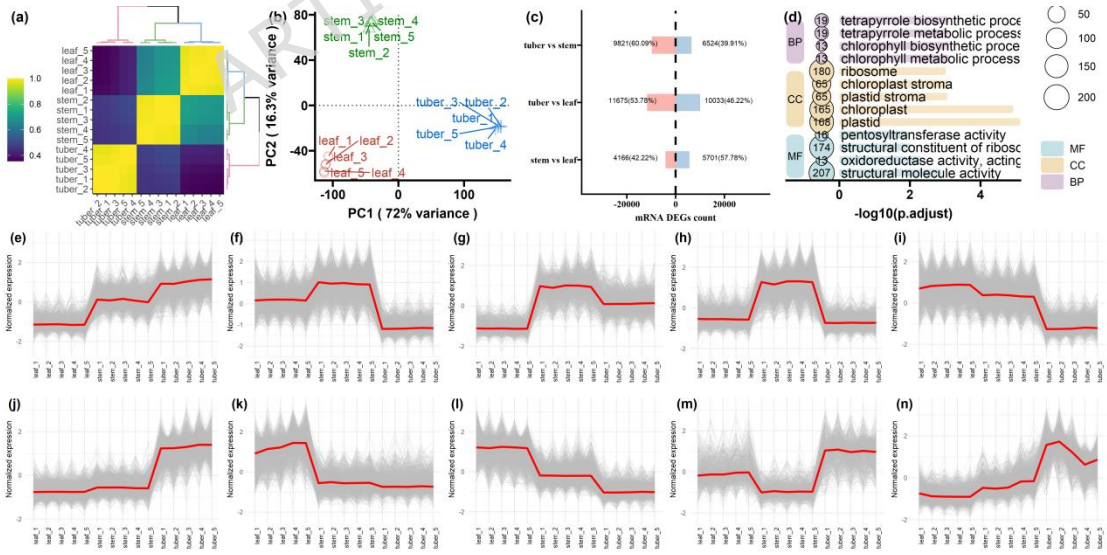

Figure 3. Transcriptome analysis across different tissues of *D. roosii* a: Hierarchical clustering heatmap of gene expression across samples; b: PCA analysis of gene expression across samples; c: Number of DEGs across different tissue comparisons; d: GO enrichment analysis of DEGs in the Tuber_vs_Leaf comparison; e–n: Cluster analysis of DEGs. Note: In subfigure b, different groups are represented by distinct

colors and shapes. In subfigures e–n, clustering was performed using the mfuzz method.

WGCNA analysis

Seven metabolites related to naringenin were selected for co-expression network analysis with genes (Fig. 4), and four modules were found to be significantly associated with different traits. The blue module showed a significant positive correlation with Naringenin 7-O-β-D-glucoside (r = 0.90, $p = 3.9 \times 10^{-6}$). Genes in this module were relatively highly expressed in leaves, and GO enrichment analysis indicated significant enrichment in processes such as photosynthesis (GO:0015979), carbohydrate metabolic process (GO:0005975), and oxazole or thiazole biosynthetic process (GO:0018131). KEGG enrichment further revealed enrichment in Photosynthesis proteins (BR:ko00194) and Phenylpropanoid biosynthesis (ko00940). The green module was significantly negatively correlated with Naringenin 7-rutinoside ($r = -0.96$, $p = 1.3 \times 10^{-8}$). Genes in this module were relatively highly expressed in stems, and key hub genes included *CYP75B1*, *CYP736*, *LOX1_5*, and *SWEET* family members. KEGG enrichment showed that pathways such as Limonene and pinene degradation (ko00903), Ascorbate and aldarate metabolism (ko00053), Glycerolipid metabolism (ko00561), and β-alanine metabolism (ko00410) were enriched. The black module was also significantly negatively correlated with Naringenin chalcone ($r = -0.94$, $p = 2.1 \times 10^{-7}$). Genes in this module were relatively highly expressed in stems, and hub genes included *AUX1*, *JAZ*, and *LOX1_5*. The brown module showed a significant positive correlation with Naringenin 7-rutinoside (r = 0.95, $p = 8.8 \times 10^{-8}$). Genes such as *GST*, *MDH2*, and *ARF1* were relatively highly expressed in tubers. GO enrichment analysis indicated significant enrichment in amide biosynthetic process (GO:0043604), translation (GO:0006412), and peptide biosynthetic process (GO:0043043).

