scientific reports

Check for updates

OPEN

Tissue-specific transcriptome analyses unveils candidate genes for flavonoid biosynthesis, regulation and transport in the medicinal plant *Ilex asprella*

Qing Ma[1,2,3,5], Saidi Wang[1,5], Haitao Tan[4], Zhongke Sun[1], Chengwei Li[1]✉ & Gaoyang Zhang[1]✉

It is not clear that the genes involved with flavonoids synthesis, regulation and transport in *Ilex asprella*. Transcriptome analysis of leaf, stem and root has uncovered 28,478 differentially expressed genes (DEGs) that are involved in various biological processes. Among these, the expression of 31 candidate synthetase genes, 19 transcription factors, and 5 transporters associated with flavonoid biosynthesis varies across tissues, encompassing seven complete biosynthetic pathways (stilbene, aurone, flavone, isoflavone, flavonol, phlobaphene, and anthocyanin) and one partial pathway (proanthocyanidin). Tissue-specific expression patterns suggest that the stilbenes, aurones, flavones and anthocyanin branches are more prominent in roots, as indicated by key genes such as *STS(Ilex_044726)*, *CH4′GT(Ilex_047989)*, *FNS(Ilex_043640)* and *UFGT(Ilex_014720)*. In leaves, the phlobaphenes and flavonols branches are dominant, determined by *CHI(Ilex_005941)*, *FNR(Ilex_039777)* and *FLS(Ilex_046424)*. The isoflavone pathway appears to be more active in stems due to the presence of *IFS(Ilex_029360)*, mirroring the accumulation of the intermediate metabolite chalcone, which is regulated by *CHS(Ilex_047537)*. The absence of *LAR* genes implies that gallocatechin, and catechin liked proanthocyanidins cannot be synthesized in *I. asprella*. Meanwhile, the general phenylpropanoid pathway is more active in roots, stems than in leaves, as evidenced by the expression of *PAL(Ilex_042231, Ilex_014816)*, *C4H(Ilex_017598)*, and *4CL(Ilex_042033)*. Flavanone, dihydroflavonol and leucoanthocyanidin, key intermediates, accumulate more rapidly in stem, stem and root, respectively, regulated by *CHI(Ilex_005941)*, *F3H(Ilex_004635)* and *DFR(Ilex_004771)*. Correlation and network analyses reveal that candidate regulators and transporters are closely associated with the synthesis genes. The study provides profound snoop into flavonoids metabolism in *I. asprella* and offers valuable refer for medicinal plant.

Keywords *Ilex asprella*, Transcriptome, Flavonoids, Synthesis, Regulation, Transport

Ilex asprella Champ. ex Benth. (Aquifoliaceae) is a deciduous shrub reaching heights of 1–2 m. It thrives in sparse forests on sloping barren land or in thickets, predominantly across the southern Chinese provinces, including Guangdong, Hunan and the Guangxi Zhuang Autonomous Region. Its roots and leaves, known as "Gangmei" in China, are well known and often used in Chinese folk medicine, either alone or in combination with other plants, to treat influenza, tonsillitis, sphagitis, trachitis, and pertussis[1].

Previous chemical investigations on *I. asprella* have led to the identification of a plethora of bioactive compounds, including triterpenoids, flavonoids, phenolic acids, glycosides and steroids[1–3]. While contemporary research has primarily focused on the structural elucidation and bioactive constituents of triterpenoids in roots,

[1]School of Biological Engineering, Henan University of Technology, No. 100 Lianhua Street, Zhengzhou High-Tech Development Zone, Zhengzhou 450001, Henan, People's Republic of China. [2]China Resources Sanjiu Medical & Pharmaceutical Co., Ltd., Shenzhen 518000, People's Republic of China. [3]Shenzhen Traditional Chinese Medicine Manufacturing, Innovation Center Co., Ltd., Shenzhen 518110, People's Republic of China. [4]Shanghai Origingene Bio-Pharm Technology Co. Ltd., Shanghai 200241, People's Republic of China. [5]Qing Ma and Saidi Wang contributed equally to this work. ✉email: lcw@haut.edu.cn; gaoyangzhang@haut.edu.cn

flavonoids have received comparatively less attention[4,5]. Researchers from Hong Kong, China, assembled a high-quality reference genome of *I. asprella* at the chromosomal level using next-generation sequencing technology[6]. These groups were subsequently subjected to biosynthetic gene profiling of triterpenoid pathway through transcriptomic comparison of two closely related Ilex species[6–8]. Up to date, the molecular basis of flavonoid biosynthesis and regulation in *I. asprella* remains largely unexplored[9].

Flavonoids are an important class of secondary compounds that are widely found in plants. They are involved in plant growth process and those products are often applied in food and medicine. Flavonoids, including anthocyanins (red, orange, blue, and purple pigments), chalcones and aurones (yellow pigments), and flavonols and flavones (white and pale-yellow pigments), which give plants a wide range of colors, are among the main pigments in plants. They are derived from the phenylpropanoid biosynthesis pathway and have a basic structure consisting of a C15 benzene ring structure of C6–C3–C6. Recent studies have shown that flavonoid biosynthesis in plants involves eight branches (stilbene, aurone, flavone, isoflavone, flavonol, phlobaphene, proanthocyanidin, and anthocyanin biosynthesis) and four important intermediate metabolites (chalcone, flavanone, dihydroflavonol, and leucoanthocyanidin)[10]. The biosynthesis of phenylpropanoid needs catalytic cascade of the structural genes, including phenylalanine ammonia lyase (*PAL*), cinnamic acid 4-hydroxylase (*C4H*), and 4-coumarate:CoA ligase (*4CL*)[11–13]. The product of this pathway *p*-coumaroyl-CoA diverges into chalcone and stilbene synthesis pathway, catalyzed by chalcone synthase (*CHS*) and stilbene synthase (*STS*), respectively. The synthetic chalcone is further processed by chalcone reductase (*CHR*), chalcone isomerase (*CHI*), and chalcone 2′-glucosyltransferase (*CH2′GT*) to synthesize various flavonoids. Additionally, chalcone 4′-O-glucosyltransferase (*CH4′GT*) and aureusidin synthase (*AS*) utilize chalcone to produce aurone. Meanwhile, *CHI*, in conjunction with flavanone 3′-hydroxylase (*F3′H*) and flavanone 3′,5′-hydroxylase (*F3′5′H*), catalyzes the conversion of chalcone into flavanones. These flavanones are then transformed into flavones by flavone synthase (*FNS*), flavonoid 6-hydroxylase (*F6H*), cinnamate-CoA ligase (*CLL-7*), *CHS*, *CHI*, and flavonoid 8-hydroxylase (*F8H*) in distinct pathways. Flavanones can also be converted into the isoflavone, phlobaphenes and dihydroflavonol by the action of isoflavone synthase (*IFS*), flavanone 4-reductase (*FNR*) and flavanone 3-hydroxylase (*F3H*) respectively. Further biosynthesis of isoflavone and dihydroflavonol requires the participation of hydroxyisoflavanone dehydratase (*HID*), isoflavanone O-methyltransferase (*IOMT*), *F3′H* and *F3′5′H*. The dihydroflavonol products are then converted into corresponding flavonols and leucoanthocyanidin by flavonol synthase (*FLS*) and dihydroflavonol 4-reductase (*DFR*). Anthocyanidin synthase (*ANS*) and UDP-glucose flavonoid 3-glucosyltransferase (*UFGT*) are also involved in the synthesis of leucoanthocyanidin, anthocyanin, which can be further converted into proanthocyanidins through the catalytic action of leucoanthocyanidin reductase (*LAR*) and anthocyanidin reductase (*ANR*). It is noteworthy that the role of all the aforementioned genes in *I. asprella* remains to be elucidated.

Transcriptional regulation plays a central role in flavonoid biosynthesis. One of the main transcriptional regulators is the MBW complex, which is composed of MYB, bHLH, and WD40 proteins. The MBW complex can act as a promoter or repressor in flavonoid regulation. For example, the R2R3-MYB transcription factor *CsMYB60* in *Cucumis sativus* and *MdMYB15L* in *Malus domestica* have been identified as important regulators[14,15]. In addition to the MBW complex, the NF-Y protein complex, which includes the NF-YA, NF-YB, and NF-YC subunit, also plays a role in flavonoid regulation. The NF-Y protein complex binds the promoter of *CHS* to regulate flavonoid synthesis[16]. Additionally, the ethylene response factors *ERF*, UV-B irradiation-mediated *bZIP1*, *PIF*, *NAC*, *SPL*, *WRKY*, *BT2*, *GRF11*, *BBX20*, *CSN5*, *ARF19*, *LOB52*, and *BES1* selectively interact with *MYB*, its promoter, *bHLH* or flavonoid synthesis genes to modulate flavonoid synthesis[17–24]. Synthetic flavonoids are subsequently transported into vacuoles for storage or to other destinations and this process are reportedly involved in vesicle trafficking, membrane transport, and glutathione S-transferase (GST) activity. The mechanisms by which ATP-binding cassette (*ABC*) transporters, *GSTs*, GREEN FLUORESCENT SEED9 (*GFS9*) and H^+-ATPase function in flavonoid transport have been partly addressed[25].

In this work, the gene transcripts of flavonoid synthesis, regulation, and transport in various tissues of *I. asprella* was investigated. Specifically, the correlation analysis of genes between flavonoid biosynthesis, regulation and transport revealed the gene pairs and regulatory network in flavonoid metabolic pathway. Our findings provided new basis into the production improvement of bioactive compound in different *I. asprella* tissues and offered a piece of valuable evidence for why roots and leaves were usually used in folk medicine.

Results

Transcriptome sequencing and assembly

For the medicinal plant *I. asprella*, a comprehensive de novo transcriptome assembly was executed utilizing DNBSEQ-T7 sequencing technology. This approach targeted mRNA from leaf, stem, and root tissues, yielding a substantial dataset (Fig. 1). The sequencing process generated a robust range of 43.69 to 45.44 million raw reads per tissue. Post-quality control, approximately 42.44 million high-quality reads, equating to an average of 6.39 Gb of clean nucleotides, was retained (Table 1). These reads were assembled into 40,129 unigenes (Table S1). The subsequent alignment of samples to the reference genome revealed an average match rate of 93.12%, while the gene set alignment averaged at 72.27%. The analysis predicted 914 novel genes. A total of 28,602 expressed genes were detected, comprising 27,691 annotated genes and 911 predicted new genes. Additionally, 18,536 new transcripts were detected, with 17,622 being novel alternative splicing variants of known protein-coding genes and 914 corresponding to new protein-coding genes (Table 1; Table S1).

Functional annotation

Annotation of the assembled transcriptome was performed using the Basic Local Alignment Search Tool (BLAST), employing a stringent e-value threshold of less than 10^{-5}, to align with the NCBI nonredundant (Nr), Gene Ontology (GO), and Encyclopedia of Genes and Genomes (KEGG) databases. The outcomes of

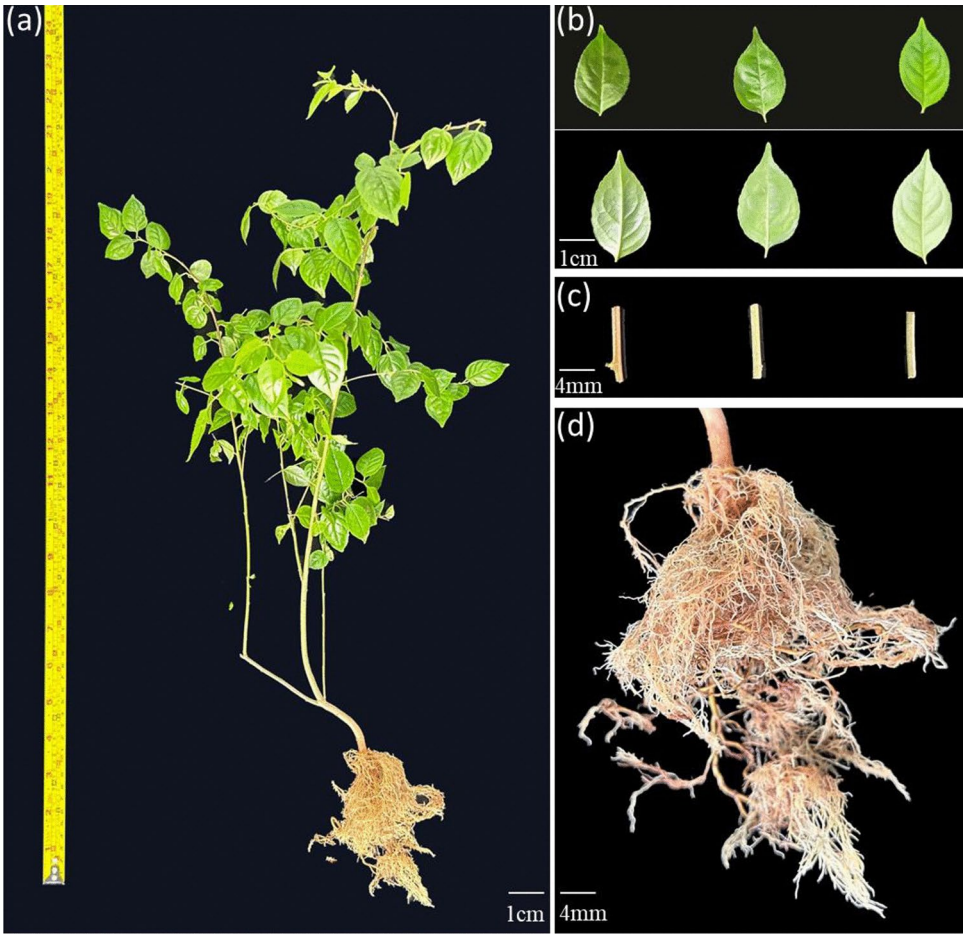


Fig. 1. Plant samples of *Ilex asprella* used in the study. (**a**) Whole plant. (**b**) Leaves. The upper represents the pressure side of blade and the lower represents opposite side. (**c**) Stem. (**d**) Root samples.

Sample	Leaf	Root	Stem
Total raw reads (M)	45.44	43.69	43.69
Total clean reads (M)	42.82	42.35	42.62
Total clean bases (Gb)	6.42	6.35	6.39
Total mapping (%) (refer to genome)	94.01	91.75	93.60
Uniquely mapping (%) (refer to genome)	92.16	89.63	91.57
Total mapping (%) (refer to gene)	73.34	71.30	72.17
Uniquely mapping (%) (refer to gene)	64.09	59.16	62.53
Total alternative splicing	15,959	18,028	17,502

Table 1. Summary of the transcriptome assembly of *Ilex asprella*.

this annotation effort were delineated in Table 2, indicating that 36,683 unigenes (91.41% of the total 40,129) were annotated with information from these databases. Specifically, 36,631 (91.28%), 28,454 (70.91%), and 26,599 (66.28%) unigenes were successfully annotated against the Nr, GO, and KEGG databases, respectively (Tables S2–S4). Functional predictions were based on the most closely related annotated sequences within these databases. The observed lower annotation rates may be attributed to the scarcity of genomic information for the *Asprella* genus in the public databases.

Gene expression

A total of 28,603 genes were detected with expressed levels quantified using fragments per kilobase of transcript per million mapped reads (FPKM), with values spanning from 0.000 to 22,959.99, which underscored high sensitivity of detection (Table S5). A correlation heatmap was employed to visually evaluate the variability in global gene expression across distinct tissue samples. The Pearson correlation coefficients indicated strong

Databases	Annotated number	Ratio
Annotated with Nr	36,631	91.28%
Annotated with GO	28,454	70.91%
Annotated with KEGG	26,599	66.28%
All annotated unigenes	36,683	91.41%

Table 2. Functional annotation statistics of *Ilex asprella* unigenes.

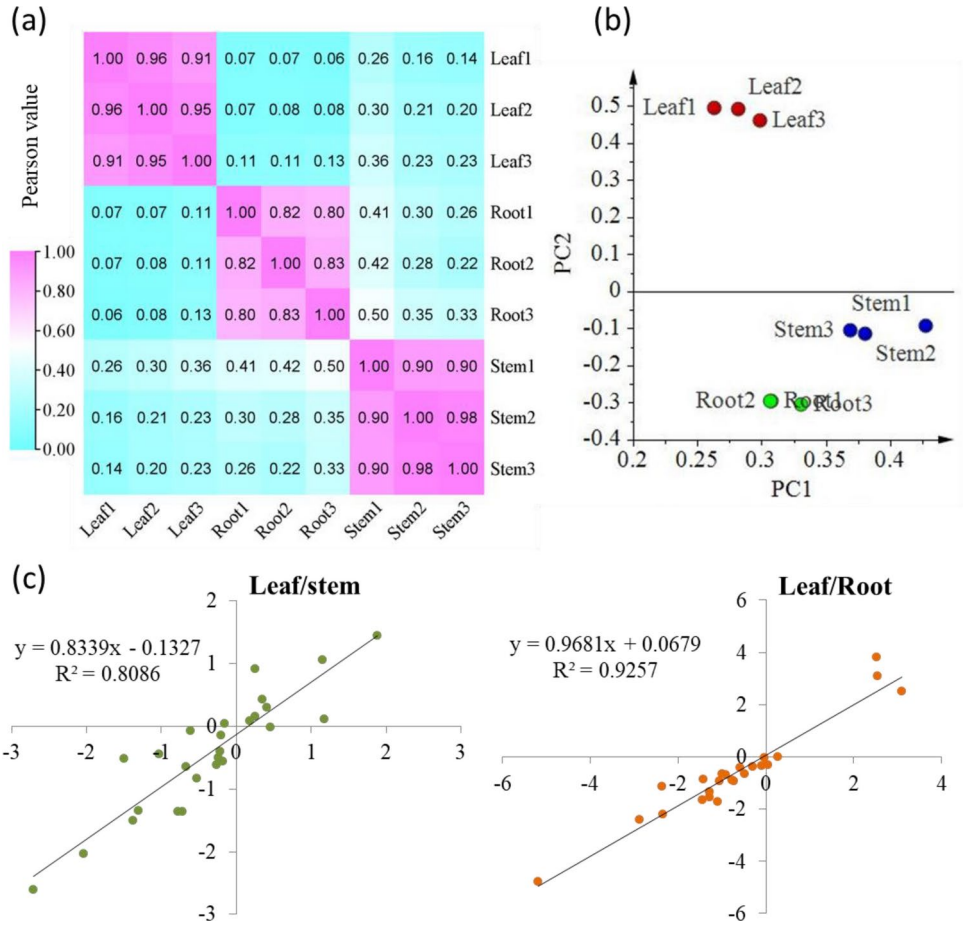


Fig. 2. Validity analysis of transcriptome in leaf1-3, root1-3 and stem1-3 samples. (**a**) Correlation heatmap of leaf1-3, root1-3 and stem1-3 samples. (**b**) Principal component analysis of leaf1-3, root1-3 and stem1-3 samples. (**c**) Correlation analysis of gene expression levels between real-time quantitative PCR (RT-qPCR) and transcriptome data for selected *Ilex asprella* genes. Each colored dot represents an expression level-based log(fold-change, 5) value from the expression ratio of two tissues.

positive correlations among tissues of the same type, while weak correlations were observed between stem, leaf, and root tissues (Fig. 2a). Principal component analysis, conducted on the FPKM values of the samples, demonstrated that the biological replicates grouped closely together, affirming the robustness of our RNA-sequencing data (Fig. 2b). To further verify the quality of our dataset, 25 differentially expressed genes across leaf, stem, and root tissues were subjected for quantitative real-time RT-qPCR. The relative expression levels of these 25 genes, as determined by transcriptome analysis, closely mirrored those derived from qPCR. The positive correlations observed ($R^2 = 0.8086$ for Leaf/stem and 0.9257 for Leaf/root) further substantiated the accuracy of gene expression estimates based on RNA sequencing (Fig. 2c). Consequently, RNA sequencing data was deemed reliable for subsequent gene expression analyses in various tissues.

Differentially expressed genes (DEGs)

Comprehensive gene expression analyses across different tissues revealed a multitude of DEGs, offering a comparative overview of gene regulation. Employing stringent criteria of an FPKM value > 1, a false discovery rate (FDR) < 0.01, and an absolute log2 fold change (|log2 (FC)|) ≥ 1, a total of 28,478 DEGs were identified among leaf, root, and stem samples. Pairwise comparisons of leaf versus stem, leaf versus root, and stem versus

root uncovered 7264 (3589 downregulated and 3675 upregulated), 11,747 (5284 downregulated and 6463 upregulated), and 9467 (4354 downregulated and 5113 upregulated) DEGs, respectively (Fig. 3). Additionally, 32,865, 28,382, and 30,662 consistently expressed genes were identified in the comparisons of leaf to stem, leaf to root, and stem to root, respectively (Fig. 3). Venn diagrams were utilized to depict the distributions and potential interrelations of DEGs across paired comparisons, highlighting 2355 DEGs that were commonly altered (Fig. 3d). Furthermore, 720, 1857, and 1418 DEGs were uniquely identified in the leaf versus stem, leaf versus root, and stem versus root comparisons, respectively.

To elucidate the biological roles of these DEGs, GO and KEGG classification and enrichment analyses were conducted. The GO categories "biological process", "cellular component", and "molecular function", which pertained to fundamental plant functions, were found to be significantly enriched in the pairwise comparisons. The DEGs across the three tissues were also notably enriched for the terms "cellular anatomical entity", and "catalytic activity" as per the GO classification analysis (Fig. S1). DEGs in leaf versus stem, leaf versus root and stem versus root were enriched in GO terms associated with "chloroplast" and "plastid" terms. Additionally, leaf versus root and stem versus root DEGs were also enriched in "tetrapyrrole binding" and "DNA-binding transcription factor activity", respectively, underscoring the intricate relationships among these tissues (Fig. S2).

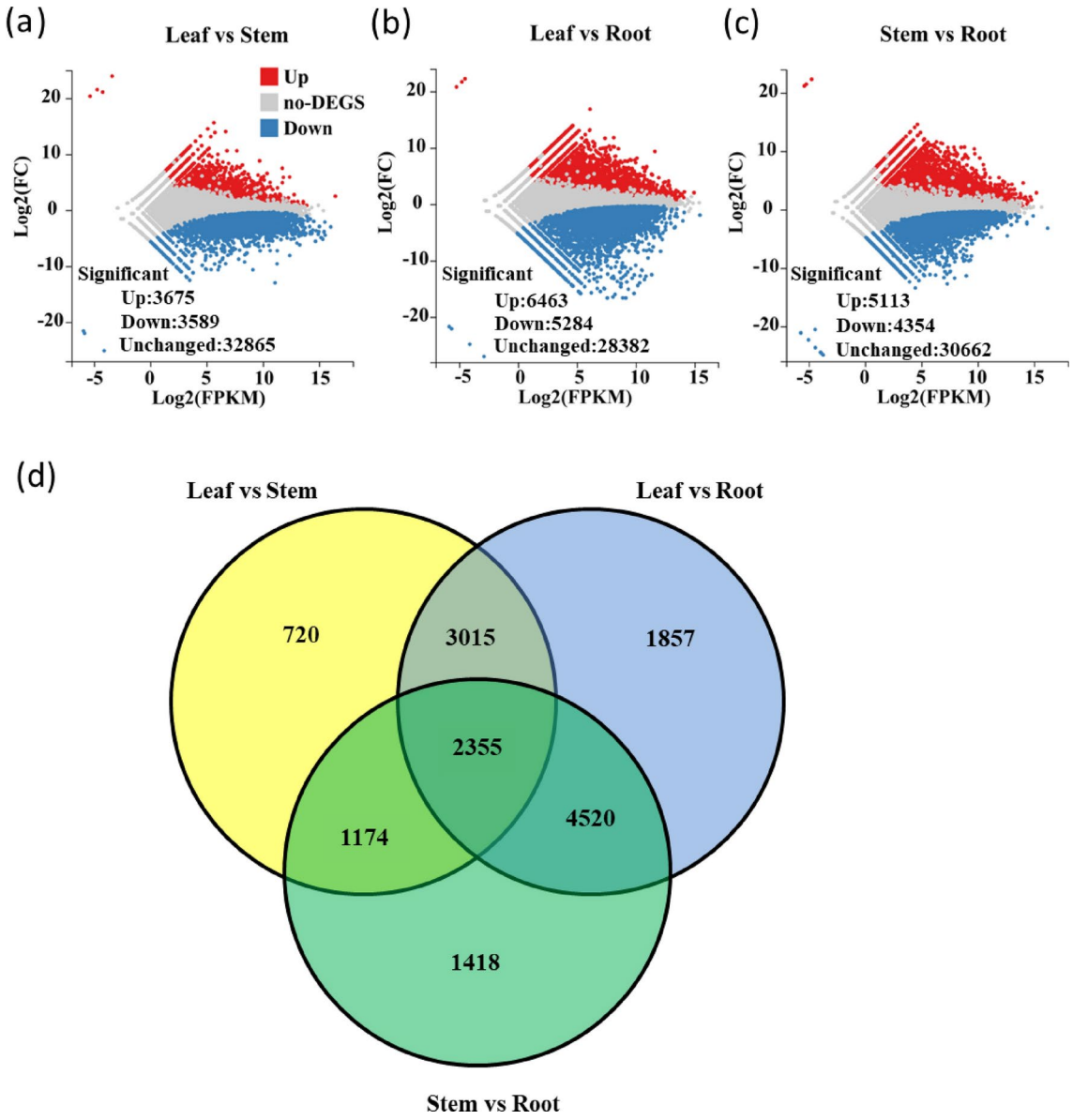


Fig. 3. MA plots of transcriptome differences and Venn diagram of DEGs between different *Ilex asprella* tissues. (**a**) Leaf vs Stem, (**b**) Leaf vs Root, and (**c**) Stem vs Root. Differentially expressed genes (DEGs) of *Ilex asprella* were identified using the criteria FDR ≤ 0.01 and |log 2 (FC)| ≥ 1. Grey dots represent gene expression that was not significantly different in the comparisons, red dots represent up-regulated genes, and blue dots represent down-regulated genes. (**d**) Venn diagram of DEGs identified in different Ilex asprella tissues. The common gene numbers are in overlapping regions by different comparisons.

KEGG pathway classification indicated DEGs were enriched in "Global and overview maps", "Carbohydrate metabolism", "Signal transduction", and "Folding, sorting and degradation" reflecting the metabolism, environmental information processing and genetic information functions of the leaf, stem and root (Fig. S3).

The DEGs were found to be enriched in 140 KEGG pathways, with the top 20 significant pathways in each pairwise comparison depicted in Fig. 4. The global maps of "plant hormone signaling" and "MAPK signaling pathway—plant" were consistently enriched across the three tissue comparisons suggesting that genes related to these pathways were expressed in all three tissues, albeit at varying levels. The leaf-root comparison uniquely identified more DEGs than the other two comparisons, such as those involved in "plant-pathogen interaction" (Fig. 4b). Moreover, the plant-pathogen interaction pathway was also enriched in the stem versus root comparison (Fig. 4c).

Flavonoid biosynthesis genes in *I. asprella*

In the canonical phenylpropanoid pathway, the aromatic amino acid phenylalanine, is converted to *p*-coumaroyl-CoA by the sequential action of PAL, C4H, and 4CL (Fig. 5a). Our study identified five *PALs*, 13 *C4Hs*, and 13 *4CLs* as unigenes, which were further validated by phylogenetic tree analysis (Table S6; Fig. S4). Except for *PAL (Ilex_030935)*, *C4H (Ilex_024374, Ilex_042319)* and *4CL (Ilex_019705, Ilex_026016)*, all identified *PALs*, *C4Hs*, and *4CLs* were expressed across the three tissues examined. Notably, certain genes, including *PAL (Ilex_042231)*, *C4H (Ilex_017598)* and *4CL (Ilex_042033)*, exhibited at high expression levels specifically in the roots (Table S6; Fig. 5bc). Additionally, among their respective isoforms, *PAL (Ilex_014816)*, *C4H (Ilex_017598)* and *4CL (Ilex_023165)* were found to be the most highly expressed in leaves (Fig. 5; Table S6).

Chalcone, the initial key intermediate in flavonoid biosynthesis, is synthesized through the action of acetyl-CoA carboxylase (ACCase), which produces malonyl-CoA, and CHS, responsible for the formation of naringenin chalcone (4,20,40,60-tetrahydroxychalcone [THC] [chalcone]). CHR further processes a CHS reaction intermediate to form isoliquiritigenin, while CHI converts THC into colorless naringenin or isomerizes it to isosalipurposide (ISP) through the action of chalcone 2′-glucosyltransferase (CH2′GT) (Fig. 5a). Eleven *ACCases* and three *CHSs* were annotated and confirmed, with the latter being expressed at very low levels across the tissues examined (Fig. 5bc; Fig. S5; Table S6). The majority of *ACCases* showed intermediate expression levels, with the exception of *ACCases (Ilex_042981, Ilex_047561, and Ilex_008769)*, which were either expressed at low levels or undetected in the samples (Fig. 5b; Fig. S6; Table S6). Furthermore, ten *CHR*, two *CHI*, and fifty-one *CH2′GT* genes were detected. Notably, *CHR (Ilex_020521)* exhibited higher expression in roots compared to the other *CHRs*, while *CHR (Ilex_046608 and Ilex_020514)* seemed to be specifically expressed in leaves. Among the *CHI* genes, only one *CHI(Ilex_005941)* showed high expression in leaves, whereas the other *CHI(Ilex_021868)* was barely detectable except in the stem. The majority of *CH2′GTs* were expressed at low levels or undetected in the tissues, with the exception of *CH2′GTs (Ilex_012606, Ilex_049682*, etc.), which were more highly expressed in roots than in stems and leaves (Fig. 5b; Fig. S6; Table S6).

STS, which also utilizes *p*-coumaroyl-CoA and malonyl-CoA as substrates, catalyzes the formation of the stilbene backbone, exemplified by resveratrol, marking the initial divergence of the flavonoid biosynthesis pathway (Fig. 5a). Only two *STSs (Ilex_044726 and Ilex_048273)* were detected, with low expression was across all tissues (Fig. 5b; Fig. S5; Table S6). In the aurone biosynthesis pathway, which yields a vibrant yellow pigment, CH4′GT catalyzes the formation of THC 4′-O-glucoside from THC, and AS transforms THC 4′-O-glucoside into aureusidin 6-O-glucoside (aurone) (Fig. 5a). Nine *ASs* and fourteen *CH4′GTs* were annotated, exhibiting differential expression in tissues (Fig. 5b; Fig. S7; Table S6). A majority of these genes exhibited preferential expression in roots, such as *AS (Ilex_009329)* and *CH4′GT (Ilex_047989)*. Flavanone, the central node in the flavonoid biosynthetic cascade, is formed through the action of CHI, which catalyzes the intramolecular cyclization of chalcones to produce flavanones; these can be converted to eriodictyol and pentahydroxyflavanone (two types of flavanones) based on activity of F3′H and F3′5′H (Fig. 5a). A single *F3′H (Ilex_006685)* was identified, with higher expression in the stem compared to the other two tissues. However, 31 *F3′5′Hs* were annotated, with most showing differential expression across the three tissues; notably, *F3′5′H (Ilex_011771)* exhibited an average 45-fold increase in expression in roots compared to leaves (Fig. 5b; Fig. S8; Table S6).

Flavones, such as naringenin, are derived from flavanones through the action of FNS. An alternative flavone synthesis pathway, commencing from cinnamic acid and leading to baicalein/norwogonin, involves CLL-7, CHS, CHI, FNS, F6H and F8H (Fig. 5a). Four *FNSs*, eight *CLL-7s*, fifty-two *F6Hs*, and four *F8Hs* were detected, with *FNS (Ilex_043640)* showing significantly higher expression in roots compared to leaves and stems. The expression of nearly all *CLL-7* genes was consistent across tissues, with the exception of *CLL-7 (Ilex_028327, Ilex_031706)*. Furthermore, most *F6H* genes displayed differential expression in all three tissues, particularly *F6H (Ilex_045743)* and *F6H (Ilex_036389)*, which were highly expressed in the roots and leaves, respectively. Similar patterns were observed for *F8H (Ilex_044784)* and *F8H (Ilex_037622)* (Fig. 5b; Fig. S9; Table S6).

IFS directs flavanone towards the isoflavone pathway, producing 2-hydroxy-2,3-dihydrogenistein and 2,7,40-trihydroxyisoflavanone, a process that requires the participation of F6H, IFS, HID, and IOMT (Fig. 5a). Transcripts for four *IFSs*, eight *HIDs*, and six *IOMTs* were identified, with *IFS (Ilex_029360)* and *IOMT (Ilex_015111)* showing higher expression in stems compared to leaves and roots. However, *HID (Ilex_044559)* exhibited greater expression in roots than in leaves or stems (Fig. 5b; Fig. S10; Table S6). FNR acts on flavanones (naringenin and eriodictyol) to produce the corresponding flanvan-4-ols (apiforol and luteoforol), which can then be further polymerized to form phlobaphenes (Fig. 5a). Only one *FNR (Ilex_039777)* was detected, with low transcript across all three tissues. Phylogenetic analysis revealed that FNR shared functions similarities with ANR, which converted anthocyanidins, pelargonidin, cyanidin, and delphinidin to the corresponding cis-flavan-3-ols, epiafzelechin, epicatechin, and epigallocatechin (Fig. 5b; Fig. S11; Table S6).

F3H converts naringenin, eriodictyol, and pentahydroxyflavanone into their respective dihydroflavonols: dihydrokaempferol (DHK), dihydroquercetin (DHQ), and dihydromyricetin (DHM). This metabolic cascade

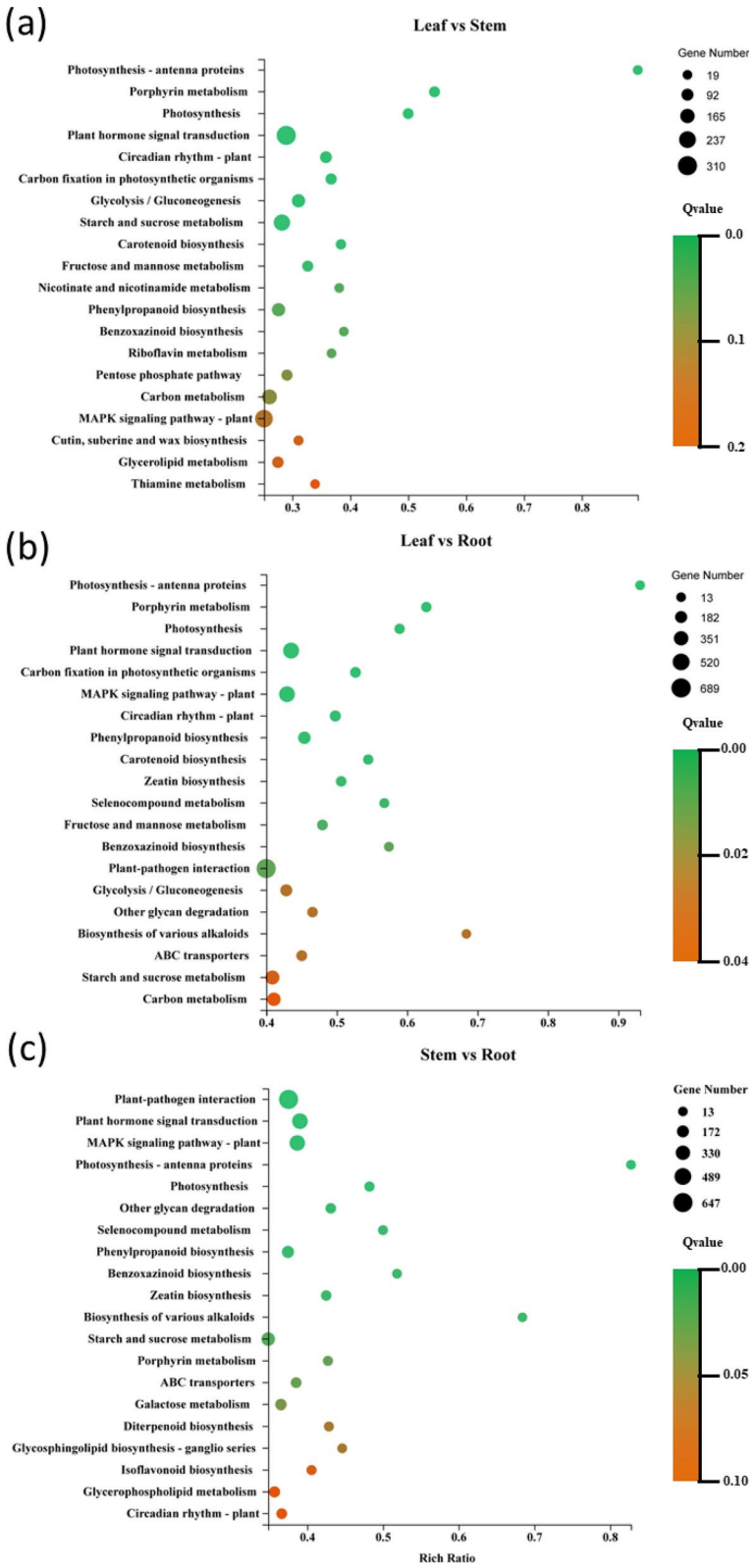


Fig. 4. Scatterplot of differentially expressed *Ilex asprella* genes in the top 20 enriched KEGG pathways. (**a**) Leaf vs Stem, (**b**) Leaf vs Root, and (**c**) Stem vs Root. The y-axis shows the KEGG pathways, and the x-axis shows the enrichment factor. The enrichment factor is the ratio of the numbers of DEGs annotated in a certain pathway to the total number of genes mapped to this pathway. The q-value represents the corrected P value. A higher enrichment factor value correlates with a more intensive pathway, and a lower q-value with a more reliable one.

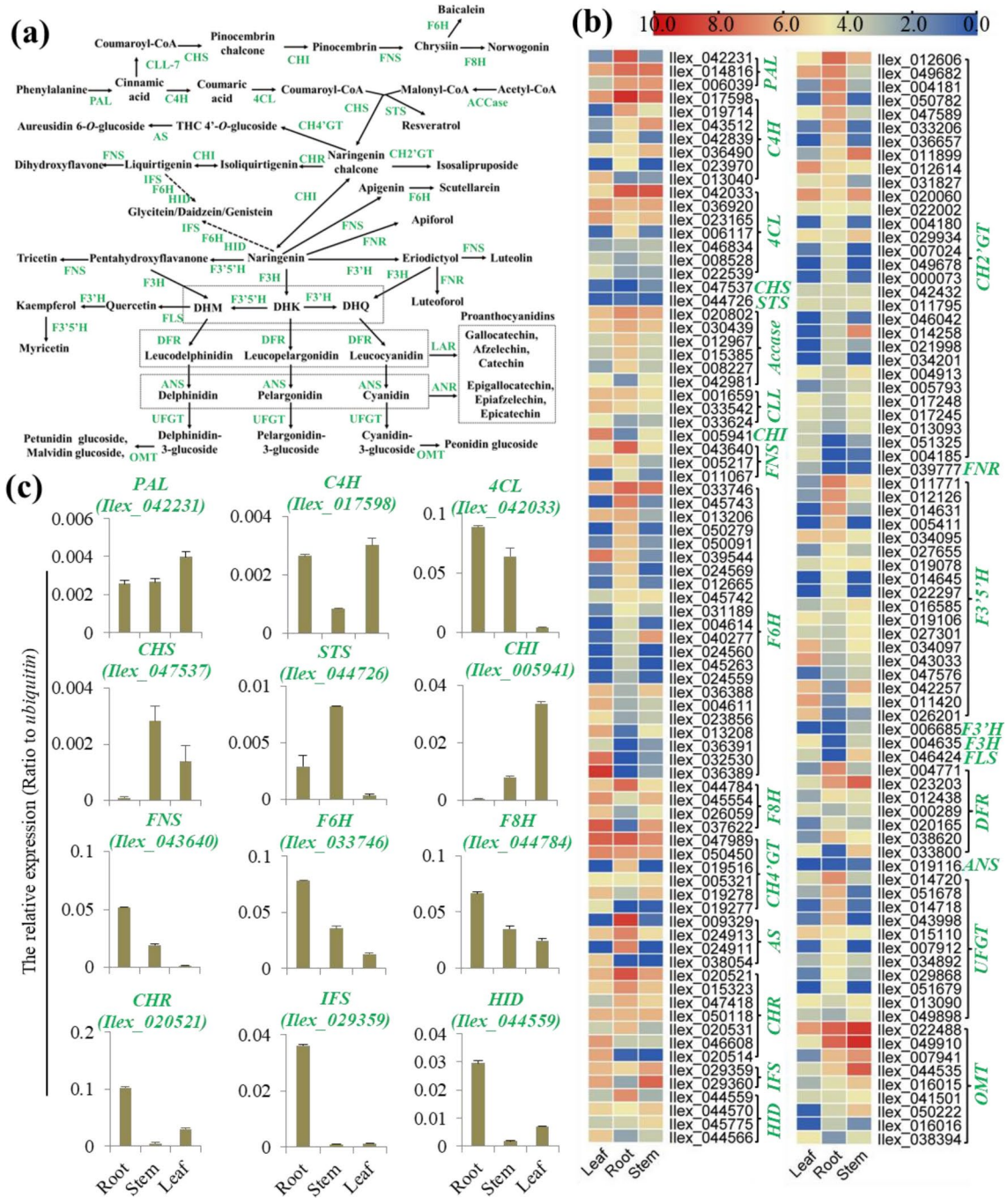


Fig. 5. The expression of key biosynthesis genes involved in flavonoid biosynthetic pathway in *Ilex asprella* tissues. (**a**) The possible flavonoid biosynthetic pathway in *Ilex asprella*. (**b**) Expression analysis of flavonoid biosynthetic pathway in leaves, roots, and stems tissues of *Ilex asprella*. Expression values are presented as log2 (FPKM). (**c**) qRT-PCR based expression analysis of a few biosynthesis genes. Data are shown as representative means ± SD from three independent experiments. The enzyme names and flavonoid compounds are abbreviated as follows: *PAL* phenylalanine ammonia lyase, *C4H* cinnamic acid 4-hydroxylase, *4CL* 4-coumarate: CoA ligase, *CHS* chalcone synthase, *STS* stilbene synthase, *ACCase* acetyl-CoA carboxylase, *CLL* cinnamate–CoA ligase, *CHI* chalcone isomerase, *FNS* flavone synthase, *F6H* flavonoid 6-hydroxylase, *F8H* flavonoid 8-hydroxylase, *CH4′GT* chalcone 4′-O-glucosyltransferase, *AS* aureusidin synthase, *CHR* chalcone reductase, *IFS* isoflavone synthase, *HID* 2-hydroxyisoflavanone dehydratase, *CH2′GT* chalcone 2′-glucosyltransferase, *FNR* flavanone 4-reductase, *F3′5′H* flavanone 3′,5′-hydroxylase, *F3′H* flavanone 3′-hydroxylase, *F3H* flavanone 3-hydroxylase, *FLS* flavonol synthase, *DFR* dihydroflavonol 4-reductase, *ANS* anthocyanidin synthase, *UFGT* UDP-glucose flavonoid 3-O-glucosyltransferase, *OMT* O-methyl transferases, *LAR* leucoanthocyanidin reductase, *ANR* anthocyanidin reductase.

is further advanced by the action of flavanone 3′-hydroxylase (F3′H) and flavanone 3′,5′-hydroxylase (F3′5′H), which facilitate the transformation of DHK into DHQ and DHQ into DHM, respectively. The dihydroflavonols are then synthesized into the flavonols kaempferol, quercetin, and myricetin by the catalytic activity of FLS, coupling with F3′H/F3′5′H (Fig. 5a). A transcript of *F3H (Ilex_004635)*, which exhibited modestly elevated expression in leaves, was identified. Among the four *FLS* transcripts detected, *FLS (Ilex_046424)* was found to have increased expression in leaves and stems relative to roots or other forms (Fig. 5b; Figs. S12, S13; Table S6). DFR is pivotal in the reduction of dihydroflavonols to leucoanthocyanidins including leucopelargonidin, leucocyanidin, and leucodelphinidin. These colorless leucoanthocyanidins are converted into anthocyanidins—pigments responsible for the colorful hues observed in plants—by ANS. The stability of anthocyanidins is enhanced through glucosylation by UFGT, yielding stable anthocyanins such as pelargonidin-3-glucoside, cyanidin-3-glucoside, and delphinidin-3-glucoside. OMT further modifies these anthocyanins, converting cyanidin-3-glucoside into peonidin glycoside and delpinidin-3-glucoside into petunidin glycoside or malvidin glycoside (Fig. 5a). Transcripts for *DFR*, *ANS*, *UFGT*, and *OMT* were quantified, with eight, one, thirty-three, and thirteen copies detected, respectively. Notably, *DFR (Ilex_004771)* and *UFGT (Ilex_014720)* exhibited heightened expression in roots, whereas *OMT (Ilex_022488)* was predominantly expressed in the stem. Conversely, *ANS (Ilex_019116)* demonstrated minimal expression across all three tissues (Fig. 5b; Fig. S14; Table S6). LAR is tasked with conversion of leucoanthocyanidins, leucopelargonidin, leucocyanidin, and leucodelphinidin to trans-flavan-3-ols, afzelechin, catechin, and gallocatechin, respectively (Fig. 6a). However, our analysis identified only one *LAR (Ilex_020124)*, which was not expressed in the tissues examined (Fig. S15; Table S6). Collectively, these findings underscored the diversity of flavonoid biosynthesis genes, many of which exhibited tissue-specific expression patterns and significant differential expression levels.

Transcriptional regulation genes of flavonoid biosynthesis in *I. asprella*

Our analysis revealed the presence of 93 *MYB*, 36 *bHLH*, and 240 *WD40* transcripts, which constituted the multifaceted MBW complex. The majority of these transcripts were detected at low expression levels across the tissues examined. However, selected genes, including *MYB (Ilex_004399)*, *bHLH (Ilex_001378)*, and *WD40 (Ilex_046093)*, demonstrated notably higher expression in the three tissues. Furthermore, A subset of MBW genes, such as *MYB (Ilex_036220, Ilex_001714)*, *bHLH (Ilex_011586)*, and *WD40 (Ilex_012309)*, exhibited enhanced expression in roots compared to leaves and stems (Fig. 6; Table S6). Additionally, 4, 11, and 6 transcripts of *NF-YA*, *NF-YB*, and *NF-YC*, respectively, were annotated. While most were expressed at low levels, *NF-YA (Ilex_012157)*, *NF-YB (Ilex_038767)*, and *NF-YC (Ilex_012555)* stood out with high expression levels in roots, leaves, and leaves, respectively (Fig. 6; Table S6). In addition, 21 *BBX* transcripts and 1 *CSN* transcript were detected, and most *BBXs*, especially *BBX (Ilex_050110, Ilex_010644, Ilex_018168)*, were differentially expressed in the three tissues. Moreover, *CSN (Ilex_010585)* was generally highly expressed, particularly in stems (Fig. 6; Table S6). A total of 36 *ERF* and 93 *WRKY* transcripts were detected, with their expression levels generally surpassing those of *CSN* or *BBX* in the tissues analyzed. Notably, *ERF (Ilex_028143, Ilex_005549)* and *WRKY (Ilex_023201)*, exhibited exceptionally expression high in roots (Fig. 6; Table S6). Four *bZIP*, 12 *NAC*, and 9 *SPL* transcripts were detected across all tissues, with *bZIP (Ilex_016445)*, *NAC (Ilex_030072)*, and *SPL (Ilex_045464)* showing increased expression in roots relative to stems and leaves (Fig. 6; Table S6). Two *PIFs* transcripts were detected, with *PIF (Ilex_036297)* in leaves showing an average fivefold higher expression compared to stems (Fig. 6; Table S6). Our predictions identified 26 *ARF*, 7 *LOB*, 2 *BES*, 14 *GRF*, and 4 *BT* transcripts, with most exhibiting differential expression across the three tissues. *ARF (Ilex_021746)*, *GRF (Ilex_002291)*, and *BT (Ilex_007791)* were particularly more highly expressed in roots. The two genes in the *BES (Ilex_033872 and Ilex_029033)* were consistently and highly expressed, while all *LOBs* showed extremely low expression across tissues but a preference for higher transcript levels in roots (Fig. 6; Table S6). Furthermore, the transcript levels of the suppressor genes were generally lower compared to those of the promoter genes for the transcriptional regulation of flavonoid biosynthesis (Fig. 6; Table S6).

Flavonoid transport genes in *I. asprella*

A total of two *GFSs*, 36 *Mates*, 20 *GSTs*, 52 *ABCCs*, and 15 *H*$^+$*-ATPases* were identified, potentially playing roles in flavonoid transport. Notably, *GFS (Ilex_042069, Ilex_050992)* exhibited no significant differential expression across the three tissues examined (Fig. 6d; Table S6). In contrast, the expression levels of *Mate (Ilex_038263, Ilex_029131)*, *GST (Ilex_036047, Ilex_023908)* and *ABCC (Ilex_037010)* were markedly higher in roots compared to other tissues. Additionally, *H*$^+$*-ATPase* genes, including *H*$^+$*-ATPase (Ilex_034304)*, generally showed lower expression levels than *GST* or *ABCC* in the tissues analyzed (Fig. 6d; Table S6). The remaining flavonoid transport genes were found to be either minimally expressed or not significantly differentially expressed.

Coexpression network analysis of flavonoid biosynthesis, regulation and transport

To elucidate the interplay between genes involved in flavonoid biosynthesis and those regulating transport, a Pearson correlation analysis was constructed. The analysis revealed significant positive or negative correlations among related genes (Fig. 7a; Figs. S16, S17). By filtering the data for Pearson correlation coefficient ≤ 0.001, several gene pairs were highly significant correlated, such as *4CL (Ilex_006117)/NF-Yb (Ilex_033014)*. In general, transporters, including *ABC*, *GST*, and *Mate*, along with regulators like *BBX*, *bHLH*, *ERF*, *LOB*, *MYB*, *NAC*, *NF-Yb*, *NF-Yc*, *WD40*, and *WRKY*, all demonstrated close associations with biosynthetic synthesis genes (Fig. 7a). Furthermore, flavonoid biosynthesis genes, such as *CHR (Ilex_005521)* with *MYB (Ilex_019318)*, *WD40 (Ilex_023245)*, and *ABC (Ilex_026902)*, exhibited multiple highly significant correlations with regulatory genes (Fig. 7a). It was evident that a single regulator or transporter could be closely linked to one or more flavonoid biosynthesis genes, exemplified by *Mate (Ilex_009191)* correlating with *I2H (Ilex_042316)*; *F6H (Ilex_038159)* and *MYB (Ilex_008490)* correlating with *UFGT (Ilex_051679)*; *STS (Ilex_048273)*; *HID (Ilex_015631)*; *F6H*

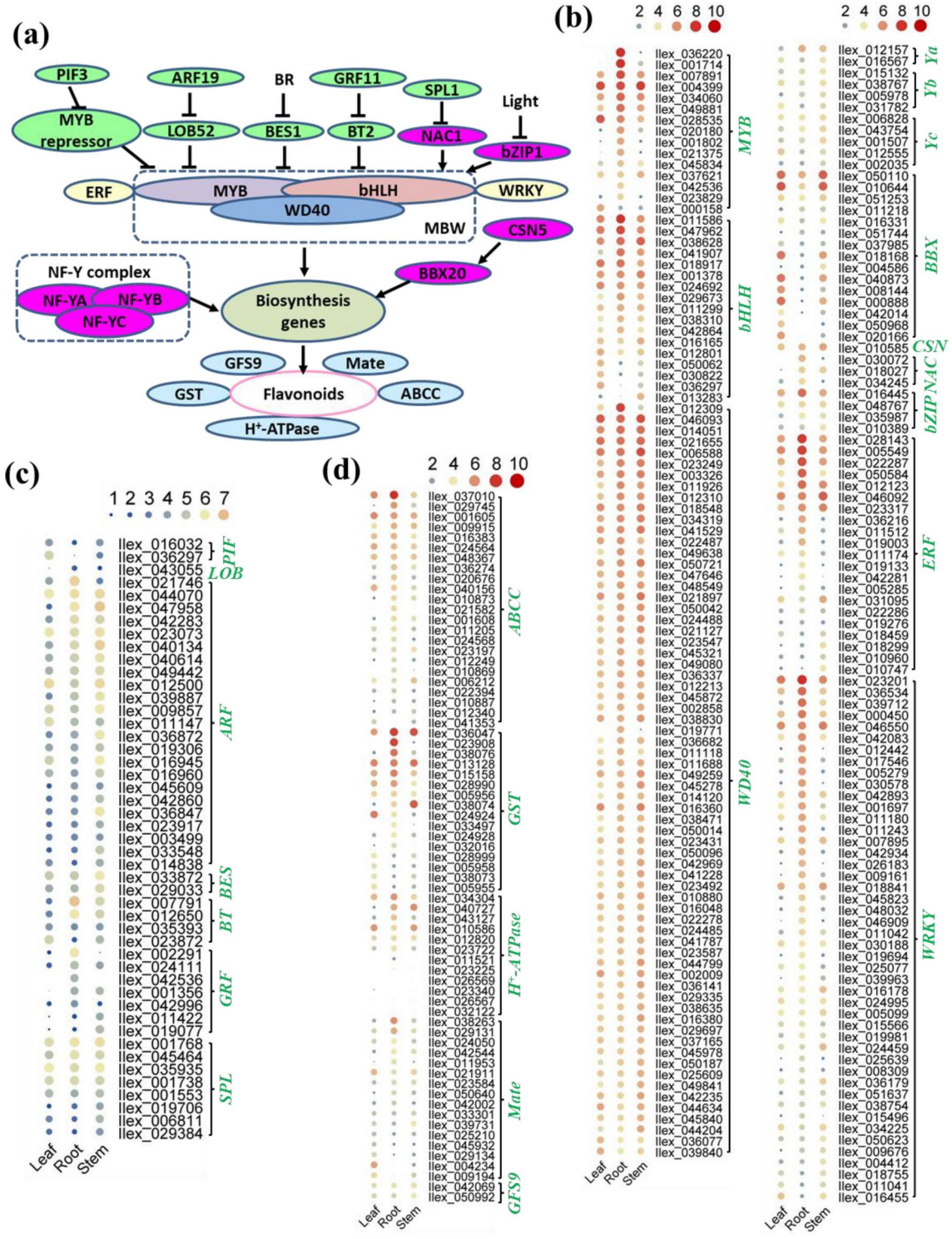


Fig. 6. The expression of transcriptional regulator genes possibly involved in the regulation of flavonoid biosynthetic pathway in *Ilex asprella* tissues. (**a**) The possible transcriptional regulation of flavonoid biosynthesis in *Ilex asprella*. (**b**) The expression of possible genes promoting flavonoid synthesis. (**c**) The expression of possible genes suppressing flavonoid synthesis. (**d**) The expression of possible genes involved in flavonoid transport. *MYBv*-myb avian myeloblastosis viral oncogene homolog, *bHLH* basic helix–loop–helix, *NF-Y* nuclear factor Y, *ERF* ethylene responsefactor, *NAC* (NAM, ATAF, CUC), *SPL* squamosa promoter binding protein-like, *GRF* growth regulating factor, *BT* BTB/TAZ, *BBX* b-box protein, *ARF* auxin response factor, *LOB* lateral organ boundaries, *BES1* BRI1-EMS-SUPPRESSOR 1, *BR* brassinosteroid, *ABC* ATP-binding cassette transporters, *GFS9* GREEN FLUORESCENT SEED9, *GST* glutathione S-transferase, *MATE* toxin extrusion. The blue dashed box represents the protein complex: MBW complex is constituted of three class of transcription factors (TFs), MYB, bHLH and WD40, while NF-Y complex is composed of TFs NF-YA, NF-YB, and NF-YC. TFs next to each other represent interaction of proteins.

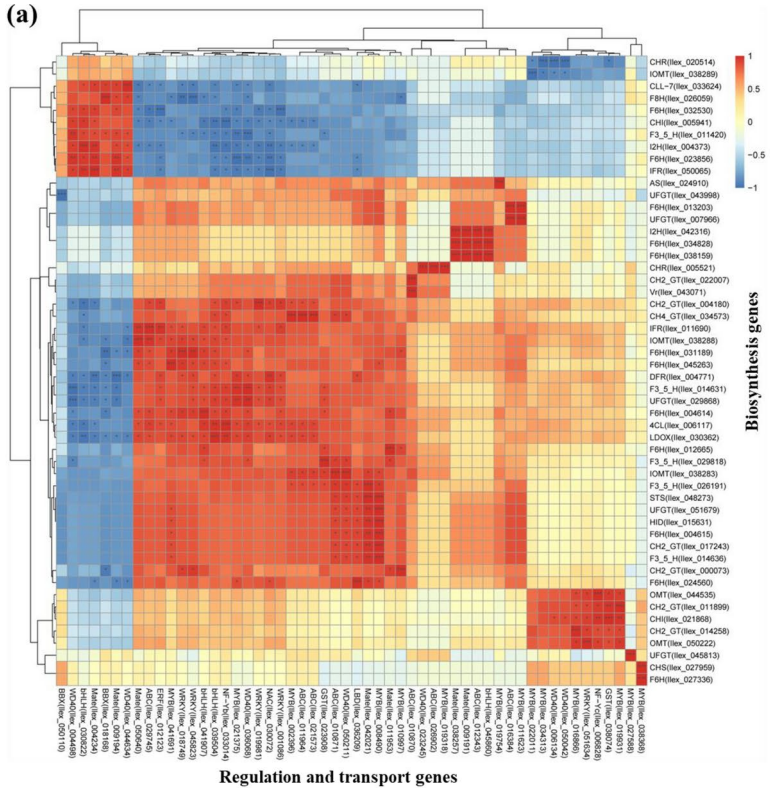
(Ilex_004615); *F3′5′H (Ilex_014636)*; and *CH2′GT (Ilex_000073)* (Fig. 7a). Further transcriptional regulatory network analysis indicated that 74, 575, 7220, and 221,891 of the transporter–gene pairs were highly significant, significant, different and not different, respectively (Fig. 7b; Fig. S18). Notably, some transporter-gene pairs, such as *MYB (Ilex_038368)-CHS (Ilex_027959)*, had a correlation coefficient equal to one, and p_adjust ≤ 0.001 (Table S7; Fig. S18).

Discussion

Functional annotation of the transcriptome data indicated that 8.59% of the unigenes remained unannotated, suggesting their potential uniqueness to *Ilex asprella* and highlighting their value for novel gene discovery. These unannotated sequences might also represent untranslated regions or noncoding RNAs, a phenomenon observed in other plant transcriptomes such as those of tomato during flowering reversion and *Sorghum bicolor* in response to *Sporisorium reilianum* colonization[26,27]. KEGG analysis of *Ilex asprella* tissues revealed an enrichment of DEGs related to "plant hormone signal transduction" and "MAPK signaling pathway". This contrasted with the enrichment of flavonoid synthesis DEGs in the leaves, stems, and roots of *Scutellaria viscidula*, and carotenoid biosynthesis DEGs in the leaves and roots of *Daucus carota*. These findings underscored the tissue-specific biosynthesis of secondary metabolites and reflected metabolic diversification during plant evolution, highlighting the distinctiveness of medicinal plants[28,29].

Our analysis identified a variety of flavonoid biosynthetic pathways in *I. asprella,* leading to the production of 12 flavonoids subgroups: chalcones, stilbenes, aurones, flavanones, flavones, isoflavones, phlobaphenes, dihydroflavonols, flavonols, leucoanthocyanidins, proanthocyanidins, and anthocyanins. Similar pathways have been reported in other plant species, such as legumes, sorghum, maize, *Gloxinia*, grape, peanut, and pine[10,30,31]. The expression patterns of these biosynthetic genes varied across leaf, stem, and root tissues (Fig. 6). The phenylpropanoid pathway genes in *I. asprella* formed a multigene family, as seen in other plants[32–34]. Tissue-specific expression differences were observed for PAL, C4H, and 4CL transcripts, as reported in *Allium sativum* and *Polygonum minus*[35,36]. The first key intermediate metabolite, chalcone, and the initial step of stilbene biosynthesis were characterized by the identification of three CHS, two STS, and ten CHR genes, showing high similarity to published sequences[37–39]. The low expression of CHS in *I. asprella* tissues suggested its role as rate-limiting enzyme in flavonoid biosynthesis. STS, competing with CHS for substrates, showed minimal differences in phylogenetic analysis and low tissue expression, distinct from that observed in *Polygonum cuspidatum*[40]. The bright yellow pigment aurone is synthesized by a limited number of species, such as snapdragon, sunflowers, and coreopsis, including those with 14 CH4′GTs, and 9 ASs in *I. asprella*[41,42]. CH4′GTs and ASs exhibited higher expression than CHS or STS, confirming the rate-limiting function. Two *CHIs*, pivotal for flavanone synthesis and central to the flavonoid pathway *I. asprella,* were the second key rate-limiting enzyme in the flavonoid biosynthetic pathway. Both showed low expression except for CHI (Ilex_005941) in leaves, potentially correlating with the flavonoid content[43]. The conversion of naringenin to eriodictyol and pentahydroxyflavanone showed marked differences based on F3′H and F3′5′H tissue expression patterns. In flavone biosynthesis, FNSs from *I. asprella* were likely of the FNSI type, with greater expression in roots, suggesting higher flavone synthesis, which was also found to be a root-specific flavone pathway in other medicinal plants[44,45]. The isoflavone pathway, primarily found in legumes, involved IFS, HID, IOMT and other enzymes, such as isoflavone 2′-hydroxylase (I2′H), isoflavone 3′-hydroxylase (I3′H), vestitone reductase (VR), pterocarpan synthase (PTS), and 7,20-dihydroxy-40-methoxyisoflavanol dehydratase (DMID)[10,46,47]. IFS transcripts accumulated more in stems, with HIDs showing higher root expression, suggesting higher levels of genistein and daidzein in roots. Phlobaphene biosynthesis, a minor flavanone pathway, required FNR, which shared Ilex_039777 with ANR, likely functioning as ANR rather than FNR. However, DFR and FNR shared the same enzyme in maize[48]. The low expression of Ilex_039777 resulted in low levels of reddish insoluble pigments in *I. asprella*. F3H expression was higher in leaves, suggesting greater flavanonol synthesis, consistent with *AgF3H* correlation with DHM content in *Ampelopsis grossedentata*[49]. Conversely, FLS transcripts accumulated more in stems, suggesting higher flavonol synthesis, aligning with observations in *Fagopyrum tataricum*[50]. Leucoanthocyanidin and anthocyanin biosynthesis seem to occur in roots and stems due to high DFR expression, while DFR and FLS may inhibit each other's transcription[51]. Although LAR was associated with proanthocyanidin biosynthesis, its expression, along with ANR and ANS, was extremely low in the tissues tested. This is consistent with LAR and ANR being key and rate-limiting enzymes in proanthocyanidin biosynthesis[52].

Emerging evidence underscores the pivotal roles of various transcription factor (TF) families in the biosynthesis of flavonoids within the plant kingdom[53]. In the case of *Ilex asprella*, a comprehensive annotation identified a plethora of TF families, with MYB being the most prevalent, encompassing 93 genes, trailed by bHLH (36 genes), WD40 (240 genes), and a suite of others including NF-Ya (4 genes), NF-Yb (11 genes), NF-Yc (6 genes), BBX (21 genes), CSN (1 gene), NAC (12 genes), bZIP (4 genes), WRKY (93 genes), ERF (36 genes), PIF (2 genes), LBD (6 genes), ARF (26 genes), BES (2 genes), BT (4 genes), GRF (14 genes), SPL (9 genes), ABC (52 genes), GST (20 genes), H^+-ATPase (15 genes), Mate (36 genes), and GF9 (2 genes). These TFs exhibited differential expression across various tissues. Correlation analyses, employing Pearson's method, discerned significant positive or negative correlations between key flavonoid biosynthetic genes, exemplified by MYB (Ilex_022011) and CHI (Ilex_021868), and the aforementioned TFs. Extensive research has delineated that MYB, either singularly or in concert with the bHLH-MYB-WD40 complex, can orchestrate the expression of CHS, CHI, F3H, FLS, LAR, and DFR by engaging their promoter regions[10,54]. In *Ilex asprella*, individual *MYB* TF could closely connected numerous synthesized genes and homologous genes. For instance, the MYB TF (Ilex_002396) was associated with STS (Ilex_048273), CHI (Ilex_005941), and F6H (Ilex_012665), as well as with a cluster of F6H genes including Ilex_012665, Ilex_031189, and Ilex_004614. Additionally, a single synthetic gene might be regulated by multiple components of the MBW complex, such as CHI(Ilex_005941) gene, which was regulated by a combination of MYB(Ilex_002396), WD40(Ilex_036068, and bHLH(Ilex_030822). Parallel



findings have highlighted the regulatory impact of the NF-Y protein complex on flavonoid synthesis through its binding to the CHS1 promoter, thereby influencing the coloration of tomato skin[16]. However, no significant connection was observed between CHS and NF-Ya/b/c homologous genes. Furthermore, a cadre of TFs, such as ERF, bZIP1, PIF3, and NAC1, have been implicated in modulating MYB activity or its promoter, potentially influencing anthocyanin accumulation. Intriguingly, WRKY TFs have been shown to synergize with bHLH

◀**Fig. 7**. Correlation analysis and transcriptional regulatory network between flavonoid biosynthesis, regulation and transport in *Ilex asprella* tissues. (**a**) Correlation analysis of genes between flavonoid biosynthesis, regulation and transport in Ilex asprella tissues. *Correlation is significant between 0.01 and 0.05, **Correlation is significant between 0.001 and 0.01, and ***correlation is significant at the 0.001 level for Pearson correlation coefficient. (**b**) Transcriptional regulatory network among pathway genes and TFs in Ilex asprella tissues. Orange circle of outer ring represents the TF, yellow circle of inner ring represents the genes of flavonoid biosynthesis. The red, green and blue line are designated as significance level of p-adjust ≤ 0.001, 0.001 < p_adjust ≤ 0.01, and 0.01 < p-adjust ≤ 0.05 between TFs and genes. The image was created by the protocol of user manual provided Cytoscape_3.9.1 (https://cytoscape.org/) software.

to enhance MYB activity, thereby propelling anthocyanin biosynthesis. Advanced studies have also indicated that certain TFs, like GRF, exert long-distance control over anthocyanin biosynthesis by negatively modulating BT, consequently augmenting MYB levels in apple[10,18,24]. TFs such as ERF (Ilex_012123), NAC (Ilex_030072), WRKY (Ilex_045823), and the MYB/bHLH/WD40 complex (Ilex_002396/Ilex_039504/Ilex_036068) were all positively associated with key enzymes, 4CL (Ilex_006117) and AS (Ilex_024910), in flavonoid biosynthesis. This suggested that these TFs could potentially cooperate with the MBW complex to regulate flavonoid synthesis in response to various environmental conditions. However, a similar association was not observed for GRF, BT, and MBW complexes, possibly indicating that they required a more extended pathway to interact with flavonoid biosynthesis. The multitude of transporters and synthesized genes involved in flavonoid biosynthesis have evolved intricate strategies to finely tune their responses to stimuli, thereby maintaining balanced flavonoid levels. Also, it is axiomatic that a thorough elucidation of the intricate transcriptional regulatory networks at play necessitates both data mining and empirical validation. The TFs characterized in *I. asprella* are poised to be instrumental in the modulation of flavonoid levels and the orchestration of their biosynthesis. These TF-gene associations present promising avenues for future investigations into the biology of *I. asprella*.

Materials and methods

Plant materials

Ilex asprella plants were cultivated in soil under environmental conditions, specifically at a temperature range of 24–26 °C and a relative humidity of 60–70%. On May 30, 2021, a systematic sampling procedure was conducted to collect three replicate samples from each of the plant's organs—roots, stems, and leaves—from two-year-old *I. asprella* seedlings. The samples were designated as root 1, root 2, root 3, stem 1, stem 2, stem 3, leaf 1, leaf 2, and leaf 3 (Fig. 1). For each tissue type, three individual plants were harvested, and approximately 1 g of material was excised using scissors. These samples were destined for both transcriptome analysis and real-time quantitative polymerase chain reaction (qRT-PCR) assays. The samples were immediately frozen in liquid nitrogen and then stored at − 80 °C for subsequent molecular studies.

RNA extraction and quality assessment

The extraction of total RNA from the plant samples was conducted using a kit specifically designed for RNA analysis, as provided by Takara, located in Dalian, Liaoning, China. The protocol supplied by the manufacturer was followed meticulously. The assessment of RNA purity and its integrity was carried out with precision instruments: a NanoDrop 2000 spectrophotometer from Thermo Fisher Scientific, based in Waltham, MA, USA, and a Bioanalyzer 2100 system from Agilent Technologies, headquartered in Palo Alto, CA, USA. To ensure the RNA's quality, any signs of degradation or contamination were carefully examined by agarose gel electrophoresis using a 1% (w/v) gel. The RNA concentration was determined with a Qubit® RNA Assay Kit using a Qubit® 2.0 fluorometer from Life Technologies (Carlsbad, CA, USA). All RNA samples met the required specifications and were transported to the BGI Corporation (Shenzhen, China) for cDNA library construction and RNA sequencing while being stored on dry ice.

cDNA library construction and sequencing

Nine distinct RNA libraries were meticulously constructed using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina®. Each library represented a unique tissue type, with triplicate samples ensuring biological redundancy. The preparation process adhered to the guidelines provided by New England Biolabs, located in Ipswich, MA, USA. The transcriptome assembly was facilitated by the Illumina NovaSeq 6000 sequencing platform, a product of Illumina Inc., based in San Diego, CA, USA. The quality of the libraries was rigorously evaluated with an Agilent Bioanalyzer 2100 system. The cDNA libraries underwent paired-end sequencing on an Illumina HiSeqX-10 instrument, yielding 150 bp reads. Post-sequencing, reads of low quality—those with adapter contamination, ambiguous nucleotides, or more than 50% of bases with a Phred quality score (Q score) ≤ 20—were meticulously filtered out to ensure the integrity of the clean reads. These clean reads were then subjected to quality metrics assessment, including the calculation of Q20, Q30, and GC content, prior to further analytical procedures. The sequencing reads were deposited at the National Center for Biotechnology Information under SRA accession number PRJNA1168255 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1168255).

De novo assembly and unigene annotation

Transcriptome assembly was performed using Trinity v2.8 software (https://trinitysoft.net/) with, minium_k-mer_coverage (min_kmer_cov) parameter set to 3 by default, and all other settings maintained at their default values[55]. To elucidate the potential functions of the assembled unigenes, a comprehensive similarity search was conducted. This involved BLAST queries against a suite of publicly accessible nucleotide and protein databases.

Specifically, the NCBI nonredundant protein database (Nr) was searched using diamond v0.8.22 with an e-value threshold of 10^-5, the NCBI nonredundant nucleotide database (Nt) was queried using BLAST 2.2.28 with an e-value of 10^-5, the protein family (Pfam) database was explored using the HMMER 3.0 package with an e-value of 10^{-2}, the Swiss-Prot database was investigated using diamond v0.8.22 with an e-value of 10^{-5}, the Kyoto Encyclopedia of Genes and Genomes (KEGG) was accessed through KAAS with an e-value of 10^{-10}, and the Gene Ontology (GO) database was queried using Blast2GO v2.5 with an e-value of 10^{-656}. Furthermore, the analysis of differential gene expression (DGE) was executed following a methodology akin to that detailed in a preceding study[57].

Expression levels

Gene expression levels across all samples were determined by aligning clean reads to the assembled transcriptome using the Bowtie v2.4.5 software (https://bowtiemaster.com/). The expression levels were quantified using the fragments per kilobase of the exon model per million mapped reads (FPKM) metric, which provided a normalized measure of transcript abundance[58,59]. To identify the DEGs between samples, the DESeq2 and EBSeq R packages were employed. Genes were classified as DEGs if they exhibited an adjusted p-value of less than 0.05 and a log2-fold change magnitude greater than 1, as indicated by DESeq. The average FPKM values from the samples were subjected to log transformation and normalization to facilitate heatmap analysis. For functional annotation, GO enrichment analysis of the DEGs was performed using the GOseq R package, which employed the Wallenius noncentral hypergeometric distribution. Additionally, the KOBAS v2.0 software (http://bioinfo.org/kobas/) was utilized to evaluate the enrichment of DEGs in KEGG pathways[60].

Identification of synthase and regulatory genes involved in flavonoid biosynthetic pathways

In order to pinpoint candidate synthase and regulatory genes within *I. asprella*, a BLAST search was conducted using the transcriptome as a reference. For this purpose, genes that have been published and identified from other plant species served as the query sequence. Each sequence of interest was further authenticated by conducting a BLAST search against the NCBI nonredundant (Nr) database. Following this, a phylogenetic analysis was performed to infer evolutionary relationships. A neighbor-joining tree was constructed using Clustal X 2.0 for alignment and MEGA 7.0 for visualization, based on the inferred amino acid sequences derived from the transcriptomes of *Ilex asprella* and other species.

cDNA synthesis and RT-qPCR

Beyond its application in transcriptome sequencing, the residual RNA was allocated for complementary DNA (cDNA) synthesis. This was achieved using the First Strand cDNA Synthesis Kit from Thermo Fisher Scientific, Shanghai, China, strictly following the manufacturer's protocol. Quantitative real-time polymerase chain reaction (qRT-PCR) was conducted on the Applied Biosystems QuantStudio™ Real-Time PCR System, Waltham, MA, USA, utilizing Hieff™ qPCR SYBR® Green Master Mix from Yeasten, Shanghai, China. The qRT-PCR was performed in triplicate for each of the three independent biological samples, with each sample being subjected to three technical replicates. The reaction mixture, totaling 20 µL, comprised 10.0 µL of 2 × Hieff™ qPCR SYBR® Green Master Mix, 0.6 µL of each primer at a concentration of 10 µM, 1.0 µL of a 1:5 diluted cDNA template, and 7.8 µL of RNase-free water. The thermal cycling conditions were initiated with a hot start at 95 °C for 5 min, followed by 40 cycles of denaturation at 95 °C for 10 s, annealing at 58 °C for 20 s, and extension at 72 °C for 20 s. For each sample, three replicates were analyzed using real-time PCR. Data analysis was conducted in accordance with previous methods[57]. The correlation between qPCR and RNA sequencing data was evaluated using OriginPro 2017 software (https://www.originlab.com/2017). The primers for all the target genes and three ubiquitin (UBIs) used as internal references were detailed in Table S8.

Statistical analysis

All the data were analyzed using SPSS 20.0 software (https://www.ibm.com/spss), and the values were expressed as the means ± standard errors of three independent samples. Significant differences ($p < 0.05$, $p < 0.01$) were analyzed using one-way analysis of variance (ANOVA) and Duncan's multiple range test.

Conclusions

In this study, we systematically identified and analyzed genes involved in flavonoid biosynthesis, regulation, and transport across the leaf, stem, and root tissues of *I. asprella* through transcriptome sequencing. Our analysis revealed 28,478 DEGs were implicated in a spectrum of biological processes, including "carbohydrate metabolism", "signal transduction", and "folding, sorting and degradation", as annotated by multiple databases. The metabolic flux of flavonoids varied significantly among the tissues. The expression patterns of key enzymes such as PAL (Ilex_042231, Ilex_014816), C4H (Ilex_017598), and 4CL (Ilex_042033) indicated that the general phenylpropanoid pathway was more active in roots and stems compared to leaves. Chalcone and isoflavone biosynthesis, driven by CHS (Ilex_047537) and IFS (Ilex_029360), appeared to be more pronounced in stems. Conversely, the synthesis of stilbenes, aurones, and flavones, governed by STS (Ilex_044726), CH4′GT (Ilex_047989), and FNS (Ilex_043640), respectively, was predicted to be higher in roots. Flavanones and phlobaphenes were likely more abundant in leaves, as indicated by the expression of CHI (Ilex_005941) and FNR (Ilex_039777). The synthesis of flavanonol in roots or stems was primarily influenced by F3H (Ilex_004635), and its downstream conversion to flavonols in leaves and stems was regulated by FLS (Ilex_046424). The biosynthesis of leucoanthocyanidin and anthocyanin was potentially controlled by DFR (Ilex_004771, Ilex_023203) in leaves and stems. Notably, the absence of LAR transcripts suggested that proanthocyanidins, such as gallocatechin, afzelechin, and catechin, were not synthesized in *I. asprella*. Furthermore, the correlation analysis of transporter-

gene pairs provided a theoretical foundation for the genetic enhancement of flavonoid metabolism and deepened our understanding of their roles and potential applications.

Data availability

Sequence data that support the findings of this study have been deposited in the European Nucleotide Archive with the primary accession code PRJNA1168255 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1168255).

Received: 19 September 2024; Accepted: 26 November 2024
Published online: 02 December 2024

References

1. Zhao, Z. X., Lin, C. Z., Zhu, C. C. & He, W. J. A new triterpenoid glycoside from the roots of *Ilex asprella*. *Chin. J. Nat. Med.* **11**, 415–418. https://doi.org/10.1016/s1875-5364(13)60062-x (2013).
2. Du, B. Z. et al. A phytochemical and pharmacological advance on *Ilex asprella*. *Zhongguo Zhong Yao Za Zhi* **42**, 20–28. https://doi.org/10.19540/j.cnki.cjcmm.20161222.025 (2017).
3. Zheng, X. et al. Characterisation of two oxidosqualene cyclases responsible for triterpenoid biosynthesis in Ilex asprella. *Int. J. Mol. Sci.* **16**, 3564–3578. https://doi.org/10.3390/ijms16023564 (2015).
4. Wen, Q., Lu, Y., Chao, Z. & Chen, D. F. Anticomplement triterpenoids from the roots of Ilex asprella. *Bioorg. Med. Chem. Lett.* **27**, 880–886. https://doi.org/10.1016/j.bmcl.2017.01.007 (2017).
5. Peng, M. H. et al. Bioactive glycosides from the roots of Ilex asprella. *Pharm. Biol.* **54**, 2127–2134. https://doi.org/10.3109/13880209.2016.1146779 (2016).
6. Kong, B. L. et al. Chromosomal level genome of Ilex asprella and insight into antiviral triterpenoid pathway. *Genomics* **114**, 110366. https://doi.org/10.1016/j.ygeno.2022.110366 (2022).
7. Zheng, X., Xu, H., Ma, X., Zhan, R. & Chen, W. Triterpenoid saponin biosynthetic pathway profiling and candidate gene mining of the Ilex asprella root using RNA-Seq. *Int. J. Mol. Sci.* **15**, 5970–5987. https://doi.org/10.3390/ijms15045970 (2014).
8. Kong, B. L. et al. Comparative analysis and phylogenetic investigation of Hong Kong Ilex chloroplast genomes. *Sci. Rep.* **11**, 5153. https://doi.org/10.1038/s41598-021-84705-9 (2021).
9. Li, L. et al. Cytochrome P450 monooxygenase/cytochrome P450 reductase bi-enzymatic system isolated from Ilex asprella for regio-specific oxidation of pentacyclic triterpenoids. *Front. Plant Sci.* **13**, 831401. https://doi.org/10.3389/fpls.2022.831401 (2022).
10. Liu, W. et al. The flavonoid biosynthesis network in plants. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms222312824 (2021).
11. Espley, R. V. & Jaakola, L. The role of environmental stress in fruit pigmentation. *Plant Cell Environ.* **46**, 3663–3679. https://doi.org/10.1111/pce.14684 (2023).
12. Zhang, X. et al. The CsHSFA-CsJAZ6 module-mediated high temperature regulates flavonoid metabolism in Camellia sinensis. *Plant Cell Environ.* **46**, 2401–2418. https://doi.org/10.1111/pce.14610 (2023).
13. Li, H. L. et al. E3 ubiquitin ligases SINA4 and SINA11 regulate anthocyanin biosynthesis by targeting the IAA29-ARF5-1-ERF3 module in apple. *Plant Cell Environ.* **46**, 3902–3918. https://doi.org/10.1111/pce.14709 (2023).
14. Li, C. et al. PyWRKY26 and PybHLH3 cotargeted the PyMYB114 promoter to regulate anthocyanin biosynthesis and transport in red-skinned pears. *Hortic. Res.* **7**, 37. https://doi.org/10.1038/s41438-020-0254-z (2020).
15. Xu, H. et al. Overexpression of a repressor MdMYB15L negatively regulates anthocyanin and cold tolerance in red-fleshed callus. *Biochem. Biophys. Res. Commun.* **500**, 405–410. https://doi.org/10.1016/j.bbrc.2018.04.088 (2018).
16. Wang, J. et al. NF-Y plays essential roles in flavonoid biosynthesis by modulating histone modifications in tomato. *New Phytol.* **229**, 3237–3252. https://doi.org/10.1111/nph.17112 (2021).
17. Ni, J. et al. Ethylene response factors Pp4ERF24 and Pp12ERF96 regulate blue light-induced anthocyanin biosynthesis in "Red Zaosu" pear fruits by interacting with MYB114. *Plant Mol. Biol.* **99**, 67–78. https://doi.org/10.1007/s11103-018-0802-1 (2019).
18. Zhao, X. et al. CsbZIP1-CsMYB12 mediates the production of bitter-tasting flavonols in tea plants (Camellia sinensis) through a coordinated activator-repressor network. *Hortic. Res.* **8**, 110. https://doi.org/10.1038/s41438-021-00545-8 (2021).
19. Zhou, H. et al. Molecular genetics of blood-fleshed peach reveals activation of anthocyanin biosynthesis by NAC transcription factors. *Plant J.* **82**, 105–121. https://doi.org/10.1111/tpj.12792 (2015).
20. Li, J. et al. CsMYB60 directly and indirectly activates structural genes to promote the biosynthesis of flavonols and proanthocyanidins in cucumber. *Hortic. Res.* **7**, 103. https://doi.org/10.1038/s41438-020-0327-z (2020).
21. Ren, Y. R. et al. The apple 14–3-3 protein MdGRF11 interacts with the BTB protein MdBT2 to regulate nitrate deficiency-induced anthocyanin accumulation. *Hortic. Res.* **8**, 22. https://doi.org/10.1038/s41438-020-00457-z (2021).
22. Luo, D. et al. SlBBX20 interacts with the COP9 signalosome subunit SlCSN5-2 to regulate anthocyanin biosynthesis by activating SlDFR expression in tomato. *Hortic. Res.* **8**, 163. https://doi.org/10.1038/s41438-021-00595-y (2021).
23. Wang, Y. et al. Nitrogen affects anthocyanin biosynthesis by regulating MdLOB52 downstream of MdARF19 in callus cultures of red-fleshed apple (Malus sieversii f. niedzwetzkyana). *J. Plant Growth Regul.* **37**, 719–729 (2017).
24. Liang, T. et al. Brassinosteroid-activated BRI1-EMS-SUPPRESSOR 1 inhibits flavonoid biosynthesis and coordinates growth and UV-B stress responses in plants. *Plant Cell* **32**, 3224–3239. https://doi.org/10.1105/tpc.20.00048 (2020).
25. Zhao, J. Flavonoid transport mechanisms: How to go, and with whom. *Trends Plant Sci.* **20**, 576–585. https://doi.org/10.1016/j.tplants.2015.06.007 (2015).
26. Poloni, A. et al. Transcriptome analysis reveals contrasting plant responses of sorghum bicolor upon colonization by two formae speciales of Sporisorium reilianum. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms23168864 (2022).
27. Sun, Y. et al. Transcriptome analysis to identify genes related to flowering reversion in tomato. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms23168992 (2022).
28. Ma, J., Li, J., Xu, Z., Wang, F. & Xiong, A. Transcriptome profiling of genes involving in carotenoid biosynthesis and accumulation between leaf and root of carrot (Daucus carota L.). *Acta Biochim. Biophys. Sin. (Shanghai)* **50**, 481–490. https://doi.org/10.1093/abbs/gmy027 (2018).
29. Bai, C., Xu, J., Cao, B., Li, X. & Li, G. Transcriptomic analysis and dynamic expression of genes reveal flavonoid synthesis in Scutellaria viscidula. *Acta Physiol. Plant* **40**, 161 (2018).
30. Winkel-Shirley, B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Phytol.* **126**, 485–493. https://doi.org/10.1104/pp.126.2.485 (2001).
31. Vogt, T. Phenylpropanoid biosynthesis. *Mol. Plant* **3**, 2–20. https://doi.org/10.1093/mp/ssp106 (2010).
32. Strygina, K. & Khlestkina, E. Flavonoid biosynthesis genes in *Triticum aestivum* L.: Methylation patterns in cis-regulatory regions of the duplicated CHI and F3H genes. *Biomolecules.* https://doi.org/10.3390/biom12050689 (2022).
33. Murota, K., Nakamura, Y. & Uehara, M. Flavonoid metabolism: The interaction of metabolites and gut microbiota. *Biosci. Biotechnol. Biochem.* **82**, 600–610. https://doi.org/10.1080/09168451.2018.1444467 (2018).
34. Shang, Q. M., Li, L. & Dong, C. J. Multiple tandem duplication of the phenylalanine ammonia-lyase genes in Cucumis sativus L. *Planta* **236**, 1093–1105. https://doi.org/10.1007/s00425-012-1659-1 (2012).

35. Tuan, P. A. et al. Molecular cloning and characterization of phenylalanine ammonia-lyase and cinnamate 4-hydroxylase in the phenylpropanoid biosynthesis pathway in garlic (Allium sativum). *J. Agric. Food. Chem.* **58**, 10911–10917. https://doi.org/10.1021/jf1021384 (2010).
36. Loke, K. K. et al. Transcriptome analysis of Polygonum minus reveals candidate genes involved in important secondary metabolic pathways of phenylpropanoids and flavonoids. *PeerJ* **5**, e2938. https://doi.org/10.7717/peerj.2938 (2017).
37. Schijlen, E. G. et al. RNA interference silencing of chalcone synthase, the first step in the flavonoid biosynthesis pathway, leads to parthenocarpic tomato fruits. *Plant Physiol.* **144**, 1520–1530. https://doi.org/10.1104/pp.107.100305 (2007).
38. Shimada, N. et al. Isolation and characterization of a cDNA encoding polyketide reductase in Lotus japonicus. *Plant Biotechnol.* **23**, 509–513 (2006).
39. Zhou, L., Yan, W. & Zhenhua, P. Advances in study on formation mechanism and genetic engineering of yellow flowers. *Sci. Silvae Sin.* **45**, 111–119 (2009).
40. Wang, X. et al. Tissue-specific transcriptome analyses reveal candidate genes for stilbene, flavonoid and anthraquinone biosynthesis in the medicinal plant Polygonum cuspidatum. *BMC Genomics* **22**, 353. https://doi.org/10.1186/s12864-021-07658-3 (2021).
41. Alsayari, A. et al. Aurone: A biologically attractive scaffold as anticancer agent. *Eur. J. Med. Chem.* **166**, 417–431. https://doi.org/10.1016/j.ejmech.2019.01.078 (2019).
42. Nakayama, T. et al. Aureusidin synthase: A polyphenol oxidase homolog responsible for flower coloration. *Science* **290**, 1163–1166. https://doi.org/10.1126/science.290.5494.1163 (2000).
43. Jiang, W. et al. Role of a chalcone isomerase-like protein in flavonoid biosynthesis in Arabidopsis thaliana. *J. Exp. Bot.* **66**, 7165–7179. https://doi.org/10.1093/jxb/erv413 (2015).
44. Wu, J. et al. Flavone synthases from *Lonicera japonica* and *L. macranthoides* reveal differential flavone accumulation. *Sci. Rep.* **6**, 19245. https://doi.org/10.1038/srep19245 (2016).
45. Zhao, Q. et al. A specialized flavone biosynthetic pathway has evolved in the medicinal plant, *Scutellaria baicalensis*. *Sci. Adv.* **2**, e1501780. https://doi.org/10.1126/sciadv.1501780 (2016).
46. Chu, S. et al. An R2R3-type MYB transcription factor, GmMYB29, regulates isoflavone biosynthesis in soybean. *PLoS Genet.* **13**, e1006770. https://doi.org/10.1371/journal.pgen.1006770 (2017).
47. Dong, N. Q. & Lin, H. X. Contribution of phenylpropanoid metabolism to plant development and plant-environment interactions. *J. Integ. Plant Biol.* **63**, 180–209. https://doi.org/10.1111/jipb.13054 (2021).
48. Halbwirth, H., Martens, S., Wienand, U., Forkmann, G. & Stich, K. Biochemical formation of anthocyanins in silk tissue of Zea mays. *Plant Sci.* **164**, 489–495 (2003).
49. Khumkarjorn, N., Thanonkeo, S., Yamada, M. & Thanonkeo, P. Cloning and expression analysis of a flavanone 3-hydroxylase gene in Ascocenda orchid. *Acta Bot. Boreali-Occident Sin.* **40**, 185–192 (2020).
50. Li, X. et al. Differential stress-response expression of two flavonol synthase genes and accumulation of flavonols in tartary buckwheat. *J. Plant Physiol.* **170**, 1630–1636. https://doi.org/10.1016/j.jplph.2013.06.010 (2013).
51. Chen, W. et al. Competition between anthocyanin and kaempferol glycosides biosynthesis affects pollen tube growth and seed set of Malus. *Hortic. Res.* **8**, 173. https://doi.org/10.1038/s41438-021-00609-9 (2021).
52. Lepiniec, L. et al. Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant.* **57**, 405–430. https://doi.org/10.1146/annurev.arplant.57.032905.105252 (2006).
53. Yang, C. Q. et al. Transcriptional regulation of plant secondary metabolism. *J. Integr. Plant Biol.* **54**, 703–712. https://doi.org/10.1111/j.1744-7909.2012.01161.x (2012).
54. Zhang, B. et al. CRISPR/Cas9-mediated targeted mutation reveals a role for AN4 rather than DPL in regulating venation formation in the corolla tube of Petunia hybrida. *Hortic. Res.* **8**, 116. https://doi.org/10.1038/s41438-021-00555-6 (2021).
55. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652. https://doi.org/10.1038/nbt.1883 (2011).
56. Ogata, H. et al. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**, 29–34. https://doi.org/10.1093/nar/27.1.29 (1999).
57. Zhang, G. et al. Transcriptome and metabolic profiling unveiled roles of peroxidases in Theaflavin production in black tea processing and determination of tea processing suitability. *J. Agric. Food. Chem.* **68**, 3528–3538. https://doi.org/10.1021/acs.jafc.9b07737 (2020).
58. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. methods* **9**, 357–359. https://doi.org/10.1038/nmeth.1923 (2012).
59. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628. https://doi.org/10.1038/nmeth.1226 (2008).
60. Mao, X., Cai, T., Olyarchuk, J. G. & Wei, L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* **21**, 3787–3793. https://doi.org/10.1093/bioinformatics/bti430 (2005).

Acknowledgements

We appreciate the patience of the reviewers and editors for the work and thank Prof. Li's laboratory members for assistance with the experiments and data analyses. We also thank Zhengzhou Han, the director of the Medical Plant Resources Department, Shenzhen Traditional Chinese Medicine Manufacturing Innovation Center Co., Ltd., Shenzhen, China, who have identified the plant (number HAUT-999-01) deposited in the School of Biological Engineering, Henan University of Technology, Zhengzhou, China.

Author contributions

Z.G., L.C. and S.Z conceived and designed the research. Z.G., M.Q. and W.S. planted the materials and collected the samples. Z.G. and M.Q. analyzed transcriptome data, and W.S. conducted the RT-qPCR experiments. T.H. performed the correlation and network analyses. Z.G. wrote the manuscript. L.C. and S.Z. revised the manuscript. All the authors have read and approved the final manuscript.

Funding

This work was supported by China Resources Sanjiu Medical & Pharmaceutical Co., Ltd. (Grant No. SJCGXY202212120002), Major Science and Technology Projects in Henan Province (231100110300) and the High Level Research Fund for Qualified People of Henan University of Technology (2021BS017).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-81319-9.

Correspondence and requests for materials should be addressed to C.L. or G.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.