Computational and Structural Biotechnology Journal 23 (2024) 1106–1116

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

Multi-omics analysis reveals promiscuous *O*-glycosyltransferases involved in the diversity of flavonoid glycosides in *Periploca forrestii* (Apocynaceae)

Xiaotong Wang [a,c,1], Lan Wu [a,b,1], Wanran Zhang [c], Shi Qiu [d], Zhichao Xu [c], Huihua Wan [a,b], Jiang He [e], Wenting Wang [a], Mengyue Wang [a,b], Qinggang Yin [a,b], Yuhua Shi [a,b], Ranran Gao [a,b,*], Li Xiang [a,b,e,f,*], Weijun Yang [e,**]

[a] *Artemisinin Research Center, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China*
[b] *Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China*
[c] *Ministry of Education, Key Laboratory of Saline-alkali Vegetation Ecology Restoration (Northeast Forestry University), Harbin 150006, China*
[d] *The SATCM Key Laboratory for New Resources & Quality Evaluation of Chinese Medicine, Institute of Chinese Materia Medica, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*
[e] *Xinjiang Institute of Materia Medica/Key Laboratory of Xinjiang Uygur Medicine, Urumqi 830004, China*
[f] *Prescription Laboratory of Xinjiang Traditional Uyghur Medicine, Xinjiang Institute of Traditional Uyghur Medicine, Urmuqi 830000, China*

ARTICLE INFO

Keywords:
Flavonoid glycoside
Glycosyltransferase
Biosynthesis
Periploca forrestii
Full-length transcriptome
Metabolome

ABSTRACT

Flavonoid glycosides are widespread in plants, and are of great interest owing to their diverse biological activities and effectiveness in preventing chronic diseases. *Periploca forrestii*, a renowned medicinal plant of the Apocynaceae family, contains diverse flavonoid glycosides and is clinically used to treat rheumatoid arthritis and traumatic injuries. However, the mechanisms underlying the biosynthesis of these flavonoid glycosides have not yet been elucidated. In this study, we used widely targeted metabolomics and full-length transcriptome sequencing to identify flavonoid diversity and biosynthetic genes in *P. forrestii*. A total of 120 flavonoid glycosides, including 21 *C*-, 96 *O*-, and 3 *C/O*-glycosides, were identified and annotated. Based on 24,123 full-length coding sequences, 99 uridine diphosphate sugar-utilizing glycosyltransferases (UGTs) were identified and classified into 14 groups. Biochemical assays revealed that four UGTs exhibited *O*-glycosyltransferase activity toward apigenin and luteolin. Among them, PfUGT74B4 and PfUGT92A8 were highly promiscuous and exhibited multisite *O*-glycosylation or consecutive glycosylation activities toward various flavonoid aglycones. These four glycosyltransferases may significantly contribute to the diversity of flavonoid glycosides in *P. forrestii*. Our findings provide a valuable genetic resource for further studies on *P. forrestii* and insights into the metabolic engineering of bioactive flavonoid glycosides.

1. Introduction

Flavonoids and their glycoside products play crucial roles in plants' response to biotic and abiotic stresses, as well as in regulating growth and development [1,2]. Compared to flavonoid aglycones, flavonoid glycosides exhibit enhanced bioactivity and water solubility, lower toxicity, and fewer side effects [3–5]. The family Apocynaceae has garnered increasing attention because of the exceptional medicinal value of its bioactive ingredients and the accumulation of a diverse array of flavonoid glycosides, including flavonol, isoflavone, and flavone glycosides [5,6]. *Periploca forrestii* Schltr, a well-known medicinal shrub within this family[7,8], has been used to treat rheumatoid arthritis,

Abbreviations: UGTs, uridine diphosphate sugar-utilizing glycosyltransferases; PAL, phenylalanine ammonia lyase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4-coumarate-CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; FNS, flavone synthase; FLS, flavonol synthase; F3H, flavanone-3β-hydroxylase; WGD, whole-genome duplication; CCS, circular consensus sequence; FLNC, full-length nonchimeric; ORFs, open reading frames; K_S, synonymous substitutions per synonymous site; FPKM, fragments per kilobase of the exon model per million mapped fragments; DEGs, differentially expressed genes.

* Corresponding authors at: Artemisinin Research Center, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China.
** Corresponding author.
E-mail addresses: rrgao1991@icmm.ac.cn (R. Gao), lxiang@icmm.ac.cn (L. Xiang), wilfred3106@163.com (W. Yang).
[1] These authors contributed equally: Xiaotong Wang, Lan Wu.

https://doi.org/10.1016/j.csbj.2024.02.028
Received 13 January 2024; Received in revised form 28 February 2024; Accepted 28 February 2024
Available online 2 March 2024

stomachache, dyspepsia, and traumatic injury with few side effects [9–11]. Various flavonoids in *P. forrestii* have been identified and characterized the content of flavonoid glycosides, including kaempferol-3-*O*-*β*-D-galactoside, quercetin-3-*O*-*β*-D-glucopyranoside (isoquercetin), quercetin-7-*O*-*β*-D-glucopyranoside, and quercetin-3-*O*-*α*-L-arabinopyranoside, have been characterized[12,13]. Moreover, the position of sugar groups in flavonoid compounds could cause differences in their bioactivities, such as quercetin-3-*O*-*β*-D-glucopyranoside and quercetin-4'-*O*-*β*-D-glucopyranoside[14,15], luteolin-7-*O*-*β*-D-glucopyranoside and luteolin-4'-*O*-*β*-D-glucopyranoside[16].

Because of the structural complexity of glycosides in *P. forrestii*, the functional identification of biosynthetic enzymes could provide insights into the metabolic engineering of bioactive flavonoid glycosides. The biosynthetic pathways of flavonoid glycosides primarily involve flavonoid skeleton formation followed by glycosylation. The basic flavonoid skeleton is derived from phenylalanine through the phenylpropanoid pathway[17,18]. Phenylalanine is enzymatically converted to *p*-coumaroyl-CoA by phenylalanine ammonia-lyase (PAL), cinnamic acid 4-hydroxylase (C4H), and 4-coumarate-CoA ligase (4CL). Chalcone, an upstream intermediate in flavonoid synthesis, is synthesized from the reaction of one molecule of 4-coumaroyl-CoA with three molecules of malonyl-CoA, which is catalyzed by chalcone synthase (CHS). Various enzymes in different plant species catalyze the production of diverse branching products from chalcone. For example, chalcone isomerase (CHI) converts chalcones to flavanones, whereas flavone synthase (FNS) transforms flavanones into flavones. Furthermore, flavanone 3β-hydroxylase (F3H) and flavonol synthase (FLS) catalyzes the conversion of flavanones to flavonols. Modification of flavonoids via glycosylation is catalyzed by uridine diphosphate sugar-utilizing glycosyltransferases (UGTs) [5,19]. Therefore, the functional identification of UGTs from different species may facilitate the discovery of diverse flavonoid glycosides[20]. Although a range of UGTs can selectively or promiscuously utilize various flavonoids for synthesizing diverse flavonoid glycosides, such as 3-*O*-glycosyltransferases (3-*O*-GTs), 5-*O*-GTs, 7-*O*-GTs, and multi-*O*-GTs[21–23], the identification of more efficient and versatile UGTs is necessary for the industrial production of novel flavonoid glycosides. However, the lack of comprehensive metabolomic analysis and genome sequencing has limited the discovery of glycosyltransferase homologs involved in the glycosylation of flavonoids in *P. forrestii*. In the absence of whole genome data, full-length transcriptome data could be good alternative for understanding the species speciation and evolutionary trajectories, such as divergence time and ancient whole genome duplication[24,25].

In this study, we established the full-length transcriptomic and metabolomic profiles of *P. forrestii* to understand its speciation and flavonoid glycosylation. We analyzed the occurrence of whole-genome duplication (WGD) events and the blueprint for the flavonoid glycoside biosynthetic pathway in *P. forrestii* based on the full-length transcriptome. We systematically screened and identified four *O*-GTs, namely PfUGT75C6, PfUGT74E7, PfUGT74B4, and PfUGT92A8, which are responsible for the biosynthesis of flavonoid-*O*-glycosides, including mono-*O*-glycosides, di-*O*-glycosides, and tri-*O*-glycosides by combining transcriptomic, metabolomic, and biochemical data. The promiscuity of PfUGT74B4 and PfUGT92A8 was further confirmed using different flavonoids as glycosyl acceptors. Our findings provide genetic information on *P. forrestii* and the key enzymes involved in the metabolic diversity of flavonoid glycosides. The multisite activity and promiscuity of PfUGT74B4 and PfUGT92A8 will benefit the future metabolic engineering of bioactive flavonoid glycosides.

2. Materials and methods

2.1. Plant materials and chemicals

P. forrestii was cultivated in Anlong County, of Southwest Guizhou Autonomous Prefecture, China. The roots, stems, and leaves of *P. forrestii* plants were collected, immediately crushed in liquid nitrogen, and frozen at − 80 °C for subsequent metabolic detection, RNA-sequencing and transcriptomic sequencing. Three biological replicates were analyzed for each sample.

The chemical standards for apigenin, phloretin, luteolin, orientin, isoorientin, vitexin, and isovitexin were purchased from Wuhan Jonk Biological Technology Co., Ltd, Wuhan, China. Vicenin-2 was purchased from Xian Qiyue Biological Technology Co., Ltd, Xian, China. UDP-glucose was obtained from Sigma-Aldrich, St. Louis, MO. Apigenin-6,8-di-*C*-glucoside and 2-hydroxynaringenin were purchased from Chengdu Caoyuankang Bio-technology Co., LTD, Chengdu, China.

2.2. RNA extraction, PacBio sequencing, full-length transcript data processing, and gene annotation

The RNeasy Plant Micro Kit (Qiagen, Germany) was used to extract total RNA, and its integrity was assessed using an Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). Complementary DNA (cDNA) libraries were prepared by Benagen Biotechnology Co. Ltd, Wuhan, China and subsequently sequenced using the PacBio Sequel platform. Polymerase reads of single molecule sequences were classified as subreads. Subreads derived from the same polymerase reads were self-corrected and combined to form a circular consensus sequence (CCS). Full-length non-chimeric (FLNC) refers to CCSs that include both sequencing primer sequences and a 3′-terminal poly-A sequence but no chimeric sequence. Calibration was performed by comparing the RNA sequencing (RNA-seq) data with FLNC data from the NGS platform. TransDecoder software was used to determine the open reading frames (ORFs) of the full-length transcripts. The assembly results were annotated using the following databases: Swiss-Prot, Pfam, Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), NonRedundant Protein Sequence Database (nr), Clusters of Eukaryotic Orthologous Groups (KOG), Translation of EMBL (TrEMBL), the Comprehensive Antibiotic Research Database (CARD), Carbohydrate-Active Enzyme Database (CAZy), Pathogen Host Interaction Database (PHI), and Virulence Factors of Pathogenic Bacteria (VFDB). Differential expression between intergroup samples was analyzed using DESeq2. R package (version 1.26.0), setting parameters with a q value < 0.05, and $|\log2FoldChange| > 1$ [26]. All up- and -down-regulated differentially expressed genes (DEGs) between the two samples were analyzed for enrichment using ClusterProfiler (version 3.14.3) [27].

2.3. Establishment of metabolite profiles using UPLC–ESI–MS/MS and data analysis

Powdered samples (100 mg) of *P. forrestii* roots, stems, and leaves were dissolved in 1.2 mL of 70% methanol. The metabolites were identified using an ultra-performance liquid chromatography-electrospray tandem mass spectrometry (UPLC–ESI–MS/MS) system (UPLC, Shim-pack UFLC SHIMADZU CBM30A system; MS, Applied Biosystems 6500 Q TRAP) at Wuhan MetWare Biotechnology Co., Ltd, Wuhan, China (www.metware.cn). Samples were injected into an Agilent SB-C18 column (1.8 μm, 2.1 mm × 100 mm) at a flow rate of 0.35 mL/min, with a column temperature of 40 °C and an injection volume of 4 μL. The liquid phase included two mobile phases: the aqueous phase was 0.1% formic in ultrapure water (A), whereas the organic phase was acetonitrile (B). The elution gradient was as follows: 5% B, 0.00 min; 5–95% B, 9.00 min; 95% B, 10.00 min; 95–5% B, 11.10 min; and 5% B, 14.00 min. The chromatographic peaks of the samples were integrated and calibrated using the MultiQuant software. The metabolites were identified basing on the extracted ion mass, MS^2 fragment, MS^2 fragment isotope distribution, and retention time (RT) through intelligent matching within the MetWare DataBase[28]. Metabolite quantification was performed via a LC/-multiple reaction monitoring (MRM)-MS. The area of each chromatographic peak represented the relative content of the corresponding metabolite.

2.4. WGD identification and phylogenomic analysis

OrthoFinder (v2.3.12) was used to detect single-copy orthologous genes from *P. forrestii* and ten other angiosperm species, i.e. *Coffea canephora* [29], *Gardenia jasminoides* [30], *Asclepias syriaca* [31], *Calotropis gigantea* [32], *Catharanthus roseus* [33], *Gentiana dahurica* [34], *Antirrhinum majus* [35], *Helianthus annuus* [36], *Vitis vinifera* [37,38], and *Olea europaea* [39,40]. We constructed a maximum likelihood phylogenomic tree using RAxML (v8.2.12) with the PROTGAMMAJTT model. The divergence time between plant species was calculated using PAML (v4.9). Data from *V. vinifera*, *H. annuus* (112.4–125.0 million years [Mya]), *H. annuus*, *O. europaea* (97.5–107.3 Mya), and *A. majus*, and *C. canephora* (72.4–104.9 Mya) were used for calibration.

2.5. Identification and phylogenetic analysis of UGT genes

BLASTP searches in the *P. forrestii* transcriptome database were performed using 107 *Arabidopsis thaliana* and 5 *Zea mays* UGT protein sequences belonging to the *Zm*O, *Zm*P and *Zm*Q groups as queries. The abovementioned protein sequences were first submitted to Pfam (http://pfam.xfam.org/) to confirm conserved domains and then submitted to MEME (https://meme-suite.org/meme/) to determine the PSPG motif.

Phylogenetic trees were constructed using the maximum likelihood method and the IQ-TREE (v. 2.0.3). Amino acid sequences were aligned using MAFFT software. ITOL (https://itol.embl.de/) was used to enhance the evolutionary trees. The following functional genes were used as query sequences to screen for homologous sequences: PAL (AAC18870.1, AED90714.1, and AED90715.1), CHS (NP_001312634.1, AAB35812.1, and AAB88208.1), 4CL (NP_001312667.1, BAD90937.1, and NP_001333770.1), CHI (CAB82707.2 and BAE48085.1), C4H (NP_180607.1, BAI77481.1, and AAB58356.1), FNS (ACV65037.1, EOY17412.1, and AMQ91113.1), and flavonoid 3′-hydroxylase (F3′H) (ADM26615.1, AAG49298.1, ACN65827.1, and ABA64468.1).

2.6. Molecular cloning, expression, and purification of PfUGTs

Total RNA from a mixture of the three tissues was reverse-transcribed using a PrimeScript RT Reagent Kit (TransGen Biotech, Beijing, China) to obtain cDNA. The cDNA of PfUGTs was amplified using specific primers (Table S1) and cloned into pET32a (+) vector at *Bam*HI (5′) and *Sal*I (3′) sites using ClonExpress II One-Step Cloning Kit (Vazyme Biotech, Nanjing, China) according to the manufacturer's instructions. Recombinant plasmids harboring the PfUGT genes were transformed into chemically competent cells of *Escherichia coli* BL21 (DE3) (TransGen Biotech, Beijing, China). We selected a positive colony to incubate in 40 mL of LB culture medium (50 μg/mL ampicillin) at 37 °C until the A_{600} reached 0.6, the temperature was reduced to 16 °C for 24 h, followed by induction with 0.3 mM isopropyl β-D-thiogalactoside. The cells were harvested via centrifugation at 7104 ×*g* at 4 °C for 5 min and were resuspended in 10 mM phosphate-buffered saline (PBS, pH 8.0). The recombinant proteins were purified using a His-tag Protein Purification Kit (Beyotime, Beijing, China) according to the manufacturer's instructions.

2.7. Functional characterization of PfUGTs

Enzymatic reactions were performed using reaction mixtures (100 μL) comprising 50–100 μg pure enzyme, 0.4 mM substrate, 1 mM UDP-glucose, and 10 mM PBS (pH 8.0). The reaction was initiated by adding the substrate, followed by incubation for 2 h at 37 °C. The reaction was terminated by adding 100 μL of methanol. The reactions for the kinetic analysis contained 10 μg pure enzyme and substrates at 0.005 mM, 0.010 mM, 0.015 mM, 0.020 mM, 0.030 mM, 0.040 mM, 0.08 mM, 0.10 mM, 0.15 mM, 0.20 mM, and 0.5 mM. The mixtures were centrifuged at 20,929 ×*g* at room temperature for 15 min and filtered through a 0.2-μm filter membrane. A 3 μL aliquot of sample was injected at a flow rate of 0.25 mL/min into the Nexera UPLC LC-20A system (SHIMADZU, Japan) equipped with a Waters Acquity UPLC BEH C18 column (1.7 μm, 2.1 ×100 mm). Acetonitrile (A) and water containing 0.1% formic acid (v/v, B) were used as mobile phases. The mobile phase gradient was as follows: 5% A for 0 min, 5% A for 2 min, 18% A for 10 min, 95% A for 12 min, 95% A for 14 min, and 5% A for 14.5 min. The detection wavelength was 340 nm, and the column temperature was 40 °C. A sample volume of 1 μL was injected into an Agilent Technologies 1290 Infinity II and 6530 Q-TOF (Agilent Technologies, CA, USA) for MS analysis. The negative ionization mode was used with a scan range of 50–1000 *m/z*. Metabolites were identified using automated MS/MS, and their retention times were compared with those of standard solutions. The ion fragments were produced at a collision energy of 30 V. Other parameters were as follows: VCap, 2000 V; desolvation temperature, 300 °C; flow rate, 6 L·min^{-1}; sheath gas, supplied at 300 °C with a flow rate of 11.0 L·min^{-1}; and nebulizer pressure, 30 psi.

3. Results

3.1. Full-length transcriptome of P. forrestii

The full-length transcriptome of *P. forrestii* generated 50,093,160 subreads with an N50 length of 2075 bp, 1170,243 CCS reads with an N50 length of 2259 bp, and 1059,088 FLNC reads with an N50 length of 2196 bp. After correction and redundancy reduction, 24,366 non-redundant high-quality transcripts with an N50 length of 2399 bp were obtained. After extracting the longest representative transcripts, we predicted 24,123 ORFs with an N50 length of 1458 bp (Table S2). Gene annotation numbers and rates revealed that 22,444 (93%), 22,433 (92%), 18,570 (76%), 18,516 (76%), 14,311 (59%), 12,083 (50%), and 9414 (39%) ORFs, were annotated in TrEMBL, Nr, Pfam, Swiss-Prot, KOG, GO, and KEGG public databases, respectively (Fig. 1A). Furthermore, 3962 ORFs related to the biosynthesis of oligosaccharides and polysaccharides and responses to stimuli were annotated as follows in four specialized databases: 98 in CARD; 1428 in CAZy; 1806 in PHI; and 711 in VFDB (Fig. 1B). Of the annotated ORFs in CAZy, 47.3% (650/1428) were glycosyltransferases (Fig. 1B; Table S3), indicating their potential involvement in the biosynthesis of various glycosides in *P. forrestii*.

