RESEARCH



Identification and analysis of *UGT* genes associated with triterpenoid saponin in soapberry (*Sapindus mukorossi* Gaertn.)



Mingzhu Zhou¹, Jialin Fan¹, Yuhan Gao¹, Chunyuan Zheng¹, Yuanyuan Xu², Liming Jia¹, Xinmin An³ and Zhong Chen^{1*}

Abstract

Background Soapberry (*Sapindus mukorossi*) is an economically important multifunctional tree species. Triterpenoid saponins have many functions in soapberry. However, the types of uridine diphosphate (UDP) glucosyltransferases (UGTs) involved in the synthesis of triterpenoid saponins in soapberry have not been clarified.

Results In this study, 42 *SmUGTs* were identified in soapberry, which were unevenly distributed on 12 chromosomes and had sequence lengths of 450 bp to 1638 bp, with an average of 1388 bp. The number of amino acids in SmUGTs was 149 to 545, with an average of 462. Most SmUGTs were acidic and hydrophilic unstable proteins, and their secondary structures were mainly α-helices and random coils. All had conserved UDPGT and PSPG-box domains. Phylogenetic analysis divided them into four subclasses, which glycosylated different carbon atoms. Prediction of *cis*-acting elements suggested roles of *SmUGTs* in plant development and responses to environmental stresses. The expression patterns of *SmUGTs* differed according to the developmental stage of fruits, as determined by transcriptomics and RT-qPCR. Co-expression network analysis of *SmUGTs* and related genes/transcription factors in the triterpenoid saponin synthesis pathway was also performed. The results indicated potential roles for many transcription factors, such as *SmERFs*, *SmGATAs* and *SmMYBs*. A correlation analysis showed that 42 *SmUGTs* were crucial in saponin synthesis in soapberry.

Conclusions Our findings suggest optimal targets for manipulating glycosylation in soapberry triterpenoid saponin biosynthesis; they also provide a theoretical foundation for further evaluation of the functions of *SmUGTs* and analyses of their biosynthetic mechanisms.

Keywords Sapindus mukorossi Gaertn., Triterpenoid saponins, UGT, Expression pattern

*Correspondence:

¹ State Key Laboratory for Efficient Production of Forest Resources, Key Laboratory of Silviculture and Conservation of the Ministry of Education, Ministry of Education of Engineering Research Centre for Forest and Grassland Carbon Sequestration, College of Forestry, Beijing Forestry University, P. O. Box 407, No.35 Qinghua East Road, Haidian District, Beijing 100083, China

 ² College of Forestry, Guangxi University, Nanning 530004, China
 ³ National Engineering Research Center of Tree Breeding and Ecological Restoration, College of Biological Sciences and Biotechnology, Beijing Forestry University, Beijing 100083, China

Background

Soapberry (*Sapindus mukorossi*) is a deciduous tree of the Sapindaceae family, which is mainly distributed in east, south, and southwest China, and is used as a source of daily chemicals, medicines, and biomass energy. In one study, the triterpenoid saponin content in the peel of soapberry ranged from 4.14% to 27.04% [1, 2]. Triterpenoid saponins are the main soapberry components with medicinal value and play an important role in plant resistance to bacteria. In addition, because of their surfactant properties saponins are used in the manufacture



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.gr/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.gr/licenses/by/4.0/. The Creative Commons Public Domain and redit line to the data.

Zhong Chen

zhongchen@bjfu.edu.cn

of chemicals such as shampoo [3]. These specialized metabolites are glycoside compounds formed by condensation of a hydrophobic triterpenoid saponin and one or more hydrophilic glycogroups [4]. They are mainly present in plants in families such as Araliaceae, Leguminosae, Cucurbitaceae, and Umbelliferae [5]. Saponins participate in physiological processes such as communication, defense, and perception in plants [6], and have antioxidant, anti-inflammatory, and anticancer effects [7, 8]. They can be used in the production of herbicides, insecticides [9], and emulsifiers [10]. Saponins are often the main bioactive components of several plants. The main pharmacological components of Panax ginseng, Panax quinquefolius, and Panax notoginseng are saponins of the ginsenoside type [11]. Glycyrrhizin in Glycyrrhiza glabra is also a pharmacologically relevant saponin [12]. Mogrosides from Siraitia grosvenorii are used in the manufacture of natural sweeteners [13]. The structural diversity of triterpenoid saponins has led to widespread interest in their applications in medicine, food, daily chemical products, agriculture, and other fields. The complex structure and frequently low content of saponins in plant biomass make it challenging to directly extract, isolate, or obtain them through chemical synthesis, thus restricting their wide application. Biotechnology methodologies have been used to study the biosynthetic pathways of triterpenoid saponins to increase the content of target metabolites or produce them using synthetic biological methods.

The biosynthesis of triterpenoid saponins begins with isopentenyl pyrophosphate (IPP, C5) and dimethylallyl pyrophosphate (DMAPP, C5). IPP and DMAPP are synthesized through two distinct pathways: the mevalonate (MVA) pathway in the cytoplasm and the 2-C-methyld-erythritol-4-phosphate (MEP) pathway in plastids. The MVA pathway utilizes acetyl-CoA as its initial substrate and undergoes a six-step condensation reaction to generate IPP. In contrast, the MEP pathway involves a seven-step reaction to synthesize IPP from pyruvate and glyceraldehyde 3-phosphate as starting substrates. The MVA pathway plays a key role in triterpenoid saponin biosynthesis. Subsequently, IPP and DMAPP are condensed by geranyl pyrophosphate synthase (GPS) to produce geranyl pyrophosphate (GPP, C10). GPP is then catalyzed by farnesyl pyrophosphate synthase (FPS); this is followed by the addition of a second IPP unit to produce farnesyl pyrophosphate (FPP, C15). The condensation of two FPP molecules by squalene synthase (SS) results in the formation of squalene (C30), which is converted to 2,3-oxidosqualene through epoxidation. Finally, 2,3-oxidosqualene undergoes cyclization mediated by oxidized squalene cyclase (OSC), oxidation mediated by cytochrome P450 enzymes, and glycosylation mediated by uridine diphosphate glucosyltransferase (UGT) [5, 14–16].

UGT glycosyltransferases of the glycosyltransferase 1 family (GT1) transfer active glycosylates from uracil nucleoside diphosphate sugars to small-molecule receptors in plants [17], and glycosyltransferases associated with triterpenoid saponin biosynthesis in soapberry belong to this class. The sequence of the N-terminal region of these enzymes is highly diverse and is responsible for the recognition of multiple substrates, whereas the C-terminal region contains a conserved plant secondary product glycosyltransferase (PSPG) motif. The UGTcatalyzed glycosylation of triterpenoids is commonly considered the final stage in the biosynthesis of triterpenoid saponins; it plays a crucial role in generating diversity and biological activity. A typical triterpenoid saponin contains 2 to 5 glycans, which usually form a sugar chain at the C-3 and/or C-28 position of the saponin. Triterpenoid saponins glycosylated at C-4, C-16, C20, C-21, C-22, and/or C-23 are also present. Common glycans include glucose (Glc), galactose (Gal), glucuronic acid (GlcA), rhamnose (Rha), xylose (Xyl), and arabinose (Ara), as well as other five- or six-membered sugars. The number, composition, and position of sugar chains on the triterpenoid skeleton impact the bioactivity, stability, solubility, and cellular uptake of triterpenoid saponins [5].

UGTs in plants are organized according to their amino acid sequence identity into gene families (identity > 40%) and subfamilies (identity > 60%) [6]. In all, 107 UGTs have been identified in Arabidopsis thaliana, distributed in 14 groups (A-N) [18]; 182 UGTs in soybean, distributed in 16 groups (A-P) [19]; 147 UGTs in maize, distributed in 17 groups (A-Q) [20]; and 241 UGTs in apple, distributed in 16 groups (A-P) [21]. The members of different subfamilies can have different functions. For example, most UGT73s have glycosylation activity at C-3 or C-28, whereas UGT74s and UGT94s typically catalyze glycosylation at C-3 [5]. Regarding functions, UGTPg1 produces ginsenoside F1 through region-specific glycosylation of C20-OH of proginsenodiol (PPD) and proginsenotriol (PPT). UGTPg100 specifically glycosylates C6-OH of PPT to produce ginsenoside Rh1 [22]. Similarly, PG3-O-UgT1 from Panax quinquefolium catalyzes the conversion of proginsenediol into ginsenoside Rh2 [23, 24]. Several UGT enzymes are related to plant resistance in Barbarea vulgaris [25]. A number of studies have focused on the UGT genes in Medicago truncatula [26], Centella asiatica [27], cucumber [28], Glycyrrhiza glabra [12], and Siraitia grosvenorii [13].

Research on triterpenoid saponins in soapberry has focused on their extraction, separation, and application. For example, Zhang and Jia investigated the pharmacological effects of soapberry [29]. Peng et al. optimized the extraction of saponins from soapberry [30]. Ma evaluated the use of soapberry extract as an herbicide [31]. However, there has been little research on triterpenoid saponin biosynthesis and the key enzymes in glycosylation. In this study, we constructed a phylogenetic tree and analyzed the expression patterns of *SmUGTs* related to triterpenoid saponin synthesis, based on available genomic and transcriptomic data of soapberry [32]. The findings provided theoretical support for functional research on soapberry genes related to triterpenoid saponin synthesis.

Results

Identification and structural analysis of SmUGT genes

In total, 42 *SmUGTs* were identified, which were successively named *SmUGT1* to *SmUGT42*. Gene structure analysis (Fig. 1) showed that 23 genes did not contain introns. Eighteen genes had one intron and one gene had two introns; most genes containing introns clustered into one group. A protein domain analysis (Figure S1) showed that the SmUGTs contained the Pfam: UDPGT domain typical of UGT proteins, which were close to the *C*-terminus, indicating structural consistency. The 42 *SmUGTs* were distributed among 12 chromosomes, and chr02 had the most genes (14).

Physicochemical properties of SmUGT proteins

The average length of these genes was 1388 bp, the longest was 1638 bp, and the shortest was 450 bp (Table 1). Except for *SmUGT14*, the genes were > 1000 bp in length. The number of amino acids ranged from 149 to 545, with an average of 462. Other than SmUGT14 and SmUGT16, the SmUGT proteins were composed of > 400 amino acids. The average molecular weight was 51,764.76 Da and the theoretical isoelectric points of 41 of the proteins were <7, indicating acidity. That of SmUGT14 was 8.43, indicating alkalinity. In all, 33 were hydrophilic and 9 were hydrophobic; 11 were stable (instability coefficient <40) and 31 were unstable (instability coefficient > 40). Most were predicted to localize in the cytoplasm, chloroplasts, and nucleus.

Conserved domains of SmUGT proteins

A sequence comparison performed in MEGA was visually analyzed in MEME v5.5.5, and the top 30 motifs were selected. The 42 SmUGTs harbored 5 to 18 motifs; all contained motif 1 and motif 4 (there was no complete motif 4 in SmUGT38, SmUGT39, and SUGT42 due to insertion of two amino acids in the sequence), the conserved PSPG-box region; all except SmUGT3, SmUGT14, SmUGT16, and SmUGT17 had motif 2 (Fig. 2). The PSPG-box is a highly conserved sequence of 44 amino acids with highly conserved

positions 1 (W), 4 (Q), 7 (I/V), 8 (L), 12, 18 (S/A) (S/T), 19 (H) and 21 (G), 23 (N/S/G), 24 (S), 27 (E), 39 (P), 43 (E/D), and 44 (Q) [33]. SmUGT15-SmUGT21 and SmUGT25 had identical motif sequences, as did SmUGT31, SmUGT32, SmUGT33, and SmUGT34. Those motif sequences were almost identical in several proteins.

Analysis of cis-acting elements of SmUGT genes

The cis-acting elements of SmUGTs with high frequencies (Figure S2) were screened and plotted (Fig. 3). The predicted functions of these elements were related to those identified by gene mining. They were categorized as elements responding to hormones, such as auxin, abscisic acid and gibberellin; stress response elements (e.g., light, drought, and oxygen); and elements regulating plant growth and development [34]. Among the selected elements, the developmental component was the A-box. Box 4, CAT-box, G-Box, GATA-motif, and WRE3 were related to plant photoperiod sensing; ARE, MYB, LTR, and W-box were associated with environmental stress regulation; O2-site and STRE were related to circadian rhythm regulation; and CCAAT-box was associated with hormone regulation. Among the 42 SmUGT promoters, 293 contained Box 4 and 280 had MYB. The numbers and positions of *cis*-acting elements differed among the SmUGTs (Figure S2).

Phylogenetic analysis of SmUGT proteins

We screened the UGT sequences of *Barbarea vulgaris*, *Arabidopsis thaliana*, *Panax ginseng*, *Medicago truncatula*, *Glycine max*, *Panax quinquefolius*, *Centella asiatica*, *Saponaria vaccaria*, and *Siraitia grosvenorii* using NCBI. A phylogenetic tree was constructed by combining their protein sequences with those of SmUGTs (Fig. 4).

The 42 SmUGTs belonged to the UGT71, UGT73, UGT74, and UGT94 subfamilies, classified into groups D, E, F, and G [33]. There were 8 members of UGT71, 8 of UGT73, 16 of UGT74, and 10 of UGT94. The 10 UGTs associated with triterpenoid saponin synthesis in *Barbarea vulgaris* were members of UGT73, whereas those in ginseng were members of UGT71, UGT74, and UGT94. In *Medicago truncatula*, four SmUGTs were members of UGT73 and one of UGT71. The four SmUGTs in soybean were members of UGT73.

Collinearity analysis (Fig. 5) showed that 8 *SmUGT* s had collinear relationships with *UGT*s in *Arabidopsis thaliana* and 10 did with *UGT*s in *Medicago truncatula*; these were mainly distributed among chr02, chr09, chr10, chr11, chr13, and chr14. Therefore, soapberry is more



Fig. 1 Structures and chromosome locations of *SmUGT* genes. **a** Exon and intron structures of *SmUGT* genes. The CDS is the coding region of the protein; the UTR is the untranslated region, a non-coding fragment at each end of the mRNA molecule; the intron, also known as spacer order, refers to a fragment of a gene or mRNA mole. **b** Chromosomal locations of *SmUGTs*. 42 *SmUGTs* were unevenly distributed on 12 chromosomes

Table 1. The physicochemical properties of the SmUGT genes

| Gene name | Gene ID | Genome location | Number of amino acids | PI/Molecula r weight (Da) | GRAVY | Instabilitly index | CDS length (bp) | Predicted subcellular localization |
|-----------|---------------------|-----------------------------|--------------------------|---------------------------------|--------|-----------------------|--------------------|--|
| SmUGT1 | Samuk01G0 040600 | Chor01:4141 244-4143406 | 486 | 5.34/54751.1 | -0.066 | 43.92 | 1461 | chlo |
| SmUGT2 | Samuk01G0 | Chor01:4148 | 545 | 5.17/61425.1 | -0.161 | 49.82 | 1638 | cyto/nucl |
| SmUCT2 | Samuk01G0 | Chor01:1168 | 437 | 6.19/49799.8 | 0.084 | 48.08 | 1214 | mito (ablo |
| 3///013 | 099100 | 4778- | 437 | 0 | -0.084 | 40.00 | 1314 | nnocino |
| SmUGT4 | Samuk01G0 099200 | 4771- | 532 | 6.49/60174.1 7 | -0.264 | 46.76 | 1599 | cyto/chlo/nuc |
| | Samuk01G0 | 11698483 Chor01:1398 | | 6 16/52250 0 | | | | arto/oblo/nuo |
| SmUGT5 | 112300 | 1735- 13983121 | 461 | 5 | -0.286 | 50.01 | 1386 | l |
| SmUGT6 | Samuk01G0 | Chor01:1399 9855- | 461 | 6.01/52290.0 | -0.284 | 51.09 | 1386 | cvto/nucl |
| | 112400 | 14001241 Chor01:1802 | | 4 | | | | 2 |
| SmUGT7 | Samuk01G0 126900 | 9868- | 453 | 6.27/50917.7 6 | -0.247 | 50.32 | 1362 | cyto/chlo |
| C. UCTR | Samuk01G0 | Chor01:1811 | 453 | 6.27/51036.8 | 0.264 | 50.07 | 12/2 | |
| 3///018 | 127000 | 18118319 | 455 | 4 | -0.204 | 50.97 | 1302 | cyto/enito |
| SmUGT9 | 029600 | 472-5352755 | 459 | 4 | 0.031 | 47.53 | 1380 | cyto/nucl |
| SmUGT10 | Samuk02G0 029800 | Chor02:5379 320-5381573 | 459 | 5.70/51242.1 9 | 0.024 | 42.93 | 1380 | cyto/nucl |
| SmUGT11 | Samuk02G0 | Chor02:3389 3474- | 484 | 5.27/54550.4 | -0.223 | 46.16 | 1455 | chlo/cyto/per |
| | Eamul-02C0 | 33895325 Chor02:3845 | | - | | | | 0 |
| SmUGT12 | 148100 | 3223- 38454657 | 477 | 1 | -0.211 | 31.24 | 1434 | cyto/chlo |
| SmUGT13 | Samuk02G0 | Chor02:3847 7198- | 477 | 6.48/53708.7 | -0.191 | 29.24 | 1434 | cyto/nucl/cys |
| | 148200 | 38478632 Chor02:3858 | | 5 | | | | k |
| SmUGT14 | Samuk02G0 149100 | 0311- | 149 | 8.43/16234.7 3 | -0.209 | 35.35 | 450 | nucl/plas |
| SmIICT15 | Samuk02G0 | Chor02:5467 | 473 | 4.81/52515.5 | 0.122 | 48.02 | 1422 | plos/E P |
| SmcGIIS | 311900 | 5289- 54676711 | 473 | 0 | 0.122 | 48.92 | 1422 | pias/E.K. |
| SmUGT16 | Samuk02G0 312000 | 2793- | 394 | 5.20/43937.3 7 | 0.033 | 47.66 | 1185 | chlo/vacu |
| | Samuk02G0 | 54684104 Chor02:5468 | | 5 39/47764 9 | | | | |
| SmUGT17 | 312100 | 5964- 54687260 | 431 | 9 | 0.026 | 47.32 | 1296 | cyto/nucl |
| SmUGT18 | Samuk02G0 | Chor02:5469 3715- | 466 | 5.78/51782.4 | -0.021 | 40.75 | 1401 | chlo/cyto |
| | Samuk02G0 | 54695116 Chor02:5471 | | 5 10/52343 7 | | | | |
| SmUGT19 | 312400 | 0111- 54711521 | 469 | 5 | 0.132 | 47.32 | 1410 | cyto/vacu |
| SmUGT20 | Samuk02G0 | Chor02:5471 2296- | 472 | 5.46/52401.9 | 0.009 | 44.39 | 1419 | vacu/chlo |
| | 312300 | 54713715 Chor02:5471 | | 5 | | | | |
| SmUGT21 | 312600 | 7428- 54718841 | 470 | 8 | 0.026 | 33.45 | 1413 | cyto/E.R. |
| SmUGT22 | Samuk02G0 | Chor02:5546 | 456 | 5.38/50720.6 | 0.057 | 49.27 | 1371 | evto/nucl |
| | 322600 | 55466934 Chor03:3772 | | 4 | | | | |
| SmUGT23 | Samuk03G0 158600 | 5929- | 471 | 5.74/53069.2 9 | -0.029 | 37.48 | 1416 | chlo |
| SmUGT24 | Samuk05G0 | Chor05:5028 | 455 | 6.25/50636.6 | -0.052 | 33.82 | 1368 | chlo |
| SmUCT25 | Samuk06G0 | Chor06:1483 | 457 | 9 5.54/51321.3 | 0.011 | 60.21 | 1274 | |
| 3m00125 | 115000 | 14835038 | 437 | 7 | -0.011 | 50.51 | 1374 | pero |
| SmUGT26 | 035700 | 141-2902738 | 449 | 2 | -0.035 | 34.12 | 1350 | chlo |
| SmUGT27 | 061600 | Chor09:4852 189-4853666 | 458 | 5.71/51/83.4 6 | -0.153 | 46.96 | 1377 | chlo/vacu |
| SmUGT28 | Samuk09G0 068300 | Chor09:5507 131-5508622 | 496 | 6.16/55012.6 4 | -0.033 | 50.08 | 1491 | chlo/cyto |
| SmUGT29 | Samuk09G0 068400 | Chor09:5509 462-5510932 | 489 | 5.64/54721.5 1 | -0.014 | 38.18 | 1470 | chlo |
| SmUGT30 | Samuk09G0 068500 | Chor09:5520 766-5522212 | 481 | 5.96/54481.0 8 | -0.127 | 42.76 | 1446 | chlo/cyto |
| SmUGT31 | Samuk10G0 005900 | Chor10:3636 58-365986 | 457 | 5.37/50772.2 8 | -0.086 | 31.18 | 1374 | cyto/nucl |
| SmUGT32 | Samuk10G0 006000 | Chor10:3685 64-370300 | 459 | 5.53/50816.0 8 | -0.098 | 41.19 | 1380 | chlo |
| SmUGT33 | Samuk10G0 | Chor10:3742 | 470 | 5.55/52053.8 | -0.121 | 42.25 | 1413 | chlo/cyto |
| SmUGT34 | Samuk10G0 | Chor10:3872 | 460 | 5.46/50971.4 | -0.173 | 40.34 | 1383 | chlo |
| SmIICT25 | Samuk10G0 | Chor10:1189 | 670 | 6.71/58124.4 | 0.268 | 49.26 | 1617 | malforito |
| 5//00155 | 135200 Samuk11C0 | 11899340 Charl118226 | 556 | 1 | -0.500 | 40.50 | 1017 | nuciveyto |
| SmUGT36 | 083900 | 529-8238495 Ch 12-7062 | 490 | 5 | -0.108 | 39.58 | 1473 | chlo/nucl |
| SmUGT37 | 007500 | 610-7864056 Chor12:20.12 | 450 | 3 | -0.01 | 43.07 | 1353 | chlo |
| SmUGT38 | Samuk12G0 091700 | 6342- | 463 | 5.85/52554.1 6 | -0.141 | 49.3 | 1392 | chlo |
| | Samuk12G0 | 20407734 Chor12:2041 | | 5.70/53014.6 | | | | |
| SmUG139 | 091800 | 4176- 20415583 | 408 | 0 | -0.18 | 51.02 | 1407 | enio/mito |
| SmUGT40 | samuk13G0 090300 | Chor13:8325 421-8326780 | 452 | 5.49/50393.3 0 | -0.163 | 44.86 | 1359 | chlo/nucl |
| SmUGT41 | Samuk14G0 058800 | 0373- | 477 | 5.81/53734.6 5 | -0.196 | 37.67 | 1434 | chlo/pero |
| 0 1/07/- | Samuk14G0 | 16893351 Chor14:1784 | 101 | 5.40/55410.5 | 0.21 | 41.54 | 170 | |
| SmUGT42 | 066500 | 0117- 17841578 | 486 | 7 | -0.21 | 41.24 | 1461 | chio/cyto |

| SmUGT1 | Name | Motif Locations |
|--|------------|--|
| SmUGT2 SmUGT3 SmUGT4 SmUGT5 SmUGT5 SmUGT6 SmUGT7 SmUGT7 SmUGT8 SmUGT9 SmUGT9 SmUGT1 SmUGT2 SmUGT3 SmUGT3 SmUGT3 SmUGT3 SmUGT3 SmUGT3 SmUGT3 SmUGT3 SmUGT3 | SmUGT1 | |
| SmUGT3 SmUGT4 SmUGT5 SmUGT6 SmUGT7 SmUGT8 SmUGT9 SmUGT1 SmUGT2 SmUGT3 | SmUGT2 | |
| SmUGT4 | SmUGT3 | |
| SmUGT5 SmUGT6 SmUGT7 SmUGT8 SmUGT9 SmUGT9 SmUGT10 SmUGT11 SmUGT12 SmUGT13 SmUGT14 SmUGT15 SmUGT14 SmUGT14 SmUGT15 SmUGT14 SmUGT15 SmUGT14 SmUGT15 SmUGT14 SmUGT14 SmUGT15 SmUGT14 SmUGT12 SmUGT2 SmUGT3 SmUGT3 <tr< td=""><td>SmUGT4</td><td></td></tr<> | SmUGT4 | |
| SmUGT6 | SmUGT5 | |
| SmUGT7 | SmUGT6 | |
| SmUGT8 | SmUGT7 | |
| SmUGT9 | SmUGT8 | |
| SmUGT10 SmUGT12 SmUGT12 SmUGT13 SmUGT14 SmUGT14 SmUGT14 SmUGT14 SmUGT16 SmUGT16 SmUGT16 SmUGT17 SmUGT20 SmUGT2 | SmUGT9 | |
| SmUGT11 SmUGT12 SmUGT13 SmUGT14 SmUGT14 SmUGT14 SmUGT14 SmUGT17 SmUGT18 SmUGT17 SmUGT18 SmUGT2 SmUGT3 SmUGT4 SmUGT | SmUGT10 | |
| SmUGT12 SmUGT14 SmUGT14 SmUGT14 SmUGT15 SmUGT17 SmUGT18 SmUGT18 SmUGT20 SmUGT20 SmUGT20 SmUGT20 SmUGT20 SmUGT20 SmUGT21 SmUGT21 SmUGT21 SmUGT24 SmUGT25 SmUGT25 SmUGT25 SmUGT26 SmUGT26 SmUGT26 SmUGT30 SmU | SmUGT11 | |
| SmUGT13 SmUGT14 SmUGT15 SmUGT17 SmUGT17 SmUGT18 SmUGT19 SmUGT19 SmUGT20 SmUGT21 SmUGT21 SmUGT22 SmUGT23 SmUGT24 SmUGT25 SmUGT26 SmUGT27 SmUGT28 SmUGT29 SmUGT29 SmUGT20 SmUGT30 SmUGT30 <td< td=""><td>SmUGT12</td><td></td></td<> | SmUGT12 | |
| SmUGT14 SmUGT15 SmUGT16 SmUGT17 SmUGT18 SmUGT19 SmUGT10 SmUGT12 SmUGT12 SmUGT2 SmUGT3 SmUGT4 < | SmUGT13 | |
| SmUGT16 SmUGT17 SmUGT17 SmUGT19 SmUGT19 SmUGT20 SmUGT21 SmUGT23 SmUGT23 SmUGT24 SmUGT25 SmUGT26 SmUGT26 SmUGT27 SmUGT27 SmUGT27 SmUGT28 SmUGT28 SmUGT29 SmUGT30 SmUGT30 SmUGT30 SmUGT31 SmUGT31 SmUGT31 SmUGT31 SmUGT32 SmUGT32 SmUGT33 SmUGT34 SmUGT34 SmUGT34 SmUGT35 SmUGT34 SmUGT34 SmUGT35 SmUGT34 SmUGT35 SmUGT34 SmUGT35 SmUGT34 SmUGT34 SmUGT35 SmUGT34 SmUGT35 SmUGT34 SmUGT34 SmUGT35 SmUGT34 SmUGT35 SmUGT36 SmUGT37 SmUGT36 SmUGT37 SmUGT37 SmUGT38 SmUGT39 SmUGT39 SmUGT39 SmUGT39 SmUGT39 SmUGT30 SmUGT40 SmU | SmUGT14 | |
| SmUGT16 SmUGT17 SmUGT17 SmUGT19 SmUGT20 SmUGT21 SmUGT21 SmUGT21 SmUGT23 SmUGT24 SmUGT25 SmUGT25 SmUGT26 SmUGT26 SmUGT27 SmUGT27 SmUGT27 SmUGT27 SmUGT28 SmUGT29 SmUGT30 SmUGT30 SmUGT30 SmUGT30 SmUGT31 SmUGT31 SmUGT31 SmUGT31 SmUGT31 SmUGT32 SmUGT34 SmUGT35 SmUGT34 SmUGT35 SmUGT36 SmUGT36 SmUGT36 SmUGT37 SmUGT37 SmUGT38 SmUGT38 SmUGT39 SmUGT40 SmUGT41 SmUGT41 SmUGT42 SmUGT40 SmUGT40 SmUGT41 SmUGT42 SmUGT40 SmUGT41 SmUGT42 SmUGT40 SmUGT40 SmUGT40 SmUGT41 SmUGT42 SmUGT40 SmUGT40 SmUGT40 SmUGT41 SmUGT42 SmUGT40 SmUGT40 SmUGT40 SmUGT40 SmUGT40 SmUGT41 SmUGT40 SmU | SmUGT15 | |
| SmUGT17 SmUGT18 SmUGT19 SmUGT20 SmUGT20 SmUGT21 SmUGT21 SmUGT23 SmUGT24 SmUGT25 SmUGT25 SmUGT26 SmUGT30 SmUGT30 SmUGT30 SmUGT30 SmUGT31 SmUGT31 SmUGT31 SmUGT34 SmUGT35 SmUGT35 SmUGT36 SmUGT36 SmUGT36 SmUGT36 SmUGT37 SmUGT37 SmUGT37 SmUGT38 SmUGT38 SmUGT39 SmUGT39 SmUGT39 SmUGT39 SmUGT30 SmUGT40 SmUGT41 SmUGT41 SmUGT42 SmUGT42 SmUGT42 SmUGT40 SmU | SmUGT16 | |
| SmUGT18 SmUGT20 SmUGT21 SmUGT21 SmUGT22 SmUGT23 SmUGT23 SmUGT24 SmUGT25 SmUGT25 SmUGT26 SmUGT26 SmUGT26 SmUGT27 SmUGT28 SmUGT28 SmUGT30 SmUGT30 SmUGT31 SmUGT31 SmUGT31 SmUGT34 SmUGT35 SmUGT35 SmUGT35 SmUGT35 SmUGT36 SmUGT37 SmUGT37 SmUGT37 SmUGT38 SmUGT37 SmUGT39 SmUGT40 SmUGT40 SmUGT41 SmUGT41 SmUGT41 SmUGT41 SmUGT42 SmUGT40 SmUGT41 SmUGT42 SmUGT41 SmUGT42 SmU | SmUGT17 | |
| SmUGT19 SmUGT20 SmUGT21 SmUGT22 SmUGT23 SmUGT23 SmUGT24 SmUGT25 SmUGT25 SmUGT25 SmUGT26 SmUGT26 SmUGT27 SmUGT28 SmUGT30 SmUGT30 SmUGT30 SmUGT31 SmUGT31 SmUGT32 SmUGT32 SmUGT34 SmUGT35 SmUGT35 SmUGT35 SmUGT36 SmUGT36 SmUGT37 SmUGT37 SmUGT38 SmUGT37 SmUGT39 SmUGT40 SmUGT40 SmUGT41 SmUGT41 SmUGT41 SmUGT41 SmUGT42 SmUGT40 SmUGT40 SmUGT41 SmUGT42 SmUGT42 SmUGT40 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT44 SmUGT42 SmUGT44 SmUGT42 SmUGT44 SmU | SmUGT18 | |
| SmUGT20 SmUGT21 SmUGT22 SmUGT23 SmUGT24 SmUGT25 SmUGT26 SmUGT26 SmUGT27 SmUGT28 SmUGT28 SmUGT28 SmUGT29 SmUGT30 SmUGT30 SmUGT31 SmUGT31 SmUGT31 SmUGT32 SmUGT33 SmUGT34 SmUGT35 SmUGT35 SmUGT36 SmUGT36 SmUGT37 SmUGT37 SmUGT37 SmUGT38 SmUGT38 SmUGT39 SmUGT39 SmUGT40 SmUGT41 SmUGT41 SmUGT41 SmUGT42 SmUGT42 SmUGT40 SmUGT41 SmUGT41 SmUGT41 SmUGT41 SmUGT41 SmUGT41 SmUGT41 SmUGT41 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT44 SmU | SmUGT19 | |
| SmUGT21 SmUGT22 SmUGT23 SmUGT24 SmUGT25 SmUGT26 SmUGT27 SmUGT27 SmUGT28 SmUGT27 SmUGT29 SmUGT30 SmUGT31 SmUGT31 SmUGT31 SmUGT33 SmUGT34 SmUGT34 SmUGT34 SmUGT35 SmUGT35 SmUGT36 SmUGT36 SmUGT37 SmUGT36 SmUGT37 SmUGT37 SmUGT38 SmUGT38 SmUGT39 SmUGT40 SmUGT42 SmUGT44 SmU | SmUGT20 | |
| SmUGT22 SmUGT23 SmUGT24 SmUGT25 SmUGT25 SmUGT26 SmUGT27 SmUGT27 SmUGT28 SmUGT29 SmUGT30 SmUGT30 SmUGT31 SmUGT31 SmUGT32 SmUGT33 SmUGT34 SmUGT34 SmUGT35 SmUGT35 SmUGT36 SmUGT36 SmUGT37 SmUGT37 SmUGT36 SmUGT37 SmUGT37 SmUGT38 SmUGT39 SmUGT40 SmUGT40 SmUGT41 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT42 SmUGT44 SmUGT42 SmUGT42 SmUGT44 SmUGT42 SmUGT44 SmUGT42 SmUGT44 SmUGT42 SmUGT44 SmUGT42 SmUGT44 SmUGT44 SmUGT42 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT44 SmUGT45 SmUGT45 SmUGT45 SmUGT45 SmUGT46 SmU | SmUGT21 | |
| SmUGT23 SmUGT24 SmUGT25 SmUGT25 SmUGT26 SmUGT27 SmUGT28 SmUGT28 SmUGT30 SmUGT30 SmUGT31 SmUGT31 SmUGT33 SmUGT33 SmUGT33 SmUGT34 SmUGT35 SmUGT35 SmUGT36 SmUGT36 SmUGT37 SmUGT36 SmUGT37 SmUGT37 SmUGT38 SmUGT38 SmUGT39 SmUGT40 SmUGT41 SmUGT42 SmU | SmUGT22 | |
| SmUGT24 SmUGT25 SmUGT26 SmUGT27 SmUGT28 SmUGT28 SmUGT29 SmUGT30 SmUGT30 SmUGT31 SmUGT31 SmUGT32 SmUGT33 SmUGT33 SmUGT34 SmUGT35 SmUGT36 SmUGT36 SmUGT36 SmUGT37 SmUGT37 SmUGT37 SmUGT38 SmUGT38 SmUGT39 SmUGT40 SmUGT42 SmU | SmUGT23 | |
| SmUGT25 SmUGT26 SmUGT27 SmUGT28 SmUGT28 SmUGT29 SmUGT30 SmUGT31 SmUGT31 SmUGT31 SmUGT32 SmUGT33 SmUGT33 SmUGT34 SmUGT35 SmUGT36 SmUGT36 SmUGT37 SmUGT36 SmUGT37 SmUGT38 SmUGT37 SmUGT38 SmUGT39 SmUGT40 SmUGT40 SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1 UKM IDSQPESSIVY VS rCSERCIS ISE SAG CLAR SUPPORT 1 UKM IDSQPESSI IST SAG SUPPORT 1 UKM IDSQPESSI IST SAG SUPPORT 2 UKM IDSQPESSI IST SAG SUPPORT 2 UKM IDSQPESSI IST SAG 2 UKM IDSCRESSI IST SAG 2 UKM IDSCRESSI IST SAG 2 UKM | SmUGT24 | |
| SmUGT26 SmUGT27 SmUGT28 SmUGT28 SmUGT29 SmUGT30 SmUGT31 SmUGT31 SmUGT32 SmUGT33 SmUGT33 SmUGT34 SmUGT35 SmUGT36 SmUGT36 SmUGT36 SmUGT37 SmUGT38 SmUGT37 SmUGT38 SmUGT39 SmUGT40 SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1 UKMAPG/DEVLAHSVCFLTHCONNSTLESJSAG 2 UKMLDSQPESSWVYVSrCSECSLSPEQ/MEEIANGLE 1 UKMAPG/DEVLAHSVCFLTHCONNSTLESJSAG 2 UKMLDSQPESSWVYVSrCSECSLSPEQ/MEEIANGLE 3 WVTTPELVERSVCFLTHCONNSTLESJSAG 2 UKMLDSQPESSWVYVSrCSECSLSPEQ/MEEIANGLE 3 WVTTPELVERSVCFLTHCONNSTLESJSAG 2 UKMLDSQPESSWVYVSrCSECSLSPEQ/MEEIANGLE 3 WVTTPELVERSVCFLTHCONNSTLESJSAG 2 UKMLDSQPESSWVYVSrCSECSLSPEQ/MEEIANGLE 3 WVTTPELVERSVVVVSrCSECSLSPEQ/MEEIANGLE 3 WVTTPELVERSVVVVSrCSECSLSPEQ/MEEIANGLE 3 WVTTPELVERSVVVVSrCSECSLSPES/MEEIANGLE 3 WVTTPELVERSVVVVSrCSECSLSPES/MARCH 4 WVTTPELVERSVVVVSrCSECSLSPES/MARCH 5 WVTTPELVERSVVVVSrCSECSLSPES/MARCH 5 WVTTPELVERSVVVVSrCSