Synthetic and Systems Biotechnology 14 (2026) 117–127

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: www.keaipublishing.com/en/journals/synthetic-and-systems-biotechnology

Original Research Article

Functional characterization of four glycosyltransferases for biosynthesis of steroidal saponins in medicinal plant *Paris polyphylla*

Hizar Subthain [a,b], Fei Guo [b], Chengjie Zhang [b], Peng Fang [b], Lingli Zhang [b], Qixuan Su [c], Dandan Tang [b], Luping Chi [b], Changning Liu [c], Vlada B. Urlacher [d], Jing Li [c,*], Lei Du [b,**], Shengying Li [b,e,***]

[a] School of Medicine, Qilu Institute of Technology, Jinan, Shandong 250200, China
[b] State Key Laboratory of Microbial Technology, Shandong University, Qingdao, Shandong 266237, China
[c] CAS Key Laboratory of Tropical Plant Resources and Sustainable Use, Yunnan Key Laboratory of Crop Wild Relatives Omics, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming 650223, China
[d] Institute of Biochemistry, Heinrich-Heine-University Düsseldorf, Universitätsstraße 1, Düsseldorf, 40225, Germany
[e] Laboratory for Marine Biology and Biotechnology, Qingdao Marine Science and Technology Center, Qingdao, Shandong 266237, China

ARTICLE INFO

Keywords:
Steroidal saponins
Glucosylation
Glycosyltransferases
Biosynthesis
Paris polyphylla

ABSTRACT

Paris species are renowned for the production of important medicinal steroidal saponins (*e.g.*, polyphyllins). UDP-dependent glycosyltransferases (UGTs) play important roles in saponin biosynthesis. However, the glycosylation steps have not been fully elucidated. Here, we investigate four candidate UGTs (*i.e.*, UGT91BP2, UGT703R1, UGT703R2 and UGT703R3) from the medicinal herb *Paris polyphylla* var. *yunnanensis*, followed by their phylogenetic, *in vitro* biochemical, and *in planta* functional characterization as saponin biosynthetic enzymes. These four recombinant *Pp*UGTs expressed in *Escherichia coli* catalyze the glucosylation of diosgenin to produce diosgenin 3-*O*-glucoside (trillin); while UGT703R1, UGT703R2 and UGT703R3 glucosylate pennogenin to pennogenin 3-*O*-glucoside *in vitro*. Transient expression of these UGTs in *Nicotiana benthamiana* leaves, supplemented with diosgenin substrate, confirms their roles in trillin biosynthesis. Subcellular localization analysis in tobacco cells reveals their presence in both cytoplasm and nucleus. Gene transcript analysis reveals that the four *PpUGTs* exhibit higher expression in leaves and flowers than in stems, indicating distinct tissue-specific patterns. This work identifies key UGTs involved in polyphyllin biosynthesis, enriching the enzymatic toolbox involved in the glucosylation of diosgenin and pennogenin aglycones, and provides a valuable reference for future studies on overproduction of bioactive steroidal glycosides in heterologous hosts.

1. Introduction

Plant saponins are structurally diverse and important bioactive natural products consisting of a steroidal or triterpenoid backbone and mono- or multi-sugar decoration [1]. These specialized molecules not only have environment adaptive properties, but also possess a wide spectrum of biological activities such as antitumor, antibacterial, antifungal, and sedative effects. Thus, plant saponins have extensively been used as pharmaceuticals and nutraceuticals in the health and food sectors [1–5].

Saponins are biosynthesized primarily via the cytosolic mevalonate (MVA) pathway alongside the plastid-localized methylerythritol-phosphate (MEP) route, leading to the key intermediate squalene. Squalene is then converted by squalene epoxidase into 2,3-oxidosqualene, which further undergoes cyclization by cycloartenol synthase (CAS) and through a series of modifications, forming cholesterol and phytosterols. The cholesterol scaffold is tailored further by a variety of cytochrome P450 monooxygenases giving rise to diverse sapogenins. These aglycon intermediates ultimately are glycosylated to produce steroidal saponins and steroidal glycoalkaloids [6–8]. Glycosylation is one of the most

Peer review under the responsibility of Editorial Board of Synthetic and Systems Biotechnology.
* Corresponding author.
** Corresponding author.
*** Corresponding author. State Key Laboratory of Microbial Technology, Shandong University, Qingdao, Shandong, China.
E-mail addresses: lijing3@xtbg.ac.cn (J. Li), lei.du@sdu.edu.cn (L. Du), lishengying@sdu.edu.cn (S. Li).

https://doi.org/10.1016/j.synbio.2026.04.002
Received 25 January 2026; Received in revised form 25 March 2026; Accepted 1 April 2026
Available online 21 April 2026

important modifications mediated usually by UDP-glycosyltransferases (UGTs) of the carbohydrate-active enzyme glycosyltransferase 1 superfamily (GT1, CAZy database, https://www.cazy.org/GlycosylTransferases.html) that catalyze the transfer of sugar moieties to various substrates including sapogenins [9,10].

The glycosylation process endows the steroidal compounds with significant biological activities. For example, oat avenacin A-1 and tomato α-tomatine in their active glycoside forms protect plants against herbivores, pathogenic microorganisms, and competing plant species; triterpene saponin glycosides from *Glycine* max determines the bitterness and anti-feedant activity [11,12]; and *Withania somnifera* sterol glycoside contributes to biotic and abiotic stresses in plants [13]. Triterpene saponins from *Quillaja saponaria* acting as immune potentiators have proven to be highly effective as human vaccine adjuvants [14]. In addition, glycosylation is considered the terminal step for biosynthesis of steroidal saponins, which impacts their aqueous solubility, stability, subcellular compartmentalization, and transportability, thus leading to easier accumulation and storage of these compounds in plant cells [12,15–17]. In particular, biosynthesis of the aglycones of steroidal saponins, for example, diosgenin, is speculated to take place in leaves and then transported through stems to other tissues [18,19]. However, many glycosylation pathways and detailed mechanisms remain obscure. Thus, it is a key to explore the tissue expression profile of UGTs and study their subcellular localization for understanding the process and regulation of saponin metabolisms.

Paris polyphylla var. *yunnanensis* is one of the most important perennial herbs and the plant organs such as the dried rhizome are typically used in Chinese traditional medicines, whereas its fruits and seeds as traditional ingredients are also used for culinary purposes [20,21]. The rich bioactive components of *P. polyphylla* are steroidal sapogenin derivatives, including diosgenin- and pennogenin-derived saponins (*e.g.*, polyphyllins), which account for approximately 80% of its total active compounds [22,23]. The aglycone scaffolds from which polyphyllins are derived, accommodate various sugar moieties including glucose, rhamnose, and arabinose linked to their hydroxyl groups at the C-3 and/or C-26 positions [24,25]. However, the long growth period (7–8 years) of the cultivated *Paris* species and excessive exploitation of the natural habitat, as well as several challenges in the chemical synthesis of steroid saponins, like such factors put major obstacles in both research and medicinal implications of the *Paris* saponins [26,27]. In recent years, metabolic engineering approaches by harnessing the biosynthetic pathway and corresponding enzymes in heterologous hosts have been providing an alternative way for stable production of polyphyllins.

Early research breakthroughs in the biosynthesis of *Paris* saponins have been made in the absence of complete genomic information, but mainly based on the growing transcriptome sequencing data of *Paris* species [28–31]. Identification of the cholesterol pathway-enriched cytochrome P450 genes involved in the major saponin (dioscin) production in *Trigonella foenum-graecum* and *Dioscorea zingiberensis*, especially in the formation of diosgenin and pennogenin (the two main aglycones of *Paris* saponins) in *P. polyphylla*, have further provided broadened access to the subsequent glycosylation steps mediated by UGTs to elucidate the downstream biosynthesis of diverse steroidal saponins [32–35]. This makes the UGTs an attractive target of the corresponding biosynthetic routes. Given the highly diverse and multigenic nature, UGTs are phylogenetically characterized into various sub-families based on their functions. For example, UGT73, UGT80 and UGT91 family genes have been identified to encode glycosyltransferases for glycosylating sterol, steroidal, and/or terpenoid saponins [33,34,36,37]. *Solanum aculeatissimum* glucosyltransferase, *Sa*GT4A, mediates the 3-*O*-glucosylation of steroid sapogenins, diosgenin, nuatigenin and tigogenin, as well as steroidal alkaloids, solanidine, solasodine and tomatidine [38]. *Dz*S3GT enzyme from *D. zingiberensis* has been characterized as a 3-*O*-sterol glycosyltransferase in the biosynthesis of diosgenin 3-*O*-glucoside [39]. *Tf*3SGT2 of *T. foenum-graecum* (fenugreek) is a steroid-specific 3-*O* glucosyltransferase involved in diosgenin or yamogenin-derived steroidal saponin biosynthesis [40]. Although the previous studies on transcriptomic data combined with the functional genetics approaches such as, *in vitro* biochemical and other complementary assays to date have substantially explained the biosynthetic mechanisms of steroidal natural products in medicinal plants and importantly in *Paris* species, how the UGT encoding genes are regulated on the cellular level remains in the exploratory stage. Specifically, not much is known about their subcellular location and potential relevance to the steroidal saponins biosynthesis and transport mechanism, as well as their *in planta* characterization. *In vivo* transient gene expression using the heterologous hosts as exemplified from *D. zingiberensis* involving *Dz*S3GT for the production of trillin [39], and from other medicinal plants for transferase genes-mediated saponins biosynthesis [41,42], could be employed for *P. polyphylla* to understand the UGT-mediated glycosylation pathway of steroidal saponins synthesis.

Here, we utilized the putative transcriptome data derived from the previously reported enormous genome of *Paris polyphylla* var. *yunnanensis* [43], and identified four candidate steroidal UDP-glycosyltransferases named as UGT91BP2, UGT703R1, UGT703R2 and UGT703R3. Further, through phylogenetic analysis of the UGTs combined with the functional approaches including, gene cloning, *in vitro* biochemical activity assays of the recombinant proteins, subcellular localization and target metabolite detection in *N. benthamiana* transiently, we characterized the four *Pp*UGTs as saponin biosynthetic enzymes. The combined findings of this work advance our understanding of the biosynthesis and regulation of saponins, and provide the basis for the metabolic production of bioactive steroidal glycosides in heterologous microbial and botanical systems.

2. Materials and methods

2.1. Chemicals and source materials

Diosgenin, pennogenin, and trillin standard compounds were purchased from Chengdu Push Biotechnology Co., Ltd., and Shanghai Yuanye Biotechnology Co., Ltd. (China). UDP-glucose was procured from Beijing Solarbio Science and Technology Co., Ltd. (China). Molecular biology kits and other biochemical reagents were obtained from Nanjing Vazyme Biotechnology Co., Ltd., and Yeasen Biotechnology Co., Ltd. (Shanghai, China) unless otherwise specified. The *Paris polyphylla* source plant materials used in this study were collected from the greenhouse field in Yunnan province, China.

2.2. Screening and bioinformatics analysis of PpUGT genes

The candidate UGT transcripts in this study were screened from the putative transcriptome data derived from *Paris polyphylla* var. *yunnanensis* genome sequencing and assembly (data have been deposited into CNGBdb under the project number CNP0001050) as previously constructed [43]. Expasy - Protparam tool (https://web.expasy.org/protparam/) was used to determine the predicted molecular weights, theoretical isoelectric point, instability index, and grand average of hydropathicity of the putative UGT transcripts. Protein transmembrane regions were predicted using TMHMM-2.0 (https://services.healthtech.dtu.dk/services/TMHMM-2.0/) and DeepTMHMM (https://dtu.biolib.com/DeepTMHMM). Signal peptide analyses were performed using SignalP-6.0 (https://services.healthtech.dtu.dk/services/SignalP-6.0/). Web logo of a putative plant secondary product glycosyltransferase (PSPG) signature motif was generated through Multiple Em for Motif Elicitation (MEME suite) (https://meme-suite.org/meme/tools/meme). The names of *Pp*UGT encoding genes were obtained with the help of the UGT nomenclature committee (https://labs.wsu.edu/ugt/submission-of-ugt-sequences-for-naming/). Subcellular localization was predicted using DeepLoc-2.0 (https://services.healthtech.dtu.dk/services/DeepLoc-2.0/), WoLF PSORT (https://wolfpsort.hgc.jp/), and CELLO v.2.5

(http://cello.life.nctu.edu.tw/).

2.3. Multiple sequence alignments and phylogenetic tree analysis

Multiple amino acid sequence alignments of the candidate *Pp*UGTs in this study and the reported homologues obtained from NCBI (http://www.ncbi.nlm.nih.gov/) were performed using ClustalW bundled in MEGA software [44]. Further analysis and shading of the transported alignment file were carried out using GeneDoc software. The phylogenetic tree was constructed using MEGA 11 by neighbor-joining method with all the default parameters. The tree reliability was evaluated using the bootstrap test with 1000 replicates. The phylogenetic tree was modified through iTOL (https://itol.embl.de/). The information of the UGTs used for the phylogenetic analysis is given in Table S2.

2.4. Molecular cloning and heterologous protein expression

To prepare the expression vectors for heterologous protein production, the coding sequences of the candidate *PpUGT* genes extracted from the putative transcriptome were amplified from the mixed cDNA of *P. polyphylla* through RT-PCR with specific primer pairs (Table S3). The PCR was performed using 2× PrimeSTAR Max premix (Takara Bio, China) with the following program: 98 °C for 3 min, 34 cycles of 98 °C for 15 s, 58 °C for 15 s, and 72 °C for 1 min, followed by 72 °C for 10 min. The purified PCR products were then cloned into the linearized pGEX-6P-1 expression vector (with a glutathione *S*-transferase tag) at *Bam*HI and *Eco*RI restriction sites through homologous recombination utilizing ClonExpress® Ultra One Step Cloning Kit (Vazyme). The *E. coli* DH5α expressed clones (confirmed by PCR and DNA sequencing) carrying the recombinant pGEX-*PpUGT* plasmids were then transformed into *E. coli* BL21(DE3) for heterologous protein production. Bacterial cell cultures of single clones were inoculated in LB liquid medium containing ampicillin (50 mg/L) and grown at 37 °C and 150 rpm until OD at 600 nm (OD_{600}) reached 0.6∼0.8. To induce the recombinant protein expression, isopropyl-*β*-D-thiogalactoside (IPTG) was added to a final concentration of 0.2 mM and grown continually for 18–24 h at 16 °C. The cells were harvested (centrifugation at 5000 *g*, 10 min, 4 °C) and resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 400 mM NaCl). The cells were disrupted through sonication (Ultrasonic Homogenizer JY92IIDN, Scientz) in an ice-chilled environment, and cell debris was removed by centrifugation (at 12000 *g* for 1 h at 4 °C). Glutathione resins (GST-Sefinose™ Resin 4FF) (Sangon Biotech) were added to the bacterial supernatant containing the target proteins for purification. After incubation for 1 h, the mixture was transferred into a disposable column and washed with 20 times column volumes of lysis buffer. The target proteins were eluted with glutathione-reduced (GSH) prepared in lysis buffer according to the manufacturer's manual. The size and purity of the *Pp*UGT enzymes were confirmed by SDS-PAGE and visualized using Coomassie Brilliant Blue staining. The purified proteins were treated using HRV 3C protease (Beyotime) to remove the GST tag according to the manufacturer's protocol. The purified proteins were stored at −80 °C till use.

2.5. In vitro enzymatic assays

For the recombinant *Pp*UGTs with and without GST tag, the enzymatic assays *in vitro* were performed in a reaction mixture containing 0.5 mM acceptor substrates (diosgenin and pennogenin), 1 mM UDP-glucose, 10 mM $MgCl_2$, and 100 μg purified recombinant UGT in 100 μL of 50 mM Tris-HCl buffer (pH 7.5). The control reactions under the same conditions were conducted by omitting the UGTs. The reactions were incubated at 37 °C for 12 h and stopped by adding 200 μL of ice-cold methanol. The reaction mixtures were centrifuged (14000 *g*, 10 min, 4 °C) and clear supernatants were collected for product analysis on an Ultimate3000-Bruker impact HD liquid chromatography time-of-flight mass spectrometry (LC-Q-TOF-MS).

For time-course analysis of the *Pp*UGTs, the enzymatic reactions with diosgenin and pennogenin were incubated at varying time points, including 1, 2, 4, 8, and 12 h under the same experimental conditions as above. The conversion rates (in μM/min) were calculated from LC-MS peak areas of the glucosylated products and substrates.

2.6. Transient gene expression in N. benthamiana and metabolite analysis

For *in planta* metabolite analysis, the candidate *PpUGTs* were amplified with the specific primers (Table S3), and introduced individually at *Pac*I-*Not*I cloning sites into *35S*-pJL-TRBO expression vector by homologous recombination. The error-free recombinant plasmids were individually transformed into *Agrobacterium tumefaciens* strain GV3101, and positive selected colonies were cultured in YEP liquid medium supplemented with 50 μg/L kanamycin and 25 μg/L rifampicin for overnight growth at 28 °C and 220 rpm, until the OD_{600} reached ∼1.0. Then, the bacterial cultures were centrifuged (5000 *g*, 5 min) and the collected cells were resuspended in a buffer containing 10 mM MES (pH 5.6), 10 mM $MgCl_2$, 100 μM acetosyringone, and the OD_{600} was adjusted to 0.5∼0.6. After 3 h of incubation at room temperature, the cell suspension culture of each gene was infiltrated into the abaxial part of the leaves of about 4–6 weeks old *N. benthamiana* plants using a needleless syringe, until the infiltrate solution covered most of the entire leaf. The *35S*-pJL-TRBO empty vector was used as a control. The infiltrated *N. benthamiana* plants were grown in a growth chamber and maintained at 25 °C with a 16 h light/8 h dark exposure period. After three days post-infiltration, the leaves transiently expressing four *PpUGTs* individually were fed-infiltrated with 1 mM diosgenin substrate. Upon 24 h of cultivation in a growth chamber, the leaves were harvested, snap-frozen in liquid nitrogen and the frozen leaf tissues were ground using Bioprep-24 Homogenizer (Allsheng) for 1 min at 30 Hz. For extraction of metabolites, 1 mL of pure isopropanol was added to the pulverized samples and the slurry was then homogenized by Bioprep-24 Homogenizer (1 min at 30 Hz). The sample mixtures were incubated at 60 °C for 30 min. Following centrifugation at 14,000 *g* for 10 min, clear supernatants were collected in glass vials for metabolite analysis.

2.7. Analytical methods

The *in vitro* enzymatic reaction mixtures of the *Pp*UGTs were analyzed for product determination on liquid chromatography time-of-flight mass spectrometry (LC-Q-TOF-MS). Liquid chromatography-mass spectrometry (LC-MS) was equipped with a YMC-Triart C18 column (250 mm × 4.6 mm, 5 μm), and the column temperature was set to 30 °C. Formic acid (0.1%) aqueous solution (solvent A) and formic acid (0.1%) in acetonitrile (solvent B) were used as mobile phases. The flow rate was 1 mL/min with the gradient elution process as follow: 60% solvent B, 0–1 min; 60–100% solvent B, 1–21 min; 100 % solvent B, 21–26 min; 60% solvent B, 26–32 min. The injection volume was 20 μL. The electrospray ionization (ESI) source was selected in positive ion mode with a mass range of *m/z* 100–1000. The Bruker Compass DataAnalysis version 4.2 software was used for data analysis.

For metabolite analysis in *N. benthamiana*, the leaf sample extracts were analyzed by ultra high-performance liquid chromatography system (SCIEX, ExionLC, UHPLC) coupled with a triple quadrupole mass spectrometer (SCIEX Triple Quad 5500+ QTrap Ready). Chromatographic separation was performed on a YMC-Triart C18 column (250 mm × 4.6 mm, 5 μm) at 40 °C with 0.1% formic acid aqueous solution (solvent A) and 0.1% formic acid in acetonitrile (solvent B) as mobile phases. The flow rate was 1 mL/min with the gradient elution process as follows: 60% solvent B, 0–1 min; 60–100% solvent B, 1–21 min; 100 % solvent B, 21–26 min; 60% solvent B, 26–32 min. The injection volume was 10 μL. The mass spectrometer was operated in positive electrospray ionization (ESI) using the multiple reaction (MRM) mode with a dwell time of 100 msec per transition. The source/gas-dependent parameters were

optimized and set as follows: curtain gas 30 psi; collision gas 9 psi; spray voltage 5500 V; temperature 650 °C; ion source gas 1 (GS1) 60 psi; ion source gas 2 (GS2) 60 psi. The mass transitions and compound-dependent parameters were optimized using the authentic standards and listed in Table S4. Analyst 1.7.1 software (SCIEX) was used for data analysis.

2.8. Determination of subcellular localization

For subcellular localization analysis, the coding region of each *Pp*UGT was amplified with the specific primers (Table S5), and cloned at the *Kpn*I site into pSuper1300-*35S*-GFP *C*-terminus expression vector by homologous recombination. Correct recombinant plasmids harboring the target genes were individually transformed into *A. tumefaciens* strain EHA105. Positive colonies after collection were cultured in LB liquid medium containing 50 μg/L kanamycin and 25 μg/L rifampicin and grown at 28 °C, 220 rpm overnight until the OD_{600} reached ~1.0. The bacterial cultures were then centrifuged at 5000 *g* for 5 min and the collected cells were resuspended in the buffer containing 10 mM MES (pH 5.6), 10 mM $MgCl_2$, 100 μM acetosyringone to the adjusted OD_{600} of 0.5~0.6. After incubation for 3 h at room temperature, the cell suspension culture corresponding to each gene was infiltrated into the abaxial part of the leaves of young (about 4–6 weeks old) *N. benthamiana* plants, using a needleless syringe. The pSuper1300-*35S*-GFP empty vector was used as a control. The infiltrated plants were kept in a growth chamber and maintained at 25 °C for a 16 h light/8 h dark exposure period. The leaf samples of each construct were collected about 3 days after infiltration and leaf discs were stained with DAPI (a nuclear-specific fluorescence dye) for subcellular location analysis. The GFP and DAPI fluorescence signals were separately visualized under a laser scanning confocal microscope (Zeiss LSM900, Germany).

2.9. Quantitative RT-PCR (qRT-PCR) and relative gene expression analysis

Total RNA was extracted from leaf, stem, and flower tissues of *P. polyphylla* using TaKaRa MiniBEST Plant RNA Extraction Kit (Takara, China), and the first-strand cDNA was synthesized through PrimeScript™ II 1st Strand cDNA Synthesis Kit (Takara, China). For the *PpUGTs* expression analysis, qRT-PCR was performed on a qTOWER^3 G Real-Time PCR Thermal Cycler (Analytik Jena AG, Germany) using the PerfectStart® Green qPCR SuperMix (TransGen Biotech, China) according to the product manual. The PCR program was as follow: 94 °C for 30 s, 40 cycles of 94 °C for 5 s 60 °C for 15 s and 72 °C for 10 s. The actin gene was used as an internal standard and relative gene expression levels were calculated using the relative quantification method. The qRT-PCR for each gene sample was performed in biological triplicate. The primers used for qRT-PCR are listed in (Table S3).

3. Results

3.1. Bioinformatics analysis and predicted properties of the putative PpUGTs

The transcriptome data derived from *Paris polyphylla* var. *yunnanensis* genome sequence (70.18 GB) were available [43] at the time of initiating this study. From the transcriptome data, we screened four transcripts, namely, transcript59365, transcript54267, transcript60374 and transcript65582 that putatively encode steroidal saponin UGTs. Meanwhile, we removed genes with significantly shorter sequence lengths and in the absence of the necessary plant secondary product glycosyltransferase (PSPG) domain, based on comparison with the available public transcriptome data for *P. polyphylla* using NCBI-BLAST algorithms. The four candidate genes were named UGT91BP2, UGT703R1, UGT703R2 and UGT703R3 by the UGT naming committee. Their GenBank accession numbers are PQ261043, PQ261044, PQ261045 and PQ261046, respectively.

UGT91BP2 and UGT703R1–3 have open reading frames of 1518, 1398, 1437 and 1371 nucleotide base pairs encoding the corresponding proteins of 505, 465, 478 and 456 amino acid residues, respectively. The predicted molecular weights of these four UGTs are 57.0, 51.7, 52.1 and 49.8 kDa, with theoretical isoelectric point (pI) values of 5.87, 5.84, 5.05 and 5.18, respectively. Their instability index and grand average of hydropathicity scores are also given (Table S1). Predicted transmembrane region and signal peptide analysis revealed that the four *PpUGT* genes encode neither transmembrane protein topologies nor signal peptides (Fig. S1a–l).

3.2. Phylogenetic analysis

Previous phylogenetic studies have grouped the sterol UGTs into different subfamilies based on their biochemical functions. For example, a few members of the GT genes that glycosylate steroidal sapogenins have been placed into UGT73, UGT80 and UGT91 subfamilies [33,34]. To investigate the four *Pp*UGTs belong to which clade, we constructed a neighbor-joining phylogenetic tree using the protein sequences and other characterized glycosyltransferases from the UGT73, 80 and 91 subfamilies (Fig. 1a). It was observed that UGT703R1–3 form a phylogenetic relationship with the members of the UGT73 subfamily, some of which have been reported as sterol 3-*O*-glucosyltransferases, as exemplified by *Pp*UGT73CR1 and *Pp*UGT6 form *P. polyphylla* [30,45], *Sa*GT4A from *Solanum aculeatissimum* [38], and UGT73C5 from *Arabidopsis thaliana* [36], including also terpenoid sapogenin glycosyltransferases (*Bv*UGT73C10–13) from *Barbarea vulgaris* [46]. Whereas, UGT91BP2 was predicted to cluster with the sugar-sugar glucosyltransferases belonging to the UGT91 subfamily. A distinct clade is formed by the UGT80 subfamily genes from those related to the UGT73 and 91 groups. Three UGTs (UGT703R1–3) share low sequence identities on protein level with the UGTs of 73 family such as *Pp*UGT73CR1 (39.0–40.26%), *Pp*UGT6 (38.26–39.19%), and *Sa*GT4A (39.44–41.05%) (Fig. S2). Similarly, UGT91BP2 also shows low amino acid sequence identities (41.98–45.79%) to the members of UGT91 subfamily including three functional steroid glucoside 6′-*O*-glucosyltransferases, UGT91AH 1–3, reported from *P. polyphylla* [34] (Fig. S3). Protein sequence alignments depicted that the four *Pp*UGTs contain a putative PSPG motif (a 44 amino acid signature domain) at their *C*-termini (Fig. 1b) [47]. The terminal residue in the PSPG motif is unanimously glutamine (Q44), which has been found to determine the glucosyl donor specificity [48]; thus, suggesting that the *Pp*UGTs may recruit UDP-glucose as a donor substrate.

3.3. Cloning and functional characterization of the PpUGT candidates

To characterize the biochemical functions of the four *Pp*UGT candidates, the open reading frames of UGT91BP2 and UGT703R1–3 were amplified from the *P. polyphylla* cDNA library (Fig. S4). Each target gene was cloned into the expression vector pGEX-6P-1 with an *N*-terminal glutathione *S* transferase (GST) affinity tag. All of the four recombinant UGTs were successfully expressed in *E. coil* BL21(DE3) and purified, and thereafter confirmed by SDS-PAGE analysis (Fig. S5a–c) before subjecting to enzymatic activity assays *in vitro*. UDP-glucose was used as a potential donor, whereas, diosgenin and pennogenin were used as acceptor substrates for the enzymatic assays of the recombinant *Pp*UGTs.

As a result, all four *Pp*UGTs were able to catalyze the glucosylation of diosgenin to form diosgenin 3-*O*-glucoside, also known as trillin (Fig. 2a–c) to different extent. Specifically, UGT91BP2 and UGT703R1–3 each in a reaction mixed with UDP-glucose and diosgenin generated a product with the retention time of 11.7 min and maximum UV-absorbance at 227 nm as that of trillin authentic standard (Fig. 2b, Fig. S6a and b, Fig. S7). The high-resolution mass spectra (HRMS) of the product of each *in vitro* reaction showed the *m/z* values of $[M+H]^+$ and $[M+Na]^+$ that were consistent with the calculated and observed

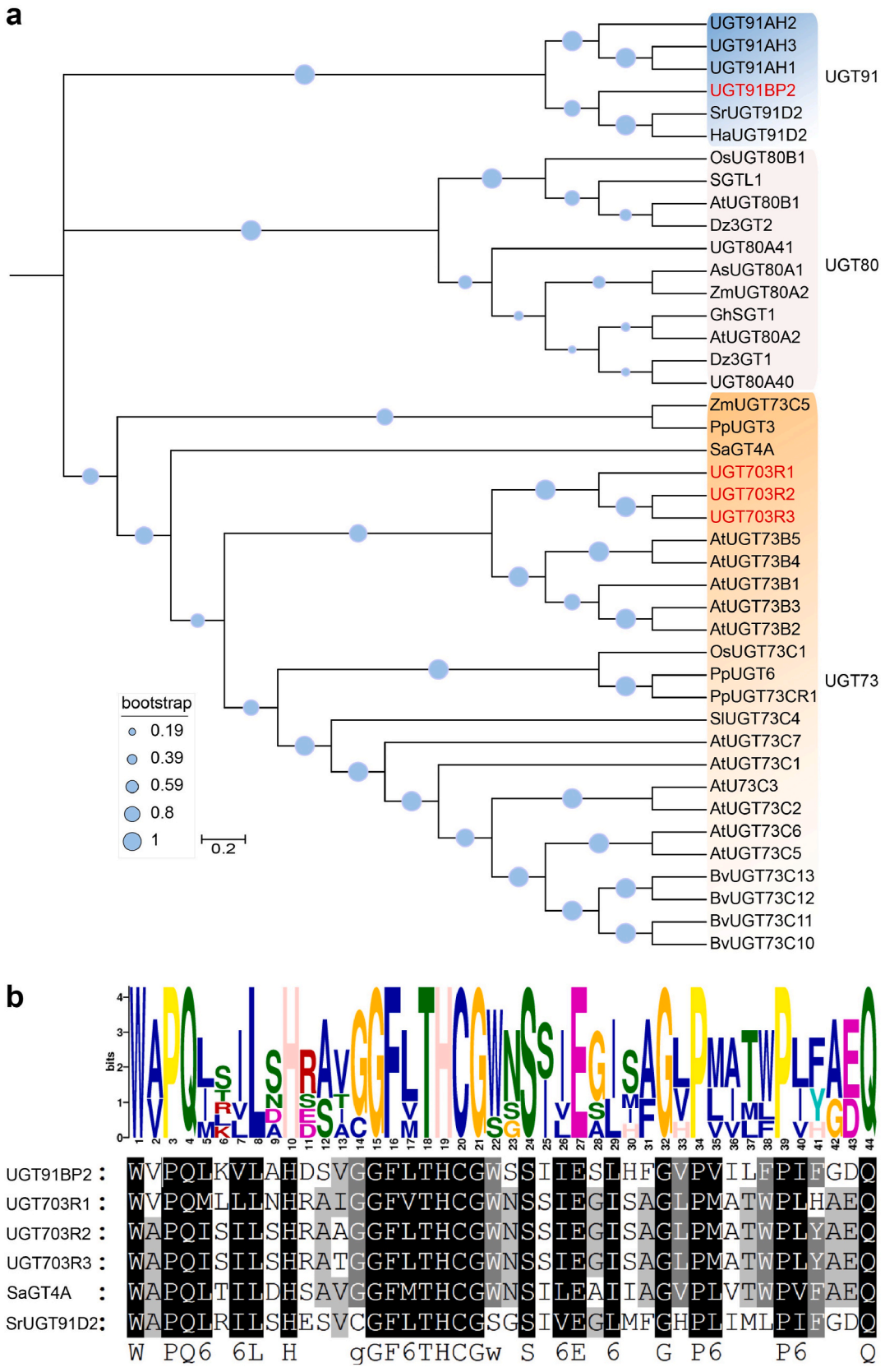
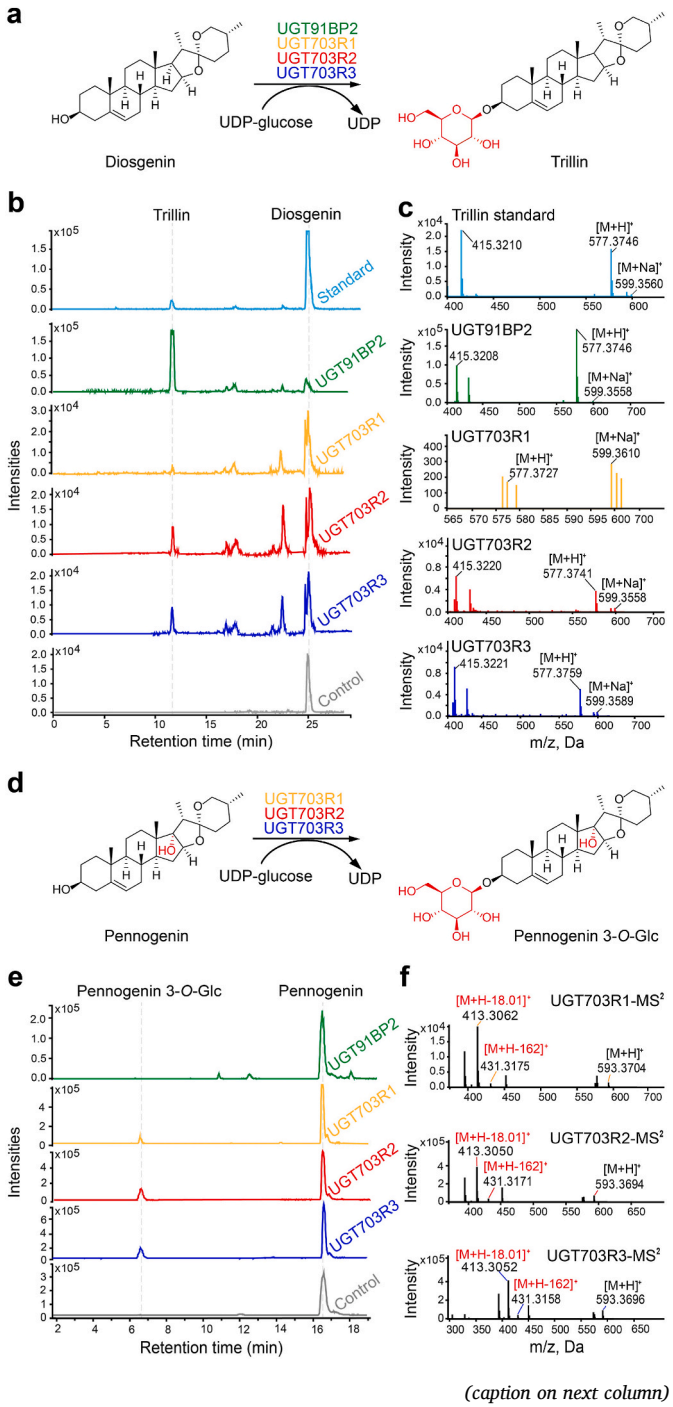


Fig. 1. Phylogenetic analysis and protein sequence alignment of *Paris polyphylla* UGTs. (**a**) Phylogenetic tree of the candidate UGTs from *P. polyphylla* and a selected number of plant UGTs. The UGT genes in this study are marked with red color. The tree is generated in MEGA software using neighbor-joining method with the bootstrap values based on 1000 replicates. The scale bar shows 0.2 amino acid substitutions per site. The UGT73, UGT80 and UGT91 families are represented in blue, green and pink backgrounds, respectively. The accession numbers of the plant UGTs used in the tree are listed in Table S2. (**b**) Web logo representing a putative PSPG conserved motif generated via MEME suite using amino acid sequences of UGT91BP2 and UGT703R1–3 with the *Solanum aculeatissimum* derived *Sa*GT4A (BAD89042.1) and *Stevia rebaudiana* derived *Sr*UGT91D2 (B3VI56).



(caption on next column)

Fig. 2. Reaction schemes and LC-MS analysis of *Pp*UGTs reactions with diosgenin and pennogenin. (**a**) Biosynthesis of trillin mediated by UGT91BP2 and UGT703R1–3 through transferring glucose from UDP-glucose to the C-3 OH position of diosgenin. (**b**) Extracted ion chromatograms (EICs) of the *in vitro* enzyme activity assays of recombinant UGT91BP2 and UGT703R1–3 with UDP-glucose and diosgenin showing the formation of trillin, as compared to the authentic trillin standard. Control is the empty expression vector. (**c**) MS spectra of the enzymatic reaction products in hydrogen and sodium ion adducts compared to the fragmentation pattern of the trillin standard. The product molecular ions $[M + H]^+$ and $[M + Na]^+$ and the feature fragment ion $[M + H - Glc]^+$ with *m/z* 415.32 are marked. Of note, additional peaks observed in the diosgenin EIC (*m/z* 415.32) for UGT703R1–3 might correspond to substrate isomeric impurities, in-source adducts, or minor non-enzymatic by-products; these peaks, which arose from the extracted ion flow of the substrate *m/z*, did not co-elute with the product, trillin (Fig. S6a and b), and therefore did not affect the interpretation of the product formation. (**d**) Biosynthesis of pennogenin 3-*O*-glucoside mediated by the UGT703R1–3 through transferring glucose from UDP-glucose to the C-3 OH position of pennogenin. (**e**) EICs of the *in vitro* enzyme activity assays of recombinant UGT91BP2 and UGT703R1–3 with UDP-glucose and pennogenin showing the formation of pennogenin 3-*O*-glucoside, compared to the control empty expression vector. (**f**) MS/MS spectra ($[M + H - Glc]^+$ and $[M + H - Glc - H_2O]^+$) of the enzymatic reaction products.

molecular weight of trillin standard (Fig. 2c). Comparatively, UGT91BP2 and UGT703R1 displayed the highest and lowest activity, respectively, as evident from the LC-MS profiles (Fig. 2b and c).

Next, we assessed the enzyme activities of the four *Pp*UGTs against pennogenin and interestingly found that UGT703R1–3 were also able to glucosylate pennogenin to produce pennogenin 3-*O* glucoside (Fig. 2d–f). LC-MS analysis demonstrated a new product peak at the retention time of 6.7 min and with $[M + H]^+ = 593.3$ compared to the control reaction, when each of the *Pp*UGTs was assayed in a reaction with pennogenin acceptor and UDP-glucose donor (Fig. 2e and Fig. S8). Due to the lack of pennogenin 3-*O*-glucoside authentic standard, we further conducted MS/MS analysis for product identification. The MS/MS spectra displayed a characteristic fragmentation pattern in which $[M+H-162]^+$ and $[M+H-162-18]^+$ ions at *m/z* 431.3 and 413.3 were observed due to loss of a glucose group and of a glucose plus water molecule, respectively (Fig. 2f). Based on these observations, the product synthesized by each of UGT703R1–3 was annotated as pennogenin 3-*O*-glucoside, also termed as floribundasaponin A, the structure of which has been determined previously by NMR spectroscopy [45,49]. UGT91BP2 was unable to accept pennogenin *in vitro*, positing that it is diosgenin-specific.

We also evaluated the diosgenin and pennogenin conversion rates of the *Pp*UGTs by assaying the reactions at 1, 2, 4, 8, and 12 h at 37 °C (Fig. 3a and b). The activity of UGT91BP2 towards diosgenin remained low during the first 2 h before significantly increasing by 4 h (2.11 μM/min) and peaking at 8 h (3.27 μM/min), followed by a slight decline to 3.00 μM/min at 12 h. UGT703R1 exhibited low overall activity with diosgenin, reaching a maximal rate of 0.04 μM/min at 8 h, with the gradual increase reflecting substrate equilibration, while sustained activity over 12 h confirming stability. UGT703R2 displayed an initial decrease in activity with diosgenin from 1 to 2 h, and thereafter, the activity increased to 0.28 μM/min at 8 h, before decreasing to 0.15 μM/min by 12 h UGT703R3 with diosgenin demonstrated enzyme activity that remained steadily high (0.21 μM/min) from the first hour and increased gradually to 0.28 μM/min by 12 h.

UGT703R1 showed the lowest overall activity against pennogenin, but with a steady increase in product formation rates over the entire time course from 0.004 μM/min at 1 h to 0.02 μM/min at 12 h. UGT703R2 exhibited an increase in activity, with pennogenin conversion rates rising from 0.02 μM/min at 1 h to 0.17 μM/min at 12 h. UGT703R3 gave a robust and consistent increase in activity, with the product formation rates starting at 0.030 μM/min in the first hour and rising to 0.21 μM/min at 12 h. These findings highlight the differences in the enzymatic performance and stability of the *Pp*UGTs for the product

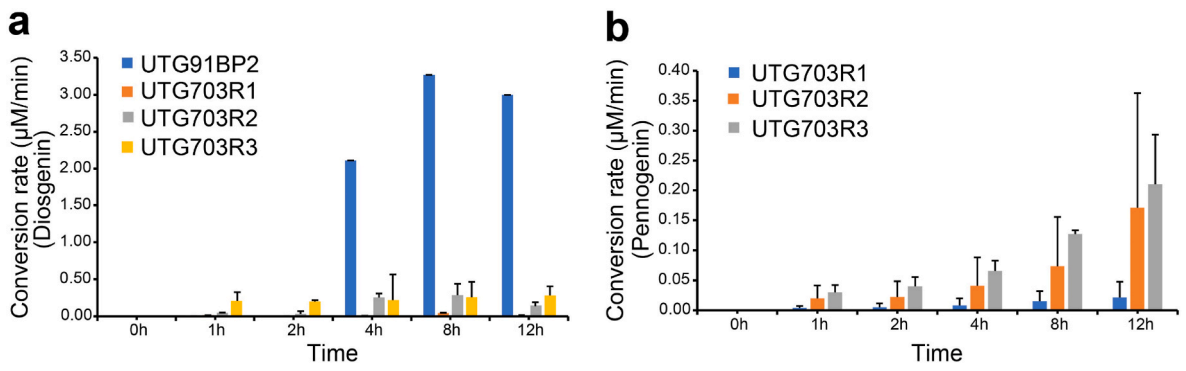


Fig. 3. Time-course activity analysis of *Pp*UGTs. Conversion rates of the *in vitro* glucosylation assays of UGT91BP2 and UGT703R1–3 with diosgenin (**a**) and of UGT703R1–3 with pennogenin (**b**) for 1, 2, 4, 8, and 12 h. UDP-glucose was used as a donor substrate. The data are recorded as mean ± SD, n = 2.

formation when utilizing diosgenin and pennogenin in response to reaction time.

3.4. Characterization of PpUGT1–4 in N. benthamiana

Having known the *in vitro* enzymatic activities of UGT91BP2 and UGT703R1–3, we next investigated the catalytic role of these four UGTs *in vivo*. We cloned and conducted Agrobacterium (*A. tumefaciens*)-mediated transient expression assays by individually transfecting each gene in *N. benthamiana* leaves and subsequently supplementing with diosgenin substrate. Next, we performed the ultra high-performance liquid chromatography-mass spectrometry (UHPLC-MS) based targeted metabolite analysis. As expected, the accumulation of trillin was observed (parental ion Q1: 577, daughter ion Q3: 253, Table S4) in the leaf extract samples transiently expressing each of the four UGT genes compared to authentic trillin standard and the empty vector control (Fig. 4a). These results demonstrate that the *Pp*UGTs can participate in the heterologous production of trillin in tobacco, consolidating their *in vitro* enzymatic activities.

3.5. Subcellular localization analysis

In silico analyses using different web tools were carried out to predict the subcellular localization of UGT91BP2 and UGT703R1–3 (Table S5). According to the computational results, the localization signals of the four UGTs were detected in different subcellular compartments such as cytoplasm, nucleus, chloroplast, and other organelles. To examine these predictions, we performed subcellular localization experiments in tobacco. We constructed pSuper1300*35S*-UGT91BP2-GFP, pSuper1300*35S*-UGT703R1-GFP, pSuper1300*35S*-UGT703R2-GFP, and pSuper1300*35S*-UGT703R3-GFP vectors for expression of *C*-terminal GFP fusion proteins. Next, we transiently agroinfiltrated each construct individually into *N. benthamiana* leaves and the GFP fluorescence signals were visualized using laser scanning confocal microscope. Interestingly, the fluorescence signals of UGT91BP2-GFP, UGT703R1-GFP, UGT703R2-GFP, and UGT703R3-GFP were observed in both cytoplasm and nucleus (Fig. 4b). Moreover, when we incubated the infiltrated *N. benthamiana* leaf cells with DAPI, the four *Pp*UGT-GFP fluorescent signals were found to colocalize with the DAPI signals in the nuclei (Fig. 4b). These findings indicated that the proteins encoded by these UGT genes were expressed in both cytoplasm and nucleus (localization in the nucleus is likely due to protein passive diffusion through the nuclear pores [50]). As reported, glycosylation is a promising way to improve water solubility, stability and bioavailability [51,52], supporting the notion of the functional roles of these biosynthetic UGTs in steroidal saponin metabolisms.

3.6. PpUGT gene expression analysis

The tissue-specific expression levels of four *PpUGTs* were investigated in leaf, stem, and flower using qRT-PCR. Their transcript levels depicted differential expression patterns in these tissues. *UGT91BP2* was expressed more in leaf than in flower, and its expression was extremely low in stem as compared to in leaf and flower (Fig. 5a). *UGT703R1* expression was high in flower compared to in leaf, while rarely detectable in stem (Fig. 5b). Relatively high transcript levels were observed for *UGT703R2* and *UGT703R3* in leaf than in flower, whereas, expression levels were lower in stem than in leaf and flower (Fig. 5c and d). Differences in the tissue-specific expression of numerous identified genes corroborate the presence of tissue-specific pathways [53], indicating that differential tissue-specific expression of *PpUGTs* may modulate the tissue-specific relative metabolite contents. In the previous study, a significant correlation of the expression pattern of a few candidate UGTs with the total polyphyllin production in different tissues of *P. polyphylla* has been established, and the authors further observed that diosgenin and pennogenin substantially accumulated in leaf and flower at specific growth phases [43]. In our findings, *UGT91BP2* and *UGT703R1–3* showed expression mostly in leaf and flower tissues, and notably, these *Pp*UGTs catalyzed the glucosylation of diosgenin and pennogenin, strongly suggesting that these UGTs are likely involved in the diosgenin and pennogenin saponins biosynthesis in their original host plant.

4. Discussion

Medicinal plants produce a variety of specialized metabolites as an important source of nutrients, energy, and pharmaceutical agents [54,55]. For thousands of years, over 10,000 medicinal plants have been used in Traditional Chinese Medicine (TCM) due to their enrichment in specialized natural products [49]. For instance, *Dioscorea zingiberensis*, an important medicinal plant in TCM with the abundance of 2–16% steroidal sapogenin (diosgenin) content in the rhizomes [56]. *Paris* plants are another valuable medicinal herb rich in steroidal saponins, the major active ingredients of a well-known traditional Chinese pharmacopeia, Rhizoma Paridis [34]. Therefore, elucidation of the molecular mechanisms behind the biosynthesis and diversity of plant specialized metabolites with medicinal importance not only improves our knowledge of these metabolites, but also provides applicable targets for metabolic engineering [57]. The majority of metabolic diversity arises due to the tailoring modifications of the basic scaffold structures [58]. Glycosylation mediated by UGTs is an important step in decorating steroidal sapogenins, such as diosgenin and pennogenin, which provides a foundation for the polyphyllins (diosgenin- and pennogenin-derived saponins) biosynthesis [34].

Herein, by mining the transcriptome of *P. polyphylla* var. *yunnanensis* with an enormous genome (70.18 Gb) and phylogenetic analysis in combination with functional approaches, we characterized four UGT

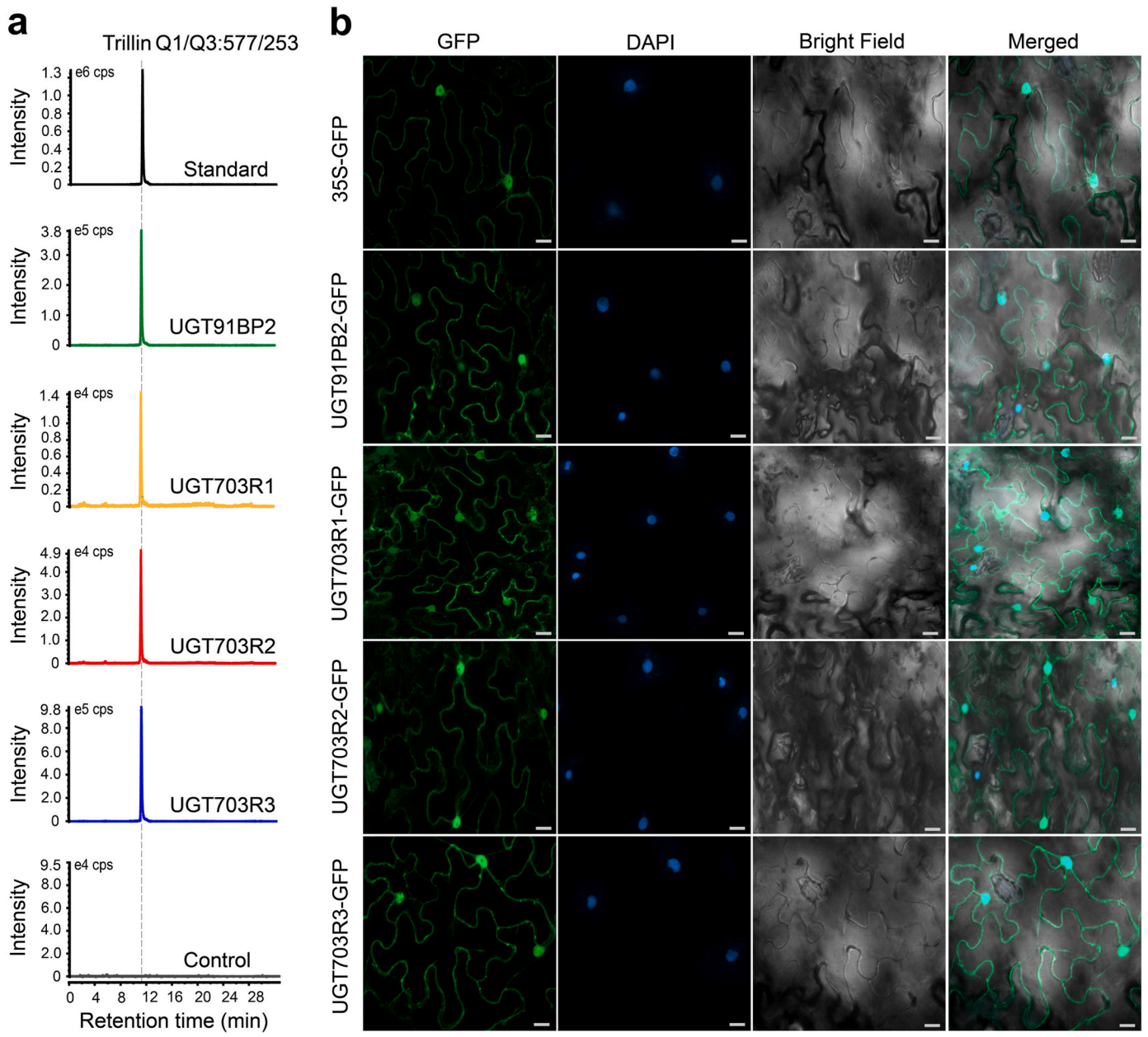


Fig. 4. Transient expression and characterization of *Pp*UGTs in *Nicotiana benthamiana*. (**a**) UHPLC-MS traces of the trillin formation in *N. benthamiana* leaf extracts transiently expressing UGT91BP2 and UGT703R1–3 with the infiltration of diosgenin substrate, compared to trillin authentic standard and empty vector control. Q1: parent ion; Q3: daughter ion. (**b**) Confocal microscopic images of the localization of UGT91BP2-GFP, UGT703R1-GFP, UGT703R2-GFP and UGT703R3-GFP in *N. benthamiana* leaves. DAPI (a nuclear-specific fluorescence dye) acts as a nucleus indicator. Scale bars, 20 μm. The experiments were repeated two times with similar results.

genes involved in the polyphyllin biosynthetic pathway. Four recombinant enzymes, UGT91BP2 and UGT703R1–3 acted as 3-*O*-glucosyltransferases that convert diosgenin to trillin in the *in vitro* enzyme assays (Fig. 2a–c), while among these four UGTs, UGT703R1–3 catalyzed the glucosylation of pennogenin to produce pennogenin 3-*O*-glucoside (floribundasaponin A) *in vitro* (Fig. 2d–f), demonstrating the potential of a heterologous prokaryotic (*e.g.*, *E. coli)* system for the production of trillin and pennogenin 3-*O*-glucoside originally from a plant source. The phylogenetic tree places these functional *Pp*UGTs into different subfamilies including UGT73 and UGT91, distinct from the UGT80 subfamily (Fig. 1a). The UGT73 subfamily presents the largest clade comprising the UGTs mostly from *Arabidopsis thaliana*, which has been reported to contain a large number of members with 13 in this subgroup [59]. UGT703R1–3 which cluster closely (UGT703R1 shares 75% and 76% identities with UGT703R2 and UGT703R3, respectively; UGT703R2 and UGT703R3 share >90% identity) in the phylogenetic tree, fall into the UGT73 subgroup, a few members of which function as sterol glycosyltransferases [30,45,60]; with diverse catalytic ability to recognize more acceptor structures. *Sa*GT4A glycosyltransferase in *Solanum aculeatissimum* shows C-3 glycosylation activity towards diosgenin, nuatigenin and tigogenin, and other steroidal alkaloids [38]. We have functionally characterized that UGT703R1–3 can convert diosgenin and pennogenin into their corresponding 3-*O*-glucosides in *P. polyphylla*, validating the phylogenetic argument. UGT91BP2 clusters phylogenetically with the UGT91 subfamily (Fig. 1a), the biochemical function of which has remained largely unexplored, with only a few identified as rhamnosyltransferases and 2′-*O*-glucosyltransferases [34, 61,62]. Recently, three functional UGT91AH1–3 from *P. polyphylla* have

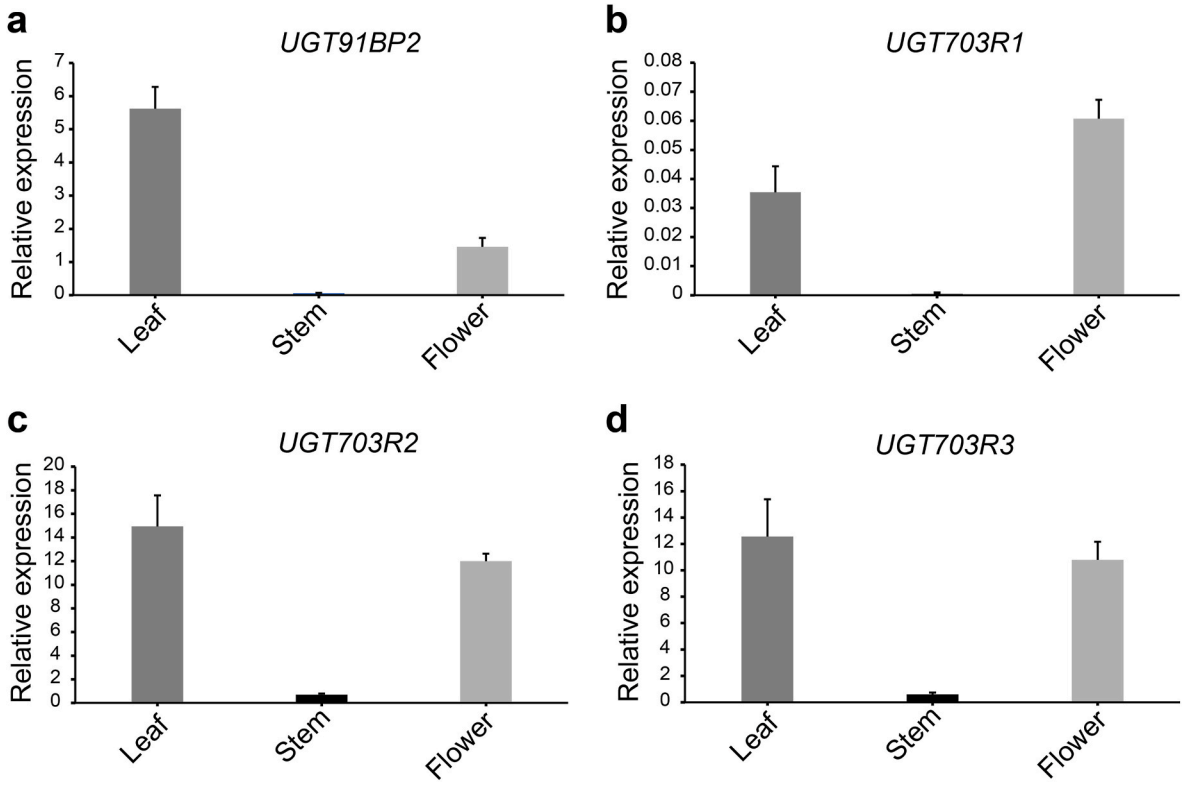


Fig. 5. qRT-PCR analysis of *P. polyphylla* UGTs expression. Bar charts of the relative expression levels for *UGT91BP2* (**a**), *UGT703R1* (**b**), *UGT703R2* (**c**) and UGT703R3 (**d**) genes in leaf, stem and flower tissues. Actin gene was used as an internal standard for data normalization. The data are presented as mean ± SE, n = 3.

been identified from the UGT91 subfamily as sugar-sugar glycosyltransferases (*i.e.*, steroid glucoside 6′-*O*-glucosyltransferases) catalyzing further 6′-*O*-glucosylation of *Paris* saponins (polyphyllins V and VI) to produce the respective triglycosides, which besides showed not to achieve activities on steroid aglycones, diosgenin and pennogenin, distinct from the sugar-sterol glycosyltransferases [34]. Our results revealed that UGT91BP2 can only glucosylate diosgenin but not pennogenin, indicating that the UGT91 subfamily can also feature a diosgenin-specific steroidal monoglucoside UGT in this phylogenetic clade. Further, UGT91BP2 and UGT703R1–3 demonstrate variable temporal activity profiles for diosgenin and pennogenin consumptions (Fig. 3a and b), highlighting the enzyme- and substrate-specific nature of their performance, which could be used for guiding the selection of these UGTs in steroidal saponin glucoside formation.

Published data demonstrated that *P. polyphylla* and *Trigonella foenum-graecum* independently employ pairs of cytochromes P450 enzymes, the heterologous expression of which in *N. benthamiana* catalyzed the cholesterol-derived biosynthesis of diosgenin [32]. Consistently, the discovery of cholesterol biosynthetic metabolic network in *P. polyphylla* for efficient synthesis of diosgenin in *N. benthamiana* has been deciphered [63], opening a route for engineering the biosynthesis of diosgenin, with further expanding the access to a diverse array of diosgenin-derived analogs through the characterization of downstream steroid-modifying enzymes in a heterologous plant synthetic chassis. With their *in vitro* enzyme activities in hand, we have demonstrated that UGT91BP2 and UGT703R1–3 transiently in *N. benthamiana* functioned as the biosynthetic enzymes for the glucosylation of diosgenin to produce trillin (Fig. 4a), indicating that these *Pp*UGTs contribute to the formation of steroidal saponins *in planta*. While the context of their *in vitro* biochemical activities for pennogenin points towards their speculated *in planta* function that these UGTs could also be good candidates to generate pennogenin saponins in tobacco, which needs to be experimentally verified in the future. The arrangement of a series of biosynthetic enzymes systematically in heterologous hosts can efficiently direct the native pathway metabolite flux or externally provide precursors to the desired product pathways [64]. Using this strategy, optimized metabolic pathway engineering involving these well-characterized UGTs for the biosynthesis of steroidal glycoproducts in tobacco could be achieved.

Identification of the subcellular location of biosynthetic enzymes importantly contributes to the optimization of plant synthetic biology and metabolic engineering approaches [65]. The cytosolic MVA and plastidial MEP pathways involving the tailoring intermediate steps have been identified as the early biosynthetic routes to contribute to the formation of the downstream plant steroidal saponins [29]. By conducting transient expression experiments involving *N. benthamiana* leaves, we verified the localization of the GFP-fused UGT91BP2 and UGT703R1–3 proteins in both cytoplasm and nucleus (Fig. 4b), highlighting the possible functions of these biosynthetic UGTs in steroidal saponin metabolisms. *D. zingiberensis Dz*S3GT, a diosgenin 3-*O* glucose transferase has been observed to present mainly in the cytoplasm of the rice protoplast [39]. *Quillajia saponaria* saponins (QS-21) glycosylation catalyzed by the two QS GTs belonging to cellulose synthase-like family enzymes (CSLMs) takes place in endoplasmic reticulum and cytoplasm upon migration, whereas UGT73CU3 and UGT73CY3 are cytosolic enzymes [66]. The nucleus-localized destination of the UGTs in our study may likely arise from the protein diffusion into the nucleus as documented for a cytosolic GH1 transglycosidase, SoGH1, a *Saponaria officinalis* saponin sugar transferase [50]. Some plant secondary compounds and steroidal triterpenoid saikosaponins are often stored in vacuoles to avoid self-poisoning. Given the speculation that modification (glycosylation/acylation) of the substrates occurs in the cytoplasm, which renders their easier transport into vacuoles for the sake of plant protection from self-toxicity caused by the excessive accumulation levels of secondary metabolites [67,68]. Therefore, UGT91BP2 and UGT703R1–3 may play an important role in improving the solubility and maintaining the stabilization of steroidal compounds metabolism in *P. polyphylla*, as common modifications of plant secondary metabolites including

glycosylation often engendering changes in their physiochemical properties [51,52].

5. Conclusion

In conclusion, we characterized four UDP-glycosyltransferases (UGT91BP2 and UGT703R1–3) from *P. polyphylla* for steroidal saponin biosynthetic functionalization, through phylogenetic analysis of the UGTs integrated with the *in vitro* biochemical activity assays and *in vivo* expression analyses. Expression of these *Pp*UGTs resulted in the production of steroidal saponins in *N. benthamiana*, providing new insight into the *in planta* characterization, with the need for efforts however, to optimize the heterologous biosynthesis for improved production of steroidal glycoproducts in tobacco. Further, the subcellular localization of the *Pp*UGTs may guide in the determination of the glycosylation site and transportation route of steroidal aglycones on cellular levels. This work expands our knowledge of UGT-catalyzed glucosylation of diosgenin and pennogenin sapogenins of the polyphyllin biosynthetic pathway and provides a novel valuable reference for biomanufacturing bioactive steroidal glycoconjugates in heterologous chassis. In addition, harnessing biosynthetic platforms involving promiscuous UGTs for producing steroidal natural compounds described here will allow for the structure-function relationship to be established in the future, and will thus favor the rational design and production of potent pharmacological *Paris* saponins.

CRediT authorship contribution statement

Hizar Subthain: Writing – original draft, Investigation, Formal analysis, Data curation, Conceptualization. **Fei Guo:** Investigation, Formal analysis. **Chengjie Zhang:** Investigation, Formal analysis. **Peng Fang:** Investigation, Formal analysis. **Lingli Zhang:** Formal analysis, Project administration. **Qixuan Su:** Investigation, Data curation. **Dandan Tang:** Investigation, Funding acquisition. **Luping Chi:** Investigation, Funding acquisition. **Changning Liu:** Writing – review & editing, Validation, Supervision, Conceptualization. **Vlada B. Urlacher:** Writing – review & editing, Validation. **Jing Li:** Writing – review & editing, Validation, Supervision, Conceptualization. **Lei Du:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis, Conceptualization. **Shengying Li:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

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.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2020YFA0907900), the National Natural Science Foundation of China (32370032, 32170088, 32300023, 32200022), and the Qingdao Postdoctoral Applied Research Project. The authors thank Prof. Yongfeng Guo from the Tobacco Research Institute of Chinese Academy of Agricultural Sciences for providing us with the *35S*-pJL-TRBO vector. We also thank Jing Zhu, Jingyao Qu, Zhifeng Li and Guannan Lin from the Core Facilities of Life and Environmental Sciences, State Key Laboratory of Microbial Technology of Shandong University for assistance in LC-MS/MS and UHPLC-MS analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2026.04.002.

References

[1] Osbourn A, Goss RJM, Field RA. The saponins – polar isoprenoids with important and diverse biological activities. Nat Prod Rep 2011;28:1261–8. https://doi.org/10.1039/C1NP00015B.
[2] Jiang QW, Chen MW, Cheng KJ, Yu PZ, Wei X, Shi Z. Therapeutic potential of steroidal alkaloids in cancer and other diseases. Med Res Rev 2015;36:119–43. https://doi.org/10.1002/med.21346.
[3] Pérez Quiñones J, Szopko R, Schmidt C, Peniche Covas C. Novel drug delivery systems: Chitosan conjugates covalently attached to steroids with potential anticancer and agrochemical activity. Carbohydr Polym 2011;84:858–64. https://doi.org/10.1016/j.carbpol.2010.12.007.
[4] Rizwan K, Zubair M, Rasool N, Riaz M, Zia-Ul-Haq M, de Feo V. Phytochemical and biological studies of *Agave attenuata*. Int J Mol Sci 2012;13:6440–51. https://doi.org/10.3390/ijms13056440.
[5] Szakiel A, Pączkowski C, Henry M. Influence of environmental biotic factors on the content of saponins in plants. Phytochem Rev 2010;10:493–502. https://doi.org/10.1007/s11101-010-9164-2.
[6] Upadhyay S, Jeena GS, Shikha Shukla RK. Recent advances in steroidal saponins biosynthesis and in vitro production. Planta 2018;248:519–44. https://doi.org/10.1007/s00425-018-2911-0.
[7] Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A. Triterpene biosynthesis in plants. Annu Rev Plant Biol 2014;65:225–57. https://doi.org/10.1146/annurev-arplant-050312-120229.
[8] Kohara A, Nakajima C, Yoshida S, Muranaka T. Characterization and engineering of glycosyltransferases responsible for steroid saponin biosynthesis in Solanaceous plants. Phytochemistry 2007;68:478–86. https://doi.org/10.1016/j.phytochem.2006.11.020.
[9] Bowles D, Isayenkova J, Lim E-K, Poppenberger B. Glycosyltransferases: managers of small molecules. Curr Opin Plant Biol 2005;8:254–63. https://doi.org/10.1016/j.pbi.2005.03.007.
[10] Y Yonekura-Sakakibara K, Hanada K. An evolutionary view of functional diversity in family 1 glycosyltransferases. Plant J 2011;66:182–93. https://doi.org/10.1111/j.1365-313X.2011.04493.x.
[11] Louveau T, Osbourn A. The sweet side of plant-specialized metabolism. Cold Spring Harbor Perspect Biol 2019;11:a034744. https://doi.org/10.1101/cshperspect.a034744.
[12] Louveau T, Orme A, Pfalzgraf H, Stephenson MJ, Melton R, Saalbach G, et al. Analysis of two new arabinosyltransferases belonging to the carbohydrate-active enzyme (CAZY) glycosyl transferase family1 provides insights into disease resistance and sugar donor specificity. Plant Cell 2018;30:3038–57. https://doi.org/10.1105/tpc.18.00641.
[13] Sharma LK, Madina BR, Chaturvedi P, Sangwan RS, Tuli R. Molecular cloning and characterization of one member of 3β-hydroxy sterol glucosyltransferase gene family in *Withania somnifera*. Arch Biochem Biophys 2007;460:48–55. https://doi.org/10.1016/j.abb.2007.01.024.
[14] Reed J, Orme A, El-Demerdash A, Owen C, Martin LBB, Misra RC, et al. Elucidation of the pathway for biosynthesis of saponin adjuvants from the soapbark tree. Science 2023;379:1252–64. https://doi.org/10.1126/science.adf3727.
[15] Chaturvedi P, Misra P, Tuli R. Sterol glycosyltransferases—the enzymes that modify sterols. Appl Biochem Biotechnol 2011;165:47–68. https://doi.org/10.1007/s12010-011-9232-0.
[16] Gleadow RM, Møller BL. Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. Annu Rev Plant Biol 2014;65:155–85. https://doi.org/10.1146/annurev-arplant-050213-040027.
[17] Kalinowska M, Zimowski J, Pączkowski C, Wojciechowski ZA. The formation of sugar chains in triterpenoid saponins and glycoalkaloids. Phytochem Rev 2005;4(2-3):237–57. https://doi.org/10.1007/s11101-005-1422-3.
[18] Kim Y, Thwe A, Li X, Tuan P, Lee S, Lee J, et al. Accumulation of astragalosides and related gene expression in different organs of *astragalus membranaceus* Bge. var *mongholicus* (Bge.). Molecules 2014;19:10922–35. https://doi.org/10.3390/molecules190810922.
[19] Li J, Hu Z. Accumulation and dynamic trends of triterpenoid saponin in vegetative organs of *Achyranthus bidentata*. J Integr Plant Biol 2009;51:122–9. https://doi.org/10.1111/j.1744-7909.2008.00764.x.
[20] Madhav KC, Phoboo S, Jha PK. Ecological study of *Paris polyphylla* Sm. Ecoprint An Int J Ecol 2011;17:87–93. https://doi.org/10.3126/ECO.V17I0.4121.
[21] Qin X-J, Ni W, Chen C-X, Liu H-Y. Seeing the light: shifting from wild rhizomes to extraction of active ingredients from above-ground parts of *Paris polyphylla* var. *yunnanensis*. J Ethnopharmacol 2018;224:134–9. https://doi.org/10.1016/j.jep.2018.05.028.
[22] Guan H-Y, Su P, Zhao Y-J, Zhang X-N, Dai Z-B, Guo J, et al. Cloning and functional analysis of two sterol-C24-methyltransferase 1 (SMT1) genes from *Paris polyphylla*. J Asian Nat Prod Res 2017;20:595–604. https://doi.org/10.1080/10286020.2016.1271791.
[23] Negi JS, Bisht VK, Bhandari AK, Bhatt VP, Singh P, Singh N. *Paris polyphylla*: Chemical and biological prospectives. Anti Cancer Agents Med Chem 2014;14:833–9. https://doi.org/10.2174/1871520614666140611101040.
[24] Kang L-p, Yu K, Zhao Y, Liu Y-x, Yu H-s, Pang X, et al. Characterization of steroidal glycosides from the extract of *Paris Polyphylla* var. *yunnanensis* by UPLC/Q-TOF MSE. J Pharm Biomed Anal 2012;62:235–49. https://doi.org/10.1016/j.jpba.2011.12.027.
[25] Seki H, Tamura K, Muranaka T. P450s and UGTs: key players in the structural diversity of triterpenoid saponins. Plant Cell Physiol 2015;56:1463–71. https://doi.org/10.1093/pcp/pcv062.

[26] Yang Y, Zhang X, Yu B. O-glycosylation methods in the total synthesis of complex natural glycosides. Nat Prod Rep 2015;32:1331–55. https://doi.org/10.1039/C5NP00033E.
[27] Patel K, Gadewar M, Tahilyani V, Patel DK. A review on pharmacological and analytical aspects of diosmetin: a concise report. Chin J Integr Med 2013;19:792–800. https://doi.org/10.1007/s11655-013-1595-3.
[28] Gao X, Su Q, Li J, Yang W, Yao B, Guo J, et al. RNA-Seq analysis reveals the important co-expressed genes associated with polyphyllin biosynthesis during the developmental stages of *Paris polyphylla*. BMC Genom 2022;23:559. https://doi.org/10.1186/s12864-022-08792-2.
[29] Gao X, Zhang X, Chen W, Li J, Yang W, Zhang X, et al. Transcriptome analysis of *Paris polyphylla* var. *yunnanensis* illuminates the biosynthesis and accumulation of steroidal saponins in rhizomes and leaves. Phytochemistry 2020;178:112460. https://doi.org/10.1016/j.phytochem.2020.112460.
[30] Hua X, Song W, Wang K, Yin X, Hao C, Duan B, et al. Effective prediction of biosynthetic pathway genes involved in bioactive polyphyllins in *Paris polyphylla*. Commun Biol 2022;5:50. https://doi.org/10.1038/s42003-022-03000-z.
[31] Yang Z, Yang L, Liu C, Qin X, Liu H, Chen J, et al. Transcriptome analyses of *Paris polyphylla* var. *chinensis*, *Ypsilandra thibetica*, and *Polygonatum kingianum* characterize their steroidal saponin biosynthesis pathway. Fitoterapia 2019;135:52–63. https://doi.org/10.1016/j.fitote.2019.04.008.
[32] Christ B, Xu C, Xu M, Li F-S, Wada N, Mitchell AJ, et al. Repeated evolution of cytochrome P450-mediated spiroketal steroid biosynthesis in plants. Nat Commun 2019;10:3206. https://doi.org/10.1038/s41467-019-11286-7.
[33] Song W, Zhang C, Wu J, Qi J, Hua X, Kang L, et al. Characterization of three *Paris polyphylla* glycosyltransferases from different UGT families for steroid functionalization. ACS Synth Biol 2022;11:1669–80. https://doi.org/10.1021/acssynbio.2c00103.
[34] Yuegui C, Qin Y, Yunheng J, Xue B, Desen L, Rongfang M, et al. Unraveling the serial glycosylation in the biosynthesis of steroidal saponins in the medicinal plant *Paris polyphylla* and their antifungal action. Acta Pharm Sin B 2023;13:4638–54. https://doi.org/10.1016/j.apsb.2023.05.033.
[35] Zhou C, Yang Y, Tian J, Wu Y, An F, Li C, et al. 22R- but not 22S-hydroxycholesterol is recruited for diosgenin biosynthesis. Plant J 2021;109:940–51. https://doi.org/10.1111/tpj.15604.
[36] Poppenberger B, Fujioka S, Soeno K, George GL, Vaistij FE, Hiranuma S, et al. The UGT73C5 of *Arabidopsis thaliana* glucosylates brassinosteroids. Proc Natl Acad Sci USA 2005;102:15253–8. https://doi.org/10.1073/pnas.0504279102.
[37] Stucky DF, Arpin JC, Schrick K. Functional diversification of two UGT80 enzymes required for steryl glucoside synthesis in Arabidopsis. J Exp Bot 2015;66:189–201. https://doi.org/10.1093/jxb/eru410.
[38] Kohara A, Nakajima C, Hashimoto K, Ikenaga T, Tanaka H, Shoyama Y, et al. A novel glucosyltransferase involved in steroid saponin biosynthesis in *Solanum aculeatissimum*. Plant Mol Biol 2005;57:225–39. https://doi.org/10.1007/s11103-004-7204-2.
[39] Ye T, Song W, Zhang J-J, An M, Feng S, Yan S, et al. Identification and functional characterization of *Dz*S3GT, a cytoplasmic glycosyltransferase catalyzing biosynthesis of diosgenin 3-*O*-glucoside in *Dioscorea zingiberensis*. Plant Cell Tissue Organ Cult 2017;129:399–410. https://doi.org/10.1007/s11240-017-1187-6.
[40] Gao J, Xu Y, Hua C, Li C, Zhang Y. Molecular cloning and functional characterization of a sterol 3-*O*-glucosyltransferase involved in biosynthesis of steroidal saponins in *Trigonella foenum-graecum*. Front Plant Sci 2021;12:809579. https://doi.org/10.3389/fpls.2021.809579.
[41] Jiang Z, Gao H, Liu R, Xia M, Lu Y, Wang J, et al. Key glycosyltransferase genes of *Panax notoginseng*: identification and engineering yeast construction ofrare ginsenosides. ACS Synth Biol 2022;11:2394–404. https://doi.org/10.1021/acssynbio.2c00094.
[42] Wang L, Jiang Z, Zhang J, Chen K, Zhang M, Wang Z, et al. Characterization and structure-based protein engineering of a regiospecific saponin acetyltransferase from *Astragalus membranaceus*. Nat Commun 2023;14:5969. https://doi.org/10.1038/s41467-023-41599-7.
[43] Li J, Lv M, Du L, Yunga A, Hao S, Zhang Y, et al. An enormous *Paris polyphylla* genome sheds light on genome size evolution and polyphyllin biogenesis. bioRxiv 2020. https://doi.org/10.1101/2020.06.01.126920.
[44] Tamura K, Stecher G, Kumar S, Battistuzzi FU. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol 2021;38:3022–7. https://doi.org/10.1093/molbev/msab120.
[45] He M, Guo S, Yin Y, Zhang C, Zhang X. A novel sterol glycosyltransferase catalyses steroidal sapogenin 3-*O* glucosylation from *Paris polyphylla* var. *yunnanensis*. Mol Biol Rep 2023;50:2137–46. https://doi.org/10.1007/s11033-022-08199-y.
[46] Augustin JM, Drok S, Shinoda T, Sanmiya K, Nielsen JK, Khakimov B, et al. UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyze sapogenin 3-*O*-glucosylation insaponin-mediated insect resistance. Plant Physiol 2012;160:1881–95. https://doi.org/10.1104/pp.112.202747.
[47] Paquette S, Moller BL, Bak S. On the origin of family 1 plant glycosyltransferases. Phytochemistry 2003;62:399–413. https://doi.org/10.1016/S0031-9422(02)00558-7.
[48] Kubo A, Arai Y, Nagashima S, Yoshikawa T. Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. Arch Biochem Biophys 2004;429:198–203. https://doi.org/10.1016/j.abb.2004.06.021.
[49] Cheng J, Chen J, Liu XN, Li XC, Zhang WX, Dai ZB, et al. The origin and evolution of the diosgenin biosynthetic pathway in yam. Plant Commun 2021;2:100079. https://doi.org/10.1016/j.xplc.2020.100079.
[50] Jo S, El-Demerdash A, Owen C, Srivastava V, Wu D, Kikuchi S, et al. Unlocking saponin biosynthesis in soapwort. Nat Chem Biol 2024. https://doi.org/10.1038/s41589-024-01681-7.
[51] Yi D, Bayer T, Badenhorst CPS, Wu S, Doerr M, Höhne M, et al. Recent trends in biocatalysis. Chem Soc Rev 2021;50:8003–49. https://doi.org/10.1039/d0cs01575j.
[52] Xiao Z, Wang Y, Liu J, Zhang S, Tan X, Zhao Y, et al. Systematic engineering of *Saccharomyces cerevisiae* chassis for efficient flavonoid-7-*O*-disaccharide biosynthesis. ACS Synth Biol 2023;12:2740–9. https://doi.org/10.1021/acssynbio.3c00348.
[53] Zerbe P. Plants against cancer: towards green taxol production through pathway discovery and metabolic engineering. aBIOTECH 2024;5:394–402. https://doi.org/10.1007/s42994-024-00170-8.
[54] Jacobowitz JR, Weng J-K. Exploring uncharted territories of plant specialized metabolism in the postgenomicera. Annu Rev Plant Biol 2020;71:631. https://doi.org/10.1146/annurev-arplant-081519-035634. 568.
[55] De Luca V, Salim V, Atsumi SM, Yu F. Mining the biodiversity of plants: a revolution in the making. Science 2012;336:1658–61. https://doi.org/10.1126/science.1217410.
[56] Heping H, Shanlin G, Lanlan C, Xiaoke J. In vitro induction and identification of autotetraploids of *Dioscorea zingiberensis*. In Vitro Cell Dev Biol Plant 2008;44:448–55. https://doi.org/10.1007/s11627-008-9177-3.
[57] Shen S, Wang S, Yang C, Wang C, Zhou Q, Zhou S, et al. Elucidation of the melitidin biosynthesis pathway in pummelo. J Integr Plant Biol 2023;65:2505–18. https://doi.org/10.1111/jipb.13564.
[58] Wang S, Alseekh S, Fernie AR, Luo J. The structure and function of major plant metabolite modifications. Mol Plant 2019;12:899–919. https://doi.org/10.1016/j.molp.2019.06.001.
[59] Yu J, Hu F, Dossa K, Wang Z, Ke T. Genome-wide analysis of UDP-glycosyltransferase super family in *Brassica rapa* and *Brassica oleracea* reveals its evolutionary history and functional characterization. BMC Genom 2017;18:474. https://doi.org/10.1186/s12864-017-3844-x.
[60] Erthmann PØ, Agerbirk N, Bak S. A tandem array of UDP-glycosyltransferases from the UGT73C subfamily glycosylate sapogenins, forming a spectrum of mono- and bisdesmosidic saponins. Plant Mol Biol 2018;97:37–55. https://doi.org/10.1007/s11103-018-0723-z.
[61] Shibuya M, Nishimura K, Yasuyama N, Ebizuka Y. Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in *Glycine max*. FEBS Lett 2010;584:2258–64. https://doi.org/10.1016/j.febslet.2010.03.037.
[62] Casas MI, Falcone-Ferreyra ML, Jiang N, Mejía-Guerra MK, Rodríguez E, Wilson T, et al. Identification and characterization of maize salmon silks genes involved in insecticidal maysin biosynthesis. Plant Cell 2016;28:1297–309. https://doi.org/10.1105/tpc.16.00003.
[63] Yin X, Liu J, Kou C, Lu J, Zhang H, Song W, et al. Deciphering the network of cholesterol biosynthesis in *Paris polyphylla* laid a base for efficient diosgenin production in plant chassis. Metab Eng 2023;76:232–46. https://doi.org/10.1016/j.ymben.2023.02.009.
[64] Cravens A, Payne J, Smolke CD. Synthetic biology strategies for microbial biosynthesis of plant natural products. Nat Commun 2019;10:2142. https://doi.org/10.1038/s41467-019-09848-w.
[65] Chen Y, Huang JP, Wang YJ, Tu ML, Li J, Xu B, et al. Identification and characterization of camptothecin tailoring enzymes in *Nothapodytes tomentosa*. J Integr Plant Biol 2024;66:1158–69. https://doi.org/10.1111/jipb.13649.
[66] Liu Y, Zhao X, Gan F, Chen X, Deng K, Crowe SA, et al. Complete biosynthesis of QS-21 in engineered yeast. Nature 2024;629:937–44. https://doi.org/10.1038/s41586-024-07345-9.
[67] Zhao XY, Zheng L, Si JJ, Miao Y, Peng Y, Cai X. Immunocytochemical localization of saikosaponin-d in vegetative organs of *Bupleurum scorzoneriflolium* Willd. Bot Stud 2013;54:32. https://doi.org/10.1186/1999-3110-54-32.
[68] Sirikantaramas S, Yamazaki M, Saito K. Mechanisms of resistance to self-produced toxic secondary metabolites in plants. Phytochem Rev 2008;7:467–77. https://doi.org/10.1007/s11101-007-9080-2.