We observed WGD events in *P. forrestii* and established a phylogenetic relationship between *P. forrestii* and ten representative plant species of Gentianales (*C. canephora*, *G. jasminoides*, *A. syriaca*, *C. gigantea*, *C. roseus*, and *G. dahurica*), Lamiales (*A. majus* and *O. europaea*), Asterales (*H. annuus*), and Vitales (*V. vinifera*). The distribution of synonymous substitutions per synonymous site (K_S) for paralogs revealed a single peak for *P. forrestii* paralogs, representing the ancestral whole-genome triplication event (γ) shared by core eudicots. No recent WGD events were observed for *P. forrestii* (Fig. 1C). We identified 26,079 orthologous groups based on the full-length transcriptome of *P. forrestii* and whole-genome data from the ten other species. Accordingly, we selected 299 single-copy orthologs to construct a phylogenetic tree, which revealed that *P. forrestii* was closely related to *A. syriaca* and *C. gigantea*, with an estimated divergence time of approximately 36.9 Mya. All three species belonged to the Apocynaceae family (Fig. 1D).

3.2. Analysis of metabolome profiles and flavonoid biosynthesis-related gene expression in different tissues of P. forrestii

A widely targeted metabolomic analysis identified 525 metabolites in the roots, stems, and leaves of *P. forrestii* (Table S4). Flavonoids accounted for 25.3% (133) of the identified metabolites, as the second largest class of secondary metabolites, after phenolic acids, which accounted for 35.5% (187) of the metabolites, followed by terpenoids (72, 13.7%), alkaloids (60, 11.4%), lignans (42, 8.0%), others (20, 3.8%), quinones (5, 0.95%), steroids (4, 0.8%), tannins (2, 0.4%), and

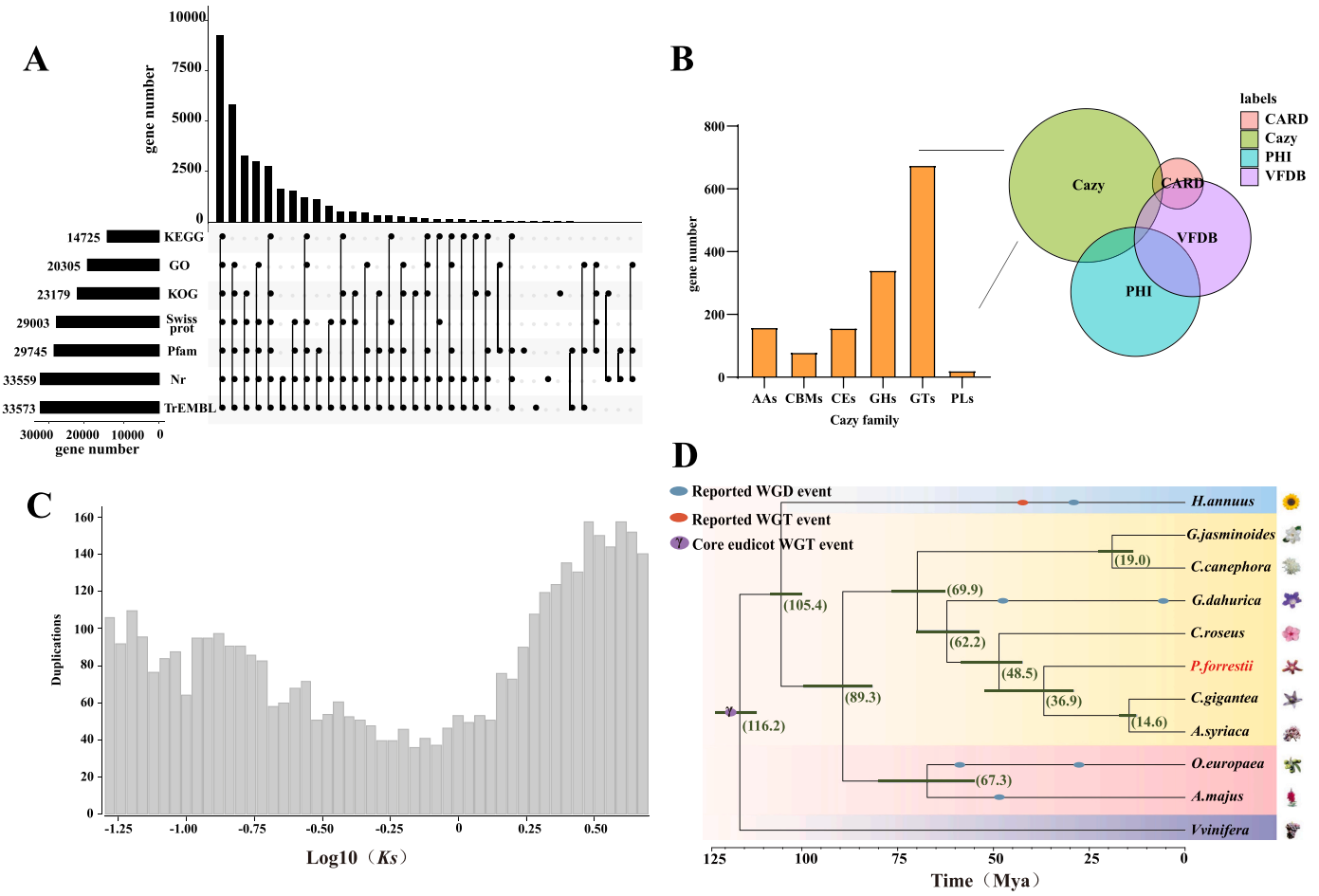

Fig. 1. Gene annotation and whole-genome duplication event analysis of *Periploca forrestii* . A. The number of protein-coding genes annotated using KEGG, GO, KOG, Swiss-Prot, Pfam, Nr, and TrEMBL databases. B. Venn diagram showing the gene annotation results in CARD, CAZy, PHI, and VFDB databases; the orange histogram represents the gene profiles annotated in the CAZy database, including auxiliary activities (AAs), carbohydrate esterases (CEs), glycoside hydrolases (GHs), glycosyl transferases (GTs), and polysaccharide lyases (PLs). C. The K_S distribution in *P. forrestii*. D. The phylogenetic tree of *P. forrestii* and 10 related or representative species based on single-copy genes. The digits represent divergence times; blue circles represent reported WGD; orange circles represent whole-genome triplication (WGT) events; and purple circles represent the core eudicot WGT (γ) event.

coumarins (1, 0.2%). The identified flavonoids corresponded to 1 aurone, 1 chalcone, 9 flavanols, 4 flavanones, 1 flavanonol, 40 flavones, 71 flavonols, 5 isoflavones, and 1 other flavonoid (Fig. 2A, Table S5). A total of 128, 132, and 132 flavonoids were identified in the roots, stems, and leaves, respectively, with 127 flavonoids shared among the three tissues. Flavonoid content was higher in the leaves than in stems and roots (Table S5). Differential accumulation analysis revealed 52 upregulated and 35 downregulated flavonoids in the roots and 9 upregulated and 13 downregulated flavonoids in the stems, indicating a varied accumulation pattern across different tissues (Fig. 2B). Most flavonoids (120, 90.22%) were glycosylated derivatives of aglycones, including quercetin, kaempferol, luteolin, and apigenin. The flavonoid glycosides included 62 monoglycosides, 49 diglycosides, and 9 triglycosides, based on the number of sugars, and more than half of the flavonoid glycosides contained glucosyl moieties. These glycosides were further categorized as 21 *C*-, 96 *O*-, and 3 *C*-*O*-glycosides, based on the connection between the sugar donor and the flavonoid skeleton (Table S5).

We performed RNA-seq analysis of the roots, stems, and leaves using a short-read sequencing protocol to determine gene expression patterns in *P. forrestii*. A total of 24,357 unigenes were obtained with an overall GC content of 39.45% and an N50 length of 2399 bp. The expression levels of these unigenes were calculated as fragments per kilobase of the exon model per million mapped fragments (FPKM). We identified 6237 DEGs by screening genes in the roots vs. leaves, stems vs. leaves, and roots vs. stems. We found that the expression of 1905 and 841 DEGs was upregulated and downregulated in roots vs. leaves, the expression of 797 and 2781 DEGs was upregulated and downregulated in stems vs. leaves, and 1958 and 1328 DEGs were upregulated and downregulated in roots vs. stems, respectively. GO enrichment analysis revealed that most DEGs of roots vs. leaves, stems vs. leaves, and roots vs. stems were enriched in the “chloroplast thylakoid membrane” category (Fig. S1A–C). In addition, the DEGs annotated in the KEGG database were significantly enriched in the “photosynthetic” pathway in stems vs. leaves and roots vs. leaves (Fig. S1D, F). In contrast, DEGs in roots vs. stems were enriched in “carbon metabolism” and “biosynthesis of amino acids” pathways (Fig. S1E).

We attempted to identify enzyme-encoding genes with high expression levels using the full-length transcriptome database for the production of flavonoid glycosides in *P. forrestii*. These genes encoded four PALs, seven C4Hs, five 4CLs, two CHIs, two FNSs, five FLSs, one F3H, and seven F3′Hs (Fig. 2C and 2D, Table S6).

3.3. Identification and phylogenetic analysis of UGT genes in P. forrestii

The ORFs of the full-length transcriptome were used to identify the *P. forrestii* UGT gene family. In total, 427 ORFs were extracted using BLAST. To ensure capture more *PfUGT* genes, all 427 sequences were characterized for features of the UGT gene family, although some had

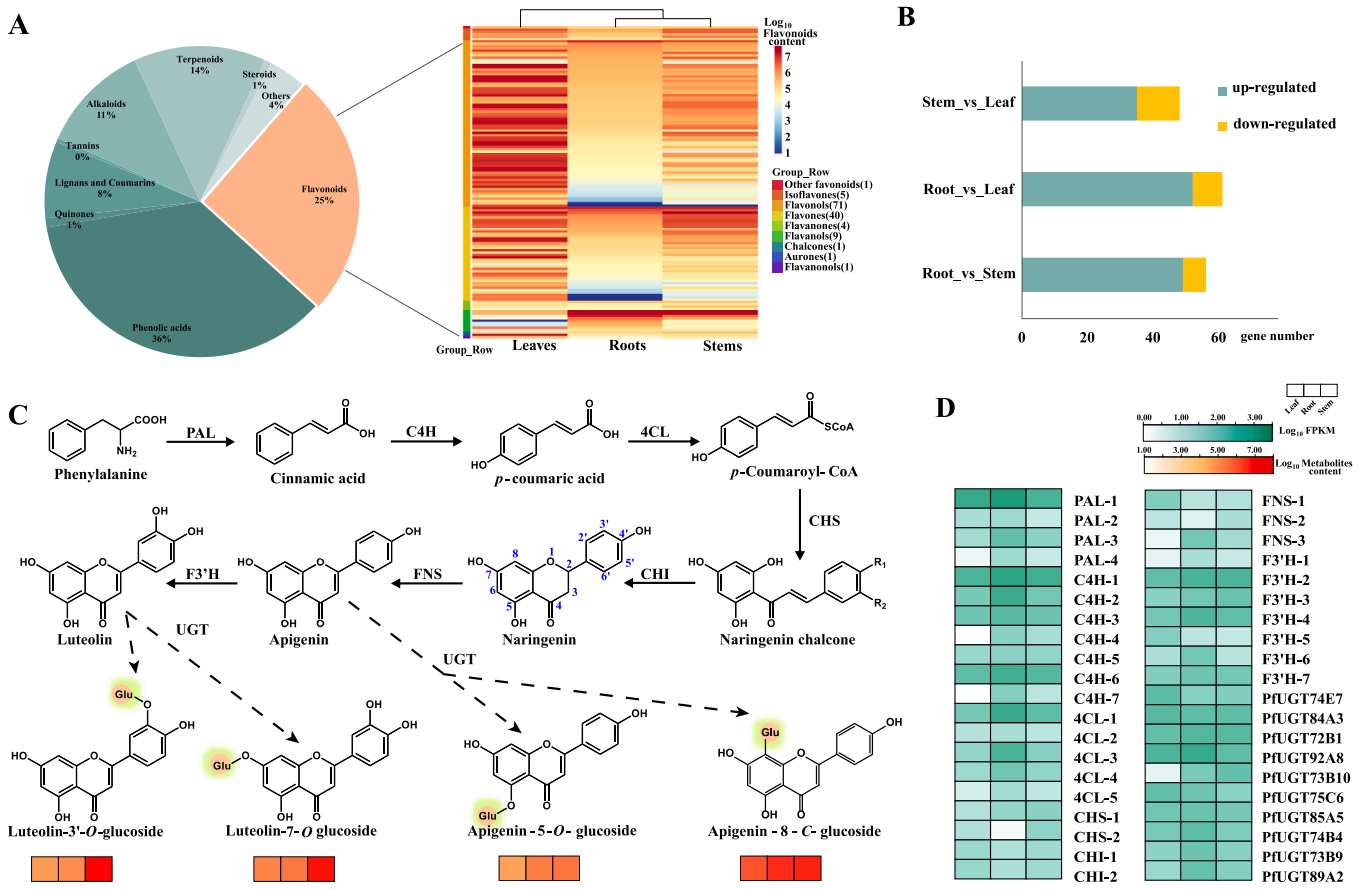

Fig. 2. Transcriptomics and widely targeted metabolome analyses of *Periploca forrestii*. A. Classification of the detected metabolites (left). Heatmap (right) showing the contents of different flavonoid compounds in the leaves, roots, and stems of *P. forrestii*. B. Upregulated and downregulated flavonoid metabolites in stem vs. leaves, roots vs. leaves, and roots vs. stems. C. Proposed biosynthetic pathway of flavonoid glycosides in *P. forrestii*. The heatmap beneath the structures of compounds represents their content in the roots, stems, and leaves. D. Expression heatmap of flavonoid biosynthesis-related genes with log10 FPKM in the roots, stems, and leaves.

low similarity to AtUGTs. After removing unfeatured sequences, 99 genes containing the PSPG signature motif near the C-terminus were identified in the full-length transcriptome database of *P. forrestii* (Table S8, Fig. S2). Sequence alignment revealed that the proteins encoded by all 99 genes contained conserved domains, with lengths ranging from 300 to 556 amino acid residues. Recent studies have indicated that UGTs from *A. thaliana* and maize (*Zea mays*) can be classified into 17 groups (A–Q)[41,42] and that a considerable increase was observed in the expansion of UGTs from E and L groups[43]. A phylogenetic tree was constructed using 99 PfUGT genes, 107 AtUGT genes, and 8 ZmUGT, and the classification of PfUGTs based on the clustering results of the tree indicated that 99 PfUGT genes were clustered into 14 groups, with most of them being clustered into Groups E (22), L (16), and A (13) groups. Other genes were clustered into Groups O (11), B (6), C (2), D (4), F (1), G (8), H (2), J (1), K (2), M (5), and P (5). However, Groups N, I, and Q did not show any clustering of PfUGT genes (Fig. 3A). Compared with *A. thaliana* and *Z. mays*, PfUGT genes from the O group showed evident expansion. We speculate that the expansion of PfUGT genes resulted from tandem or segmental duplication[44,45], as no recent WGD events have been reported in *P. forrestii*.

Several UGT groups, including A, B, D, E, F, L, and K, are associated with flavonoid glycosylation. We analyzed the expression levels of the 99 PfUGT genes to gain further insight into the biosynthesis of flavonoid glycosylation in *P. forrestii*. Among them, ten PfUGT genes (*PfUGT72B1*, *PfUGT73B10*, *PfUGT74B4*, *PfUGT73B9*, *PfUGT75C6*, *PfUGT74E7*, *PfUGT92A8*, *PfUGT89A2*, *PfUGT84A3*, and *PfUGT85A5*) with the highest FPKM values were selected as candidate genes (Fig. 2D, Table S6). The amino acid sequences encoded by these ten candidate PfUGT genes (Table S7) and 30 characterized genes, including 3-*O*, 5-*O*, 7-*O*, *C*-, and multisite glycosyltransferases from previous studies, were used to construct a functional phylogenetic tree (Fig. 3B, Table S7). The results revealed that the seven PfUGT genes can be classified into four groups: *C*GT (1), 5-*O* (1), 7-*O* (1), and multifunctional (4) glycosyltransferases. No PfUGT genes were clustered with the 3-OH branch of the phylogenetic tree; however, *PfUGT84A3*, *PfUGT75C6*, and *PfUGT89A2* showed the potential to carry out the catalytic reaction of flavonoid *C*-, 5-OH, and 7-OH glycosylation, respectively. In addition, *PfUGT74B4*, *PfUGT74E7*, and *PfUGT85A5* did not cluster into any group. All ten PfUGTs were selected for further analysis of their catalytic activity in *P. forrestii*.

3.4. Functional characterization of candidate UGTs of *P. forrestii* in vitro

Of the ten candidate PfUGT genes with the highest FPKM values, eight (*PfUGT72B1*, *PfUGT84A3*, *PfUGT73B9*, *PfUGT73B10*, *PfUGT75C6*, *PfUGT74E7*, *PfUGT74B4*, and *PfUGT92A8*) were successfully cloned and expressed. Because glucosyl-flavonoids are the most abundant metabolites of *P. forrestii*, catalytic reactions were performed using UDP-glucose as a sugar donor and apigenin and luteolin as glucosyl acceptors. Compared to the negative control (Fig. S3), PfUGT75C6 and PfUGT74E7 produced one monoglycoside (1 A), when apigenin was used as the glucosyl acceptor. However, PfUGT74B4 and PfUGT92A8 produced one monoglycoside (1 A) and one diglycoside (2 A). Product 1 A was identified as apigenin-7-*O*-glucoside after comparison of its retention time

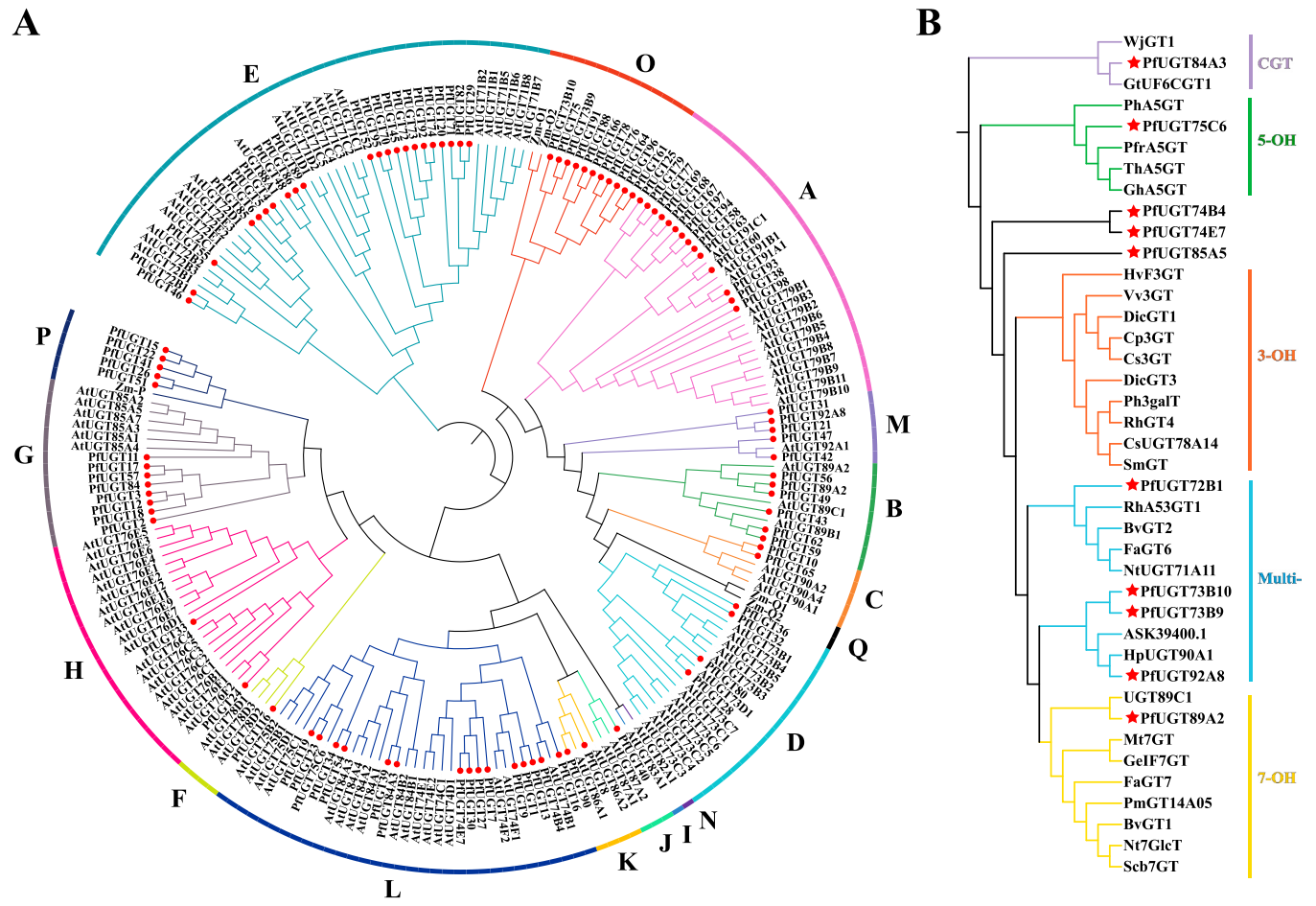

Fig. 3. Phylogenetic analysis of UGTs in *Periploca forrestii*. A. Phylogenetic tree showing UGTs in *Periploca forrestii*, *Arabidopsis thaliana*, and *Zea mays*. B. Phylogenetic tree of previously characterized UGTs involved in flavonoid biosynthesis and 10 PfUGTs isolated in this study (highlighted using red dots). The functional protein sequences are listed in Table S7.

with that of the standard solutions and evaluation of the second-order mass spectra (Fig. 4A). PfUGT75C6 and PfUGT74E7 produced only one product (2 L), which was identified as luteolin-7-*O*-glucoside, when luteolin was used as the glucosyl acceptor. PfUGT74B4 and PfUGT92A8 exhibited multisite glycosylation or consecutive glycosylation activity toward luteolin. PfUGT74B4 produced luteolin-7-*O*-glucoside (2 L) and four new peaks (1 L, 3 L–5 L), whereas PfUGT92A8 produced luteolin-7-*O*-glucoside (2 L) and five new peaks (1 L, 3 L–6 L) (Fig. 4B). The glycosylated products 1 L and 3 L were identified as mono-*O*-glycoside luteolin based on the characteristic fragment ion ([M-H-162]$^-$), 4 L and 5 L were identified as di-*O*-glycoside luteolin based on the fragment ions ([M-H-162]$^-$) and ([M-H-162–162]$^-$), and 6 L was identified as tri-*O*-glycoside luteolin based on the characteristic fragment ions ([M-H-162]$^-$), ([M-H-162–162]$^-$), and ([M-H-162–162-162]$^-$). In addition, the four recombinant PfUGT proteins, i.e., PfUGT72B1, PfUGT84A3, PfUGT73B9, and PfUGT73B10, did not exhibit any significant glycosyltransferase activity toward apigenin or luteolin (Fig. S3). Comparison of enzyme properties of PfUGT74B4, PfUGT92A8, PfUGT75C6 and PfUGT74E7 showed that the affinity of PfUGT74B4 and PfUGT92A8 for apigenin ($K_{m\ PfUGT74B4}$ = 50.23 μM, $K_{m\ PfUGT92A8}$ = 58.28 μM) and luteolin ($K_{m\ PfUGT74B4}$ = 147.6 μM, $K_{m\ PfUGT92A8}$ = 251.7 μM) was lower than PfUGT75C6 and PfUGT74E7 (apigenin: $K_{m\ PfUGT75C6}$ = 29.03 μM, $K_{m\ PfUGT74E7}$ = 5.97 μM; luteolin: $K_{m\ PfUGT75C6}$ = 28.4 μM, $K_{m\ PfUGT74E7}$ = 7.75 μM), whereas the values of specific activity were the opposite, which may be associated with their corresponding number of products (Fig S4, Table 1).

3.5. Substrate promiscuity of PfUGT74B4 and PfUGT92A8 in vitro

We examined the substrate selectivity of the recombinant proteins based on the multiple glycosylation activities of PfUGT74B4 and PfUGT92A8. Seven diverse flavonoids, including one chalcone (2-hydroxynaringenin), five flavones (vitexin, isovitexin, acacetin, orientin, and isoorientin), and one dihydrochalcone (phloretin) were selected as sugar acceptors, with UDP-glucose serving as a sugar donor. The results indicated that both PfUGT74B4 and PfUGT92A8 exhibited glycosylation activity toward all seven substrates (Fig. 5). 2-Hydroxynaringenin is a chalcone with a common chemical scaffold 1,3-diaryl-2-propen-1-one. PfUGT92A8 and PfUGT74B4 catalyzed the conversion of 2-hydroxynaringenin into two mono-*O*-glycoside products (6a and 6b) with *m/z* 287.0560 [M-H-162]$^-$. In addition, PfUGT92A8 generated two additional di-*O*-glycosides (6c and 6d) with *m/z* 449.1080 [M-H-162]$^-$ and 287.0554 [M-H-162–162]$^-$ as well as an additional tri-*O*-glycoside (6e) with *m/z* 611.1598 [M-H-162] and 449.1100 [M-H-162–162]$^-$ (Fig. 5A).

Acacetin is a flavone aglycone, whereas vitexin, isovitexin, orientin, and isoorientin are *C*-glycosides. Both PfUGT92A8 and PfUGT74B4 catalyzed the synthesis of the same products using these flavones as substrates. Compared with apigenin, acacetin contains an additional methoxy group on C-4′. Only one product (3a) was generated, which was identified as mono-*O*-glycoside-acacetin. PfUGT92A8 and PfUGT74B4 produced a single product, i.e., 1a or 2a, respectively, when vitexin or isovitexin was used as the substrate. The fragmented ions of 1a showed *m/z* 431.0972 [M-H-162]$^-$ and 311.0588 [M-H-162–120]$^-$, whereas 2a

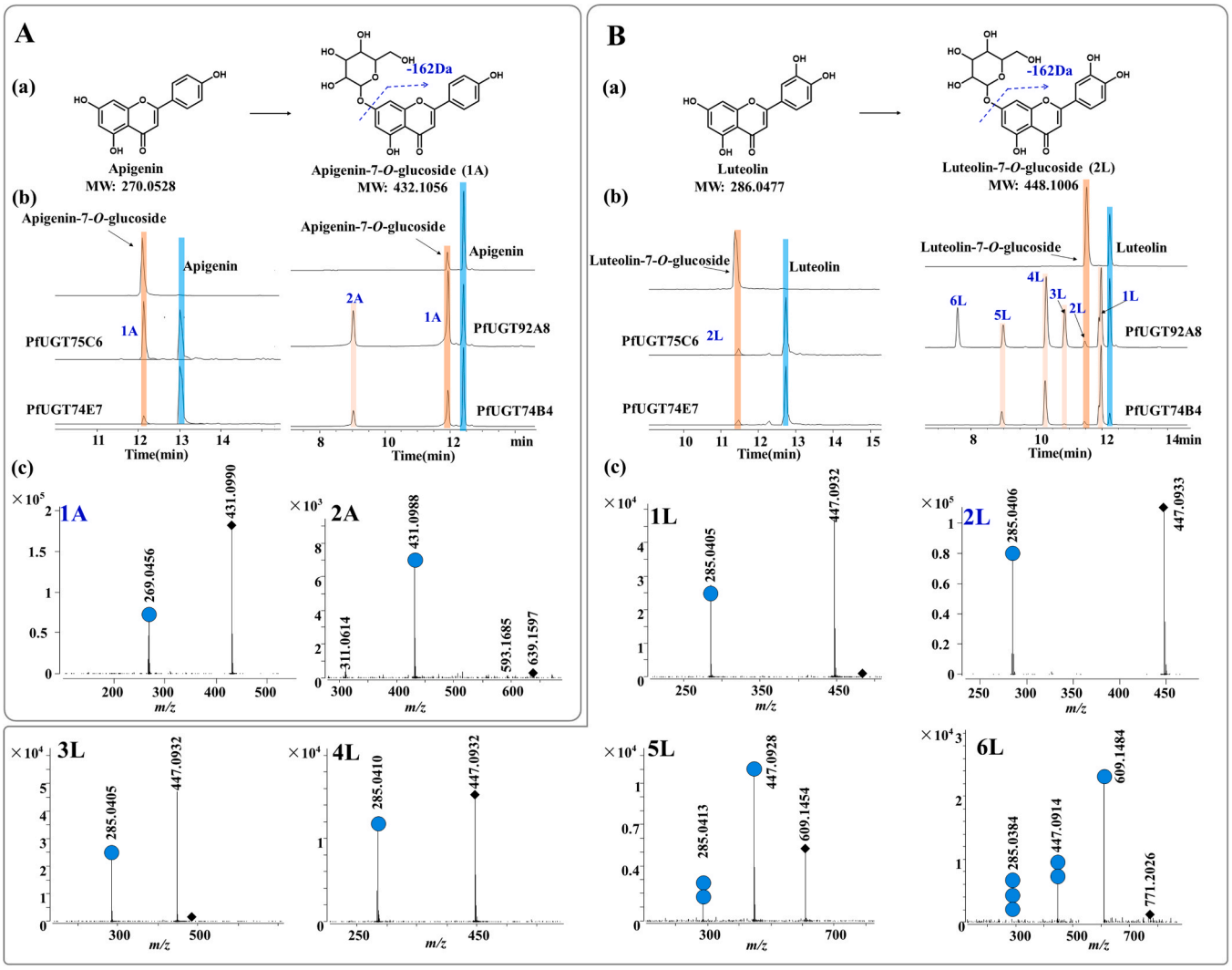

Fig. 4. Functional characterization of recombinant proteins PfUGT75C6, PfUGT74E7, PfUGT74B4, and PfUGT92A8 from *P. forrestii*. A. UPLC–TOF–MS/MS analysis of apigenin reaction mixtures catalyzed by PfUGT75C6, PfUGT74E7, PfUGT74B4, and PfUGT92A8. (a) Glycosylation of apigenin resulting in the formation of apigenin-7-*O*-glucoside (1 A); (b) Extracted ion chromatogram (EIC) of apigenin reaction mixtures catalyzed by recombinant proteins; (c) (−)-MS/MS spectrum of products in 1 A and 2 A. The blue circles indicate a decrease of 162 in molecular mass. B. UPLC–TOF–MS/MS analysis of luteolin reaction mixtures catalyzed by PfUGT92A8, PfUGT74B4, PfUGT92A8, and PfUGT74B4 proteins. (a) Glycosylation of luteolin resulting in the formation of luteolin-7-*O*-glucoside (2 L); (b) EIC of luteolin reaction mixtures catalyzed by recombinant proteins; (c) (−)-MS/MS spectra of products 1 L–6 L. The blue circles indicate a decrease of 162in molecular mass.

showed *m/z* 473.1084 $[M\text{-}H\text{-}120]^-$ and 311.0561 $[M\text{-}H\text{-}162\text{–}120]^-$, indicating the addition of one *O*-glycoside. Three new products (4a, 4b, and 4c for orientin and 5a, 5b, and 5c for isoorientin) when orientin or isoorientin was used as the substrate. Based on the fragmented ions in the mass spectrum, 4a, 4b, 5a, and 5b added one *O*-glycoside corresponding to the substrate (orientin or isoorientin), with characteristic fragment ions showing *m/z* 609.1451 $[M\text{-}H]^-$ and 447.0933 $[M\text{-}H\text{-}162]^-$. However, 4c and 5c added two *O*-glycosides, with characteristic fragment ions showing *m/z* 771.1978 $[M\text{-}H]^-$, 609.1451 $[M\text{-}H\text{-}162]^-$, and 447.0933 $[M\text{-}H\text{-}162\text{–}162]^-$ (Fig. 5B). PfUGT92A8 and PfUGT74B4 catalyzed the conversion of phloretin into one product of mono-*O*-glycoside (7a) with *m/z* 273.0772 $[M\text{-}H\text{-}162]^-$; however, PfUGT92A8 produced an additional di-*O*-glycoside (7b) with *m/z* 435.129 $[M\text{-}H\text{-}162]^-$ (Fig. 5C). Although the site at which the glucosyl group was added to the flavonoid remains unknown, the characteristic fragment ions indicated that PfUGT74B4 and PfUGT92A8 exhibit substrate promiscuity with glycosylation activity toward flavones, chalcones, and dihydrochalcones.

Substrate specificities of PfUGT74B4 and PfUGT92A8 were determined using nine flavonoid substrates and UDP-glucose donors (Fig. S4, Table 1). PfUGT74B4 had the highest affinity and catalytic efficiency toward acacetin (K_m = 2.23 μM, K_{cat}/K_m = 7.13 ×10^3 mM^{-1} sec^{-1}), and PfUGT92A8 had the highest affinity and catalytic efficiency toward phloretin (K_m = 34.7 μM, K_{cat}/K_m = 7.13 ×10^3 mM^{-1} sec^{-1}). These results indicated that these enzymes have different substrate preferences, which may be attributed to the differences in their structures.

4. Discussion

Members of the Apocynaceae family produce large quantities of flavonoid glycosides [6]; however, the molecular mechanisms underlying this structural diversity remain unknown. *P. forrestii* is a vine shrub that produces diverse flavonoid glycosides, and our metabolomic data revealed that > 90% of the flavonoids in *P. forrestii* are glycosylated. Thus, *P. forrestii* may be an ideal model for understanding complex glycosylation mechanisms. In the present study, the full-length transcriptome provided a robust genetic resource for identifying new

Table 1
Glycosyltransferase activity of purified recombinant PfUGT74B4, PfUGT92A8, PfUGT75C6 and PfUGT74E7 toward related flavonoids and corresponding kinetic parameters.

Enzyme	Flavonoid type	Sugar acceptor	Specific activity (pkat mg^{-1})	K_m (μM)	K_{cat} (sec^{-1})	K_{cat}/K_m (10^3 mM^{-1} sec^{-1})
PfUGT74B4	2-Hydroxyflavanones	2-Hydroxynaringenin	20.83	47.83	23.73	4.96
	Flavones	Vitexin	7.86	307.1	8.95	0.03
		Isovitexin	12.21	58.6	13.91	0.24
		Acacetin	13.99	2.23	15.93	7.13
		Orientin	trace	-	-	-
		Isoorientin	9.71	65.21	11.06	0.17
		Apigenin	111	50.23	126.43	2.52
		Luteolin	115.2	58.28	131.21	2.25
	Dihydrochalcone	Phloretin	13.97	45.55	15.91	0.35
PfUGT92A8	2-Hydroxyflavanones	2-Hydroxynaringenin	40.37	51.36	52	1.01
	Flavones	Vitexin	12.28	94.15	15.82	0.17
		Isovitexin	92.96	78.24	119.75	1.53
		Acacetin	trace	-	-	-
		Orientin	439.8	218.1	566.55	2.6
		Isoorientin	137	134.1	176.48	1.32
		Apigenin	217.1	147.6	279.67	1.89
		Luteolin	322.5	251.7	415.44	1.65
	Dihydrochalcone	Phloretin	19.73	34.7	25.47	7.32
PfUGT75C6	Flavones	Apigenin	84.06	29.03	100.47	3.46
		Luteolin	38.1	28.4	45.54	1.6
PfUGT74E7	Flavones	Apigenin	5	5.97	5.73	0.96
		Luteolin	5.45	7.75	6.25	0.81

P. forrestii genes and for future molecular breeding of the species. In addition, no WGD events were observed in the full-length transcriptome of *P. forrestii*. The shared ancestral paleohexaploid (γ) event by core eudicots suggested that the unique flavonoid diversity in Apocynaceae could be influenced by lineage-specific tandem gene duplication and neofunctionalization, particularly among the UGT superfamily members.

We first identified the UGTs involved in flavonoid glycoside biosynthesis in Apocynaceae species using the full-length transcriptome of *P. forrestii*. Four UGTs (PfUGT74B4, PfUGT74E7, PfUGT75C6, and PfUGT92A8) exhibited glycosylation activity toward apigenin and luteolin *in vitro* , whereas the other four enzymes did not. This may be attributed to our inability to fully mimic the in *vivo* environment of the plants; or to the limited substrates and sugar donors in our study, which were not suitable for the other four UGTs. In addition, PfUGT75C6 and PfUGT74E7 showed considerable regiospecificity toward the 7-OH group of apigenin and luteolin, whereas PfUGT74B4 and PfUGT92A8 showed high promiscuity. Interestingly, PfUGT75C6 clustered with the 5-OH branch of a phylogenetic tree constructed using previously documented UGTs involved in flavonoid biosynthesis (Fig. 3B). This is in contrast to our findings regarding the 7-OH glycosylation function. Similarly, CsUGT75L12 from *Camellia sinensis* and NTGT2 from *Nicotiana tabacum* showed increased similarity with 5-*O*-glucosyltransferase, but were characterized by glycosylation activity toward the 7-hydroxyl group of flavonoids [46,47]. This suggested that the position of the glycosylation modification is not closely related to the relationship of the UGTs in distant species. Furthermore, PfUGT74B4 and PfUGT92A8 showed variable catalytic site selectivity for substrates containing different numbers of phenolic hydroxyl groups. Only one mono--*O*-glycoside and one di-*O*-glycoside products were generated using apigenin, which contains three hydroxyl groups (5,7,4′ position) as acceptors. Three mono-*O*-glycosides and two di-*O*-glycosides were produced by PfUGT74B4 and PfUGT92A8 when luteolin was used as an acceptor, whereas one tri-*O*-glycoside was produced by PfUGT92A8 alone. This finding showed that the 3′-OH of luteolin may influence the catalytic site selectivity of PfUGT74B4 and PfUGT92A8. Furthermore, *C*-glycosylation products account for more than 20% of flavonoid glycosides in *P. forrestii*; however, only a few functional *C*-glycoside glycosyltransferases (*C*GTs) from plants have been reported, and new plant-derived *C*GTs must have been identified[48]. PfUGT84A3 was grouped into a previously described *C*GT branch based on our study results. However, its function was not confirmed, indicating that new *C*GTs may exist in *P. forrestii*. The discovery of PfCGTs provided insights into the diversity of flavonoid glycosides in Apocynaceae.

An increasing number of UGTs have been identified in plants, particularly those with flavonoid acceptors, including the UGT72, UGT75, UGT88, UGT89, and UGT90 families[49,50]; however, members of the UGT74 and UGT92 families exhibit glycosyltransferase activity toward plant-specific metabolites, particularly flavonoids. UGT74AC1 from *Siraitia grosvenorii* catalyzes the glycosylation of quercetin and naringenin without product identification[51], whereas UGT74F1 from *A. thaliana* glycosylates quercetin using UDP-glucose as the substrate at 7-OH and 4′-OH sites[52]. Moreover, only UGT92G2 from *Glycine max* has been identified as a flavonol-specific glycosyltransferase that cannot utilize apigenin or luteolin as substrates[53]. PfUGT74B4 and PfUGT92A8 catalyze multisite or consecutive glycosylations using flavones, chalcones, and dihydrochalcones as substrates. To the best of our knowledge, this is the first study to demonstrate that the UGT92A subfamily exhibits promiscuous glycosyltransferase activity toward flavonoids.

Plant flavonoid glycosides are the main active ingredients of some important traditional Chinese medicines. Glycosylation is often the last step in flavonoid glycoside biosynthesis and not only changes the polarity of flavonoid compounds but also affects their pharmacodynamic activity. The versatility of promiscuous enzymes leads to diverse compounds and facilitates organismal adaptation [54–59]. In this study, two UGTs with promiscuous catalytic activities on various flavonoid skeletons produced diverse flavonoid glycosides (Figs. 4 and 5). These products generally have important pharmacological activities. For example, luteolin-7-*O*-glucoside may be effective against SARS-CoV-2 [60], and apigenin-7-*O*-glucoside has been reported to possess remarkable antispasmodic, anti-inflammatory, antioxidant, and anticarcinogenic properties [61–63]. The lack of regioselectivity and substrate specificity of promiscuous UGTs may result in lower catalytic efficiency for the desired glycosides and the generation of unwanted byproducts compared to specific UGTs. However, promiscuous UGTs provide a starting point for new enzymes that generate different bioactive substances [64]. Simultaneously, mutations in promiscuous enzymes can improve biosynthetic pathway efficiency in metabolic engineering [65,66]. However, the catalytic mechanisms of these extremely promiscuous PfUGTs remain unclear. Protein crystal structure analysis and key amino acid mutations in PfUGT74B4 and PfUGT92A8 will aid in a better

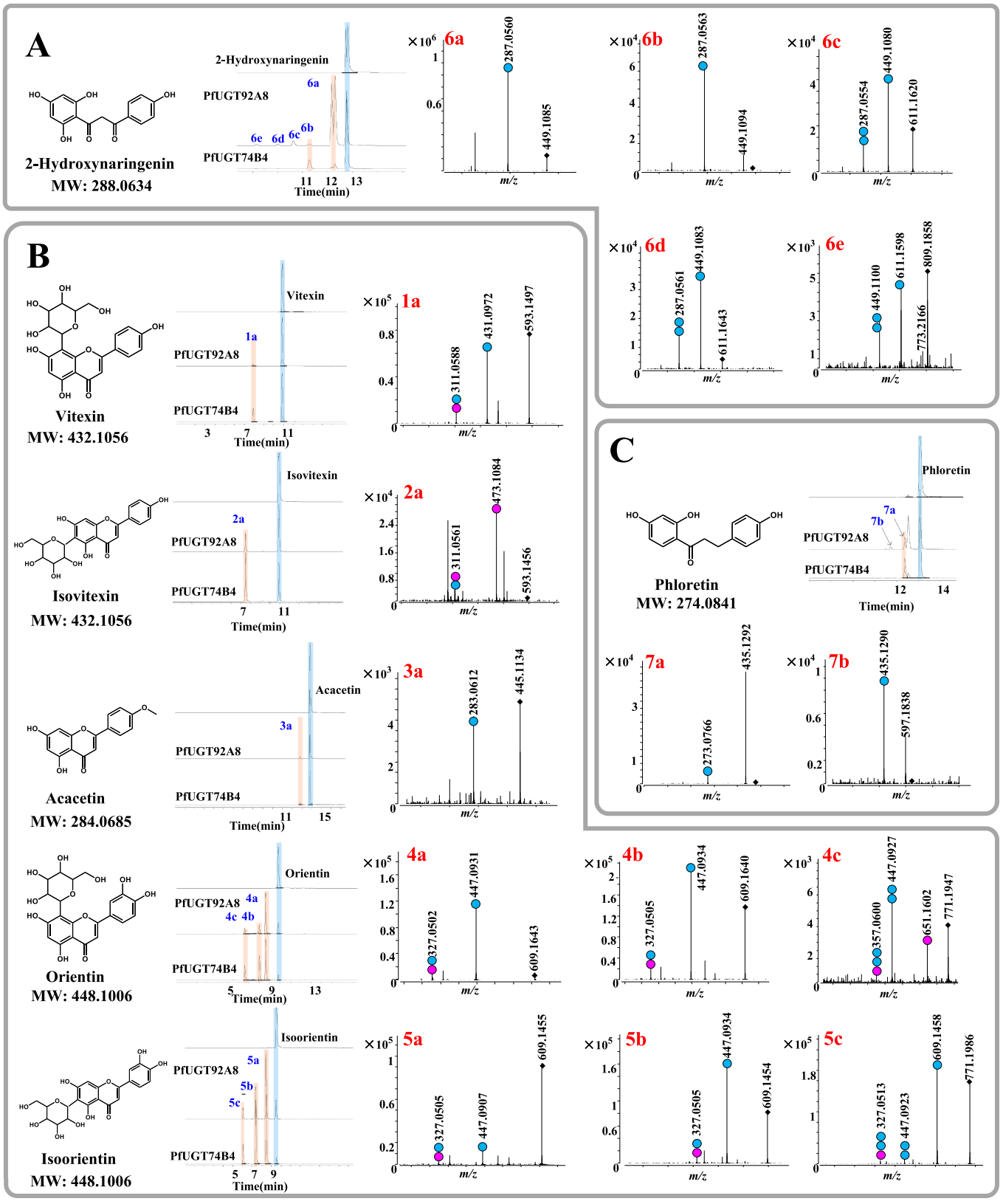

Fig. 5. Promiscuity of PfUGT92A8 and PfUGT74B4 toward flavones, chalcone, and dihydrochalcone. A. Extracted ion chromatogram (EIC) and (−)-MS/MS spectra of the products formed through reactions catalyzed by PfUGT92A8 and PfUGT74B4 using chalcone (2-hydroxynaringenin) as a substrate. B. EIC and (−)-MS/MS spectra of the products formed through reactions catalyzed by PfUGT92A8 and PfUGT74B4 using flavones (vitexin, isovitexin, acacetin, orientin, and isoorientin) as substrates. C. EIC and (−)-MS/MS spectra of the products formed through reactions catalyzed by PfUGT92A8 and PfUGT74B4 using dihydrochalcone (phloretin) as a substrate. The blue circles indicate minus *m/z* 162 molecular mass, whereas the pink circles indicate minus *m/z* 120 molecular mass.

understanding of this promiscuity, thereby providing new insights into the metabolic engineering of novel flavonoid glycosides.

Funding

This work was supported by the National Natural Science Foundation of China (U1812403-1), the CACMS Innovation Fund (CI2021A05103), and the Fundamental Research Funds for the Central Public Welfare Research Institutes (ZZ16-YQ-047, ZZ13-YQ-102).

CRediT authorship contribution statement

Xiaotong Wang and Ranran Gao performed the experiments and wrote this manuscript; Lan Wu, Li Xiang, and Weijun Yang designed and supervised the work; Wanran Zhang, Shi Qiu, and Zhichao Xu analyzed the data; Wenting Wang, Mengyue Wang, Li Xiang, and Jiang He collected the plant samples; and Qinggang Yin and Yuhua Shi revised this manuscript and contributed to the discussion.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The RNA-Seq and full-length transcriptome datasets of *Periploca forrestii* were deposited in the NCBI Sequence Read Archive (SRA) database (bioproject no. PRJNA987534).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.02.028.

References

[1] Zhuang W-B, Li Y-H, Shu X-C, Pu Y-T, Wang X-J, Wang T, et al. The classification, molecular structure and biological biosynthesis of flavonoids, and their roles in biotic and abiotic stresses. Molecules 2023;28:3599.
[2] Shomali A, Das S, Arif N, Sarraf M, Zahra N, Yadav V, et al. Diverse physiological roles of flavonoids in plant environmental stress responses and tolerance. Plants 2022;11:3158.
[3] Badshah SL, Faisal S, Muhammad A, Poulson BG, Emwas AH, Jaremko M. Antiviral activities of flavonoids. Biomed Pharmacother 2021;140:111596.
[4] Ji Y, Li B, Qiao M, Li J, Xu H, Zhang L, et al. Advances on the in *vivo* and in *vitro* glycosylations of flavonoids. Appl Microbiol Biotechnol 2020;104:6587–600.
[5] Yang B, Liu H, Yang J, Gupta VK, Jiang Y. New insights on bioactivities and biosynthesis of flavonoid glycosides. Trends Food Sci Technol 2018;79:116–24.
[6] Bhadane BS, Patil MP, Maheshwari VL, Patil RH. Ethnopharmacology, phytochemistry, and biotechnological advances of family Apocynaceae: a review. Phytother Res 2018;32:1181–210.
[7] Li P. Asclepiadaceae. Flora China 1995;16:189–270.
[8] Byng JW, Chase MW, Christenhusz MJM, Fay MF, Judd WS. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. BOTANICAL J Linn Soc 2016;181:1–20.
[9] Wang Q, Yao X, Ling Y, Huang Y, Chen C. Investigation of the mechanism of *Periploca forrestii* against rheumatoid arthritis with network pharmacology-based analysis. Evid-Based Complement Altern Med 2022;2022:2993374.
[10] Yan X, Chen H, Zhou X. Research progress of *Periploca forrestii* Schltr. Chin Pharm J 2019;54:259–63.
[11] Zhu C, Zhou X, Chen H. Advances on Treatment of Rheumatoid Arthritis with Miao Medicine. Chin J Mod Appl Pharm 2020;37:2669–77.
[12] Chen H, Liang Q, Zhou X, Wang X. Preparative separation of the flavonoid fractions from *Periploca forrestii* Schltr. ethanol extracts using macroporous resin combined with HPLC analysis and evaluation of their biological activities. J Sep Sci 2019;42:650–61.
[13] Zhang NL, Cai JZ, Huang RM, Hu YJ, Xue Y, Qiu SX. Chemical constituents from *Periploca forestti*. Chin Tradit Herb Drugs 2011;42:1909–12.
[14] Sohn US, Lee SE, Lee SH, Nam Y, Hwang WK, Sohn UD. The protective mechanism of quercetin-3-*O*-β-D-glucuronopyranoside (QGC) in H_2O_2-induced injury of feline esophageal epithelial cells. Arch Pharm Res 2016;39:1324–34.
[15] Lin C, Wu M, Dong J. Quercetin-4′-*O*-β-D-glucopyranoside (QODG) inhibits angiogenesis by suppressing VEGFR2-mediated signaling in Zebrafish and endothelial cells. PloS One 2012;7:e31708.
[16] Ngnokam Jouogo DCE, Tamokou Paul. A new phenolic glycoside from the leaves of Flacourtia flavescens Willd. Nat Prod Res 2023:1–11.
[17] Fowler ZL, Koffas MAG. Biosynthesis and biotechnological production of flavanones: current state and perspectives. Appl Microbiol Biotechnol 2009;83:799–808.
[18] Liu W, Feng Y, Yu S, Fan Z, Li X, Li J, et al. The flavonoid biosynthesis network in plants. Int J Mol Sci 2021;22:12824.
[19] Alseekh S, Perez de Souza L, Benina M, Fernie AR. The style and substance of plant flavonoid decoration: towards defining both structure and function. Phytochemistry 2020;174:112347.
[20] Choi SH, Kim HS, Yoon YJ, Kim D-M, Lee EY. Glycosyltransferase and its application to glycodiversification of natural products. J Ind Eng Chem 2012;18:1208–12.
[21] Cui LL, Yao SB, Dai XL, Yin QG. Identification of UDP-glycosyltransferases involved in the biosynthesis of astringent taste compounds in tea (*Camellia sinensis*). J Exp BOTANY 2016;67:2285–97.
[22] Li P, Li Y-J, Zhang F-J, Zhang G-Z, Jiang X-Y, Yu H-M, et al. The arabidopsis UDP-glycosyltransferases UGT79B2 and UGT79B3, contribute to cold, salt and drought stress tolerance via modulating anthocyanin accumulation. Plant J 2017;89:85–103.
[23] Wang Z, Wang S, Xu Z, Li M, Chen K, Zhang Y, et al. Highly promiscuous flavonoid 3-*O*-Glycosyltransferase from *Scutellaria baicalensis*. Org Lett 2019;21:2241–5.
[24] Y. Luo, N. Ding, X. Shi, Y. Wu, Generation and comparative analysis of full-length transcriptomes in sweetpotato and its putative wild ancestor I. trifida, bioRxiv; 2017. 112425.
[25] Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, et al. Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nat Commun 2013;4:1827.
[26] Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11:R106.
[27] Yu GC, Wang LG, Han YY, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics 2012;16:284–7.
[28] Chen W, Gong L, Guo ZL, Wang WS, Zhang HY, Liu XQ, et al. Integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. Mol Plant 2013;6:1769–80.
[29] Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science 2014;345:1181–4.
[30] Xu Z, Pu X, Gao R, Demurtas OC, Fleck SJ, Richter M, et al. Tandem gene duplications drive divergent evolution of caffeine and crocin biosynthetic pathways in plants. BMC Biol 2020;18:63.
[31] Weitemier K, Straub SCK, Fishbein M, Bailey CD, Cronn RC, Liston A. A draft genome and transcriptome of common milkweed (*Asclepias syriaca*) as resources for evolutionary, ecological, and molecular studies in milkweeds and Apocynaceae. PEERJ 2019;7:e7649.
[32] Hoopes GM, Hamilton JP, Kim J, Zhao D, Wiegert-Rininger K, Crisovan E, et al. Genome assembly and annotation of the medicinal plant *Calotropis gigantea*, a producer of anticancer and antimalarial cardenolides. G3 Genes|Genomes|Genet 2018;8:385–91.
[33] Kellner F, Kim J, Clavijo BJ, Hamilton JP, Childs KL, Vaillancourt B, et al. Genome-guided investigation of plant natural product biosynthesis. PLANT J 2015;82:680–92.
[34] Li T, Yu X, Ren Y, Kang M, Yang W, Feng L, et al. The chromosome-level genome assembly of *Gentiana dahurica* (Gentianaceae) provides insights into gentiopicroside biosynthesis. DNA Res 2022;29:1–10.
[35] Li M, Zhang D, Gao Q, Luo Y, Zhang H, Ma B, et al. Genome structure and evolution of antirrhinum majus L. Nat Plants 2019;5:174–83.
[36] Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. Nature 2017;546:148–52.
[37] Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 2007;449:463–7.
[38] Massonnet M, Cochetel N, Minio A, Vondras AM, Lin J, Muyle A, et al. The genetic basis of sex determination in grapes. Nat Commun 2020;11:2902.
[39] Rao G, Zhang J, Liu X, Lin C, Xin H, Xue L, et al. De novo assembly of a new Olea europaea genome accession using nanopore sequencing. Hortic Res 2021;8:64.
[40] Unver T, Wu Z, Sterck L, Turktas M, Lohaus R, Li Z, et al. Genome of wild olive and the evolution of oil biosynthesis. Proc Natl Acad Sci 2017;114:E9413–22.
[41] Li Y, Baldauf S, Lim E-K, Bowles DJ. Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. J Biol Chem 2001;276:4338–43.
[42] Li Y, Li P, Wang Y, Dong R, Yu H, Hou B. Genome-wide identification and phylogenetic analysis of Family-1 UDP glycosyltransferases in maize (*Zea mays*). Planta 2014;239:1265–79.
[43] Caputi L, Malnoy M, Goremykin V, Nikiforova S, Martens S. A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. PLANT J 2012;69:1030–42.
[44] Kawai Y, Ono E, Mizutani M. Expansion of specialized metabolism-related superfamily genes via whole genome duplications during angiosperm evolution, Plant. Biotechnology 2014;31:579–84.

[45] Ramamoorthy R, Jiang SY, Kumar N, Venkatesh PN, Ramachandran S. A comprehensive transcriptional profiling of the WRKY gene family in rice under various abiotic and phytohormone treatments. PLANT CELL Physiol 2008;49:865–79.
[46] Dai X, Shi X, Yang C, Zhao X, Xia T. Two UDP-glycosyltransferases catalyze the biosynthesis of bitter flavonoid 7-*O*-Neohesperidoside through Sequential Glycosylation in Tea Plants. J Agric Food Chem 2022;70:2354–65.
[47] Taguchi G, Ubukata T, Hayashida N, Yamamoto H, Okazaki M. Cloning and characterization of a glucosyltransferase that reacts on 7-hydroxyl group of flavonol and 3-hydroxyl group of coumarin from tobacco cells. Arch Biochem Biophys 2003;420:95–102.
[48] Zhang Y-Q, Zhang M, Wang Z-L, Qiao X, Ye M. Advances in plant-derived *C*-glycosides: phytochemistry, bioactivities, and biotechnological production. Biotechnol Adv 2022;60:108030.
[49] Wilson AE, Tian L. Phylogenomic analysis of UDP-dependent glycosyltransferases provides insights into the evolutionary landscape of glycosylation in plant metabolism. Plant J 2019;100:1273–88.
[50] Liu Y, Wang Q, Liu X, Cheng J, Zhang L, Chu H, et al. pUGTdb: a comprehensive database of plant UDP-dependent glycosyltransferases. Mol Plant 2023;16:643–6.
[51] Dai L, Liu C, Zhu Y, Zhang J, Men Y, Zeng Y, et al. Functional characterization of cucurbitadienol synthase and triterpene glycosyltransferase Involved in biosynthesis of mogrosides from *Siraitia grosvenorii*. Plant Cell Physiol 2015;56:1172–82.
[52] Lim E-K, Ashford DA, Hou B, Jackson RG, Bowles DJ. Arabidopsis glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. Biotechnol Bioeng 2004;87:623–31.
[53] Rehman HM, Nawaz MA, Shah ZH, Chung G, Yang SH. Molecular elucidation of two novel seed specific flavonoid glycosyltransferases in Soybean. J Plant Biol 2018;61:320–9.
[54] Miner BE, Kulling PM, Beer KD, Kerr B. Divergence in DNA photorepair efficiency among genotypes from contrasting UV radiation environments in nature. Mol Ecol 2015;24:6177–87.
[55] Sengupta A, Carrara F, Stocker R. Phytoplankton can actively diversify their migration strategy in response to turbulent cues. Nature 2017;543:555.
[56] Exposito-Alonso M, Vasseur F, Ding W, Wang G, Burbano HA, Weigel D. Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis thaliana*. Nat Ecol Evol 2018;2:352–8.
[57] Kazlauskas RJ. Enhancing catalytic promiscuity for biocatalysis. Curr Opin Chem Biol 2005;9:195–201.
[58] Hughes AL. The evolution of functionally novel proteins after gene duplication. Proc Biol Sci 1994;256:119–24.
[59] Chen K, Arnold FH. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proc Natl Acad Sci USA 1993;90:5618–22.
[60] Yan Y-M, Shen X, Li Y-P, Cao Y-K, Zhang J, Wang Y, et al. Discovery of anti-SARS-CoV-2 agents from 38 Chinese patent drugs toward respiratory diseases via docking screening. Med Plant Biol 2023;2.
[61] Bolzon LB, dos Santos JS, Silva DB, Crevelin EJ, Moraes LAB, Lopes NP, et al. Apigenin-7-*O*-glucoside oxidation catalyzed by P450-bioinspired systems. J Inorg Biochem 2017;170:117–24.
[62] Nasr Bouzaiene N, Chaabane F, Sassi A, Chekir-Ghedira L, Ghedira K. Effect of apigenin-7-glucoside, genkwanin and naringenin on tyrosinase activity and melanin synthesis in B16F10 melanoma cells. Life Sci 2016;144:80–5.
[63] Samet I, Villareal MO, Motojima H, Han J, Sayadi S, Isoda H. Olive leaf components apigenin 7-glucoside and luteolin 7-glucoside direct human hematopoietic stem cell differentiation towards erythroid lineage. Differentiation 2015;89:146–55.
[64] Bornscheuer UT, Kazlauskas RJ. Catalytic promiscuity in biocatalysis: using Old enzymes to form new bonds and follow new pathways. Angew Chem Int Ed 2004;43:6032–40.
[65] Khanal A, McLoughlin SY, Kershner JP, Copley SD. Differential effects of a mutation on the normal and promiscuous activities of orthologs: implications for natural and directed evolution. Mol Biol Evol 2015;32:100–8.
[66] Rahimi M, van der Meer JY, Geertsema EM, Poddar H, Baas BJ, Poelarends GJ. Mutations closer to the active site improve the promiscuous aldolase activity of 4-oxalocrotonate tautomerase more effectively than distant mutations. Chembiochem 2016;17:1225–8.