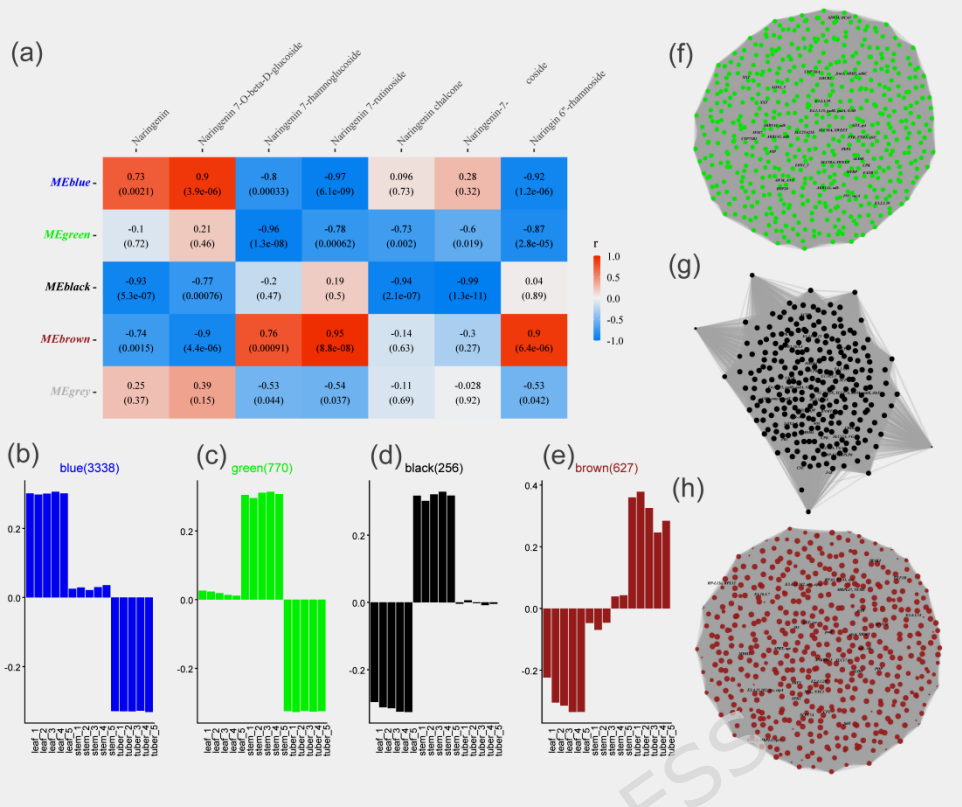

Figure 4. WGCNA (Weighted Gene Co-expression Network Analysis).
a: Module–trait relationship plots. b: Expression of eigengenes in the blue module. c: Expression of eigengenes in the green module. d: Expression of eigengenes in the black module. e: Expression of eigengenes in the brown module. f: Network visualization of the green module. g: Network visualization of the black module. h: Network visualization of the brown module. Note: In panel (a), the numbers represent Pearson's correlation coefficients, with positive correlations shown in red and negative correlations shown in blue.

Correlation Network Analysis

Seven metabolites related to naringenin were selected, and a correlation network was constructed between their relative abundances and gene expression levels using batch correlation analysis. The results showed that genes such as *Rapid ALkalinization Factor* (RALF) and *Lipoxygenase* were significantly positively correlated with Naringenin content, while genes like *UDP-glycosyltransferase 83A1-like*, *4CL* were significantly negatively correlated. Naringenin chalcone showed significant negative correlations with *Cytochrome P450 CYP90B30* (Dr_61012) and *JAZ* (Dr_174001). In addition, genes such as *Glycosyltransferase (UGT85A)*, *4CL* and DELLA were significantly correlated with Naringin 6''-rhamnoside and Naringenin 7-rutinoside (Fig. 5). To verify the full-length RNA-Seq data, eight genes were randomly selected for qRT-PCR analysis. The results showed that the expression trends of genes were consistent with the results of RNA-seq analysis (Fig. S1), which proved that the RNA-Seq data were reliable

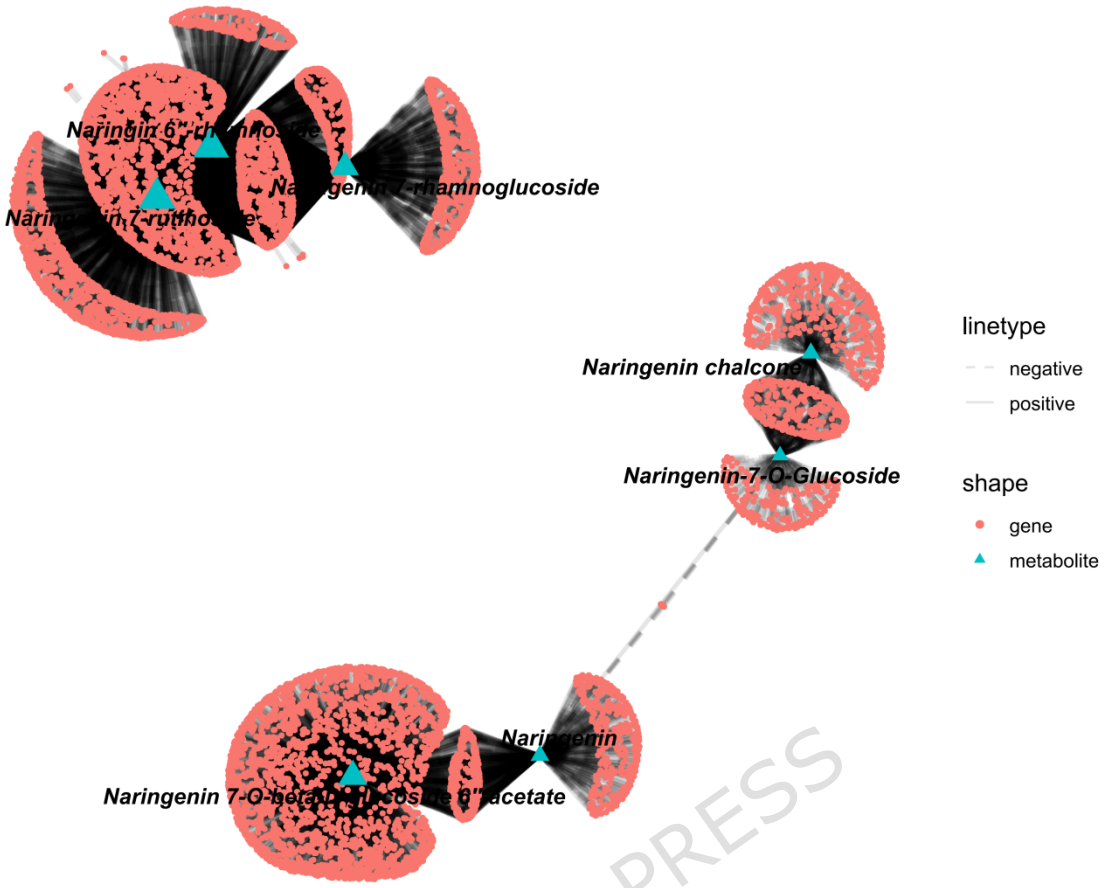

Figure 5. Correlation network analysis between naringenin-related metabolites and genes. Note: Metabolites and genes were represented by different shapes. The type of line (linetype) indicated the direction of correlation (positive or negative). Correlations were filtered using the criteria $p < 0.01$ and $|r| > 0.95$.

Discussion

D. roosii, a traditional medicinal plant in Guizhou, China, is known for promoting blood circulation, removing blood stasis, and strengthening the kidneys and bones [6]. This species has a broad suitable habitat range, mainly distributed across the southwestern, central, southern, and eastern regions of China. Phylogenetic analysis based on chloroplast genomes revealed that *Lygodium japonicum*, *Diplopterygium glaucum*, and *Osmundastrum cinnamomeum* were identified *D. roosii* belongs to Polypodiales [30]. However, research on the distribution of medicinal components, gene expression patterns, and biosynthetic pathways in different tissues of *D. roosii* remains relatively limited. Previous studies on *D. roosii*, particularly Sun et al. [17], primarily focused on temporal variation in the biosynthesis of naringin and neoeriocitrin across different growth years, aiming to elucidate year-dependent regulatory mechanisms of these key flavonoids. In contrast, the present study emphasizes spatial variation in metabolite accumulation by systematically comparing different tissues of *D. roosii*, with particular focus on the rhizome, which represents the primary medicinal organ. Through untargeted LC–MS profiling, we characterized tissue-specific metabolic landscapes and identified medicinal compounds preferentially accumulated in the rhizome, this study identifies candidate genes potentially involved in the biosynthesis of rhizome-enriched metabolites. This tissue-centered framework complements previous time-series–based analyses and provides new insights into the spatial regulation of medicinal compound biosynthesis in *D. roosii*.

In this study, greenhouse-grown plants were selected to minimize environmental heterogeneity and to ensure biological reproducibility, which is critical for comparative metabolomic profiling across tissues. In contrast, wild populations often exhibit substantial heterogeneity in genetic background and growth environments, making it difficult to control provenance and maintain consistent living conditions across samples. Non-targeted metabolomic profiling was conducted on different tissues of *D. roosii*, revealing significant metabolic differences among tissues. Samples from the same tissue type clustered closely together, indicating good biological reproducibility (Fig. 1c). A total of 917 metabolites were identified in positive ion mode and 488 in negative ion mode. Based on HMDB database annotation, a total of 1,151 metabolites were identified, including 319 lipids and lipid-like molecules, 267 phenylpropanoids and polyketides, and 152 organic acids and derivatives. According to plant metabolite classification, 203 flavonoid-related compounds were identified, among which flavonoid glycosides were the most abundant group (102 compounds). Seven naringenin-related compounds, such as "Naringenin" and "Naringenin chalcone," were also identified. The highest number of differentially accumulated metabolites (DAMs) was observed in the comparison between tuber and leaf tissues. Cluster analysis of DAMs from the three comparison groups (Fig. 1h-q) revealed that metabolites in Cluster 1 and Cluster 7 were relatively abundant in leaf tissues. Cluster 4 metabolites were highly accumulated in stem, including 2,3-Dihydro-4-(4-methoxyphenyl)-1H-phenalene-1,2,3-triol, Glu-Tyr, Neogitogenin, and Syringic acid. Metabolites in Cluster 8 were highly accumulated in tuber tissues, including 1-Ipomeanol, 5-Deoxykievitone, and Buntansin A, among a total of 12 compounds. A heatmap analysis of flavonoid-related metabolites showed high consistency within biological replicates from the same tissue type. Notably, 31 flavonoid metabolites—including quercetin 7-glucoside, tamarixetin, and naringenin 7-rutinoside—were found at relatively high levels in the tuber tissue.

A high-quality reference transcriptome can be obtained using PacBio SMRT sequencing. In this study, a total of 6,955,473 HiFi reads and 6,942,837 FLNC reads were generated, resulting in 151,192 consensus reads (Fig. 2a). After redundancy removal, 56,773 transcripts were obtained (Fig. 2d). Transcription factor (TF) analysis identified a large number of members in the C3H (229), bHLH (181), and bZIP (140) families (Fig. 2e). Transcriptional regulator (TR) analysis showed that SET (86), PHD (68), and mTERF (66) families were the most abundant (Fig. 2f). A total of 15 transcriptome libraries were constructed from different tissues of *D. roosii*, and expression quantification was performed based on the PacBio SMRT non-redundant full-length transcriptome. The mapping rate ranged from 84.44% to 85.66%, indicating high alignment efficiency and supporting the use of the PacBio SMRT full-length transcriptome as a reference for downstream analyses. GO enrichment analysis of DEGs in the tuber_vs_leaf comparison revealed enrichment in biological processes such as chlorophyll metabolic process (GO:0015994) and tetrapyrrole metabolic process (GO:0033013). KEGG pathway enrichment analysis indicated significant enrichment in pathways such as photosynthesis proteins (ko00194), steroid biosynthesis (ko00100), and phenylpropanoid biosynthesis (ko00940), suggesting that differential expression of these pathways is associated with the biosynthesis of differentially accumulated metabolites.

Flavonoids, as one of the most widespread classes of secondary metabolites, typically possess a C6-C3-C6 carbon skeleton. They are water-soluble pigments stored in vacuoles and are responsible for the coloration of flowers, fruits, leaves,

tubers, and other plant tissues [31]. Based on the degree of oxidation of the C-ring and the number of hydroxyl or methyl groups on the aromatic rings, flavonoids can be classified into 12 subgroups, including chalcones, flavanones, flavones, isoflavones, leucoanthocyanidins, proanthocyanidins, and anthocyanins [32,33]. Flavonoid biosynthesis originates from phenylalanine through the phenylpropanoid pathway, with phenylalanine itself synthesized via the shikimate pathway. Through the WGCNA analysis, *CHI*, *C4H*, *PAL*, and *F3H* may play crucial roles in the accumulation of flavonoids in *Litsea coreana* var. *sinensis* [34] .In this study, seven metabolites related to naringenin were selected, and a correlation network was constructed between their relative metabolite levels and gene expression using WGCNA and batch correlation analysis. The results showed that four modules were found to be significantly associated with different traits (Fig. 4). The blue module showed a significant positive correlation with Naringenin 7-O-β-D-glucoside (r = 0.90, $p = 3.9 \times 10^{-6}$). The green module was significantly negatively correlated with Naringenin 7-rutinoside ($r = -0.96$, $p = 1.3 \times 10^{-8}$). Genes in green module were relatively highly expressed in stems, and hub genes included *AUX1*, *JAZ*, and *LOX1_5*.The black module was also significantly negatively correlated with Naringenin chalcone ($r = -0.94$, $p = 2.1 \times 10^{-7}$). Genes such as *GST*, *MDH2*, and *ARF1* were relatively highly expressed in tubers. *ARF2* directly regulates the expression of *MYB12* and *FLS* genes, and indirectly regulates *MYB11* and *MYB111* genes increased flavonol content in *Arabidopsis thaliana* [35]. Genes significantly positively correlated with naringenin content included *Rapid ALkalinization Factor* (RALF) and Lipoxygenase. Genes such as *UGT85A*, *4CL* and DELLA were significantly correlated with Naringin 6''-rhamnoside and Naringenin 7-rutinoside (Fig. 5) . The identification of these correlation-associated genes lays the groundwork for subsequent investigations into the biosynthesis and functional analysis of the related metabolites. Future work integrating transcriptomics, metabolite flux analysis, and functional validation (*e.g.,* gene editing and overexpression) will be essential to clarify causal relationships and to uncover the molecular mechanisms underlying tissue-specific flavonoid deposition. Ultimately, such insights may facilitate targeted improvement of nutritional quality and medicinal properties traits in economically important plant species.

Conclusions

D. roosii possesses significant medicinal and economic value; however, the differences in medicinal components and gene expression across its various tissues remain unclear. Using LC-MS, a total of 1,151 metabolites were identified, including 203 flavonoid-related compounds. Among them, 31 flavonoid metabolites such as quercetin 7-glucoside, tamarixetin, and naringenin 7-rutinoside were detected. PacBio SMRT sequencing generated 56,773 non-redundant transcripts. Compared to leaf, the tuber exhibited 11,675 (53.78%) downregulated and 10,033 (46.22%) upregulated DEGs. GO enrichment analysis revealed significant enrichment in biological processes such as microtubule-based process (GO:0007017), calcium-mediated signaling (GO:0019722), and second-messenger-mediated signaling (GO:0019932).

WGCNA and correlation network analysis identified genes associated with naringenin-related metabolites. Notably, *Glycosyltransferase* and *DELLA* showed significant correlations with the biosynthesis of Naringin 6''-rhamnoside and Naringenin 7-rutinoside.

Data availability

The raw bam file from PacBio SMRT sequencing has been deposited in the NCBI SRA database (BioProject acc. PRJNA1291835) (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1291835/). The raw reads generated from Illumina sequencing have been deposited in the NCBI SRA database (BioProject acc. PRJNA1291623) (https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1291623).

Contributions

Conceptualization, Z.X. and C.X.; methodology, Z.X. and C.X..; software, Z.X..; validation, W.Y, W.C. and Z.T.; formal analysis, Z.X. and C.X.; investigation, Z.X. and C.X.; resources, W.Y, L.M., W.C. and Z.T.; data curation, X.X.; writing—original draft preparation, Z.X. and C.X.; writing—review and editing, Z.X. and C.X..; visualization, Z.X.; supervision, Z.X.; project administration, Z.X.; funding acquisition, Z.X. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by Guizhou Provincial Science and Technology Plan Project (QKH Support [2023] General 003; Talent Base Project of Guizhou Provincial Committee Organization Department (RCJD2020-21); Guangzhou Science and Technology Plan Project (2023B03J1292); Bijie Science and Technology Innovation Platform and Talent Team (BKH [2023] No. 66); Scientific Research Team Project of Bijie Medical College (BJYZXT202401).

Conflict of interest

The authors have no relevant financial or non-financial interests to disclose.

References

1 Chang, H.-C. *et al.* In vitro culture of Drynaria fortunei, a fern species source of Chinese medicine “Gu-Sui-Bu”. *In Vitro Cellular & Developmental Biology-Plant* **43**, 133-139 (2007).

2 Chang, N. *et al.* Epiphytic Patterns Impacting Metabolite Diversity of Drynaria roosii Rhizomes Based on Widely Targeted Metabolomics. *Metabolites* **14**, 409 (2024).

3 Sun, M., Li, J., Li, D. & Shi, L. Complete chloroplast genome sequence of the medical fern Drynaria roosii and its phylogenetic analysis. *Mitochondrial Dna Part B* **2**, 7-8 (2017).

4 Han, F. *et al.* Action Mechanism and Clinical Research Progress of Gusuibu (Rhizoma Drynariae) Fighting Against Osteoporosis. *Acta Chinese Medicine* **40**, 812-819, doi:10.16368/j.issn.1674-8999.2025.04.131 (2025).

5 Yu, X. *et al.* Research progress on the clinical efficacy mechanism of total flavonoids of rhizoma drynariae in the treatment of osteoporotic fracture. *Chinese Journal of Osteoporosis* **31**, 149-156 (2025).

6 Dong, Y. *et al.* Metabolite profiling of Drynariae Rhizoma using 1H NMR and HPLC coupled with multivariate statistical analysis. *Journal of Natural Medicines* **77**, 839-857 (2023).

7 Chen, R., Qi, Q.-L., Wang, M.-T. & Li, Q.-Y. Therapeutic potential of naringin: an overview. *Pharmaceutical biology* **54**, 3203-3210 (2016).

8 Song, S.-h. *et al.* Effects of total flavonoids from Drynariae Rhizoma prevent bone loss in vivo and in vitro. *Bone reports* **5**, 262-273 (2016).

9 Song, S. *et al.* Total flavonoids of Drynariae Rhizoma prevent bone loss induced by Hindlimb unloading in rats. *Molecules* **22**, 1033 (2017).

10 Zhang, Y. *et al.* Total flavonoids from Rhizoma Drynariae (Gusuibu) for treating osteoporotic fractures: implication in clinical practice. *Drug Design, Development and Therapy*, 1881-1890 (2017).

11 Li, W. *et al.* Effects of total flavonoids of Rhizoma Drynariae on biochemical indicators of bone metabolism: a systematic review and meta-analysis. *Frontiers in Pharmacology* **15**, 1443235 (2024).

12 Zhang, F. *et al.* Total flavonoids of drynariae rhizoma improve glucocorticoid-induced osteoporosis of rats: UHPLC-MS-based qualitative analysis, network pharmacology strategy and pharmacodynamic validation. *Frontiers in endocrinology* **13**, 920931 (2022).

13 Qiu, D., Luo, Y., Li, C., Du, C. & Yuan, X. Comparative Analysis of Broad-targeted Metabolomics Between Greenhouse Cultivated and Wild Collected Drynaria fortune. *Guangdong Agricultural Sciences* **51**, 30-43, doi:10.16768/j.issn.1004-874X.2024.05.003 (2024).

14 Rhoads, A. & Au, K. F. PacBio sequencing and its applications. *Genomics, proteomics & bioinformatics* **13**, 278-289 (2015).

15 Liao, T. *et al.* Full-length transcriptome characterization of Platycladus orientalis based on the PacBio platform. *Frontiers in Genetics* **15**, 1345039 (2024).

16 Li, Z. *et al.* Comprehensive analysis of Metacrinus rotundus full length transcriptome. *Scientific Reports* **15**, 6723 (2025).

17 Sun, M.-Y. *et al.* Full-length transcriptome sequencing and modular organization analysis of the naringin/neoeriocitrin-related gene expression pattern in Drynaria roosii. *Plant and Cell Physiology* **59**, 1398-1414 (2018).

18 Mirdita, M., Steinegger, M., Breitwieser, F., Söding, J. & Levy Karin, E. Fast and sensitive taxonomic assignment to metagenomic contigs. *Bioinformatics* **37**, 3029-3031 (2021).

19 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).

20 Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature methods* **18**, 366-368 (2021).

21 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27-30 (2000).

22 Jensen, L. J. *et al.* eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic acids research* **36**, D250-D254 (2007).

23 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890, doi:10.1093/bioinformatics/bty560 (2018).

24 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359 (2012).

25 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* **12**, 1-16 (2011).

26 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 1-21 (2014).

27 Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The innovation* **2** (2021).

28 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**, 559 (2008).

29 Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *methods* **25**, 402-408 (2001).

30 Sun, M., Li, J., Li, D. & Shi, L. Complete chloroplast genome sequence of the medical fern Drynaria roosii and its phylogenetic analysis. *Mitochondrial DNA Part B*, doi:10.1080/23802359.2016.1275835 (2017).

31 Dong, N. Q. & Lin, H. X. Contribution of phenylpropanoid metabolism to plant development and plant–environment interactions. *Journal of integrative plant biology* **63**, 180-209 (2021).

32 Tanaka, Y., Sasaki, N. & Ohmiya, A. Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *The Plant Journal* **54**, 733-749 (2008).

33 Sasaki, N. & Nakayama, T. Achievements and perspectives in biochemistry concerning anthocyanin modification for blue flower coloration. *Plant and Cell Physiology* **56**, 28-40 (2015).

34 Xie, N. *et al.* Integrated transcriptomic and WGCNA analyses reveal candidate genes regulating mainly flavonoid biosynthesis in Litsea coreana var. sinensis. *BMC Plant Biology* **24**, 231 (2024).

35 Jiang, W., Xia, Y., Su, X. & Pang, Y. ARF2 positively regulates flavonols and proanthocyanidins biosynthesis in Arabidopsis thaliana. *Planta* **256**, 44 (2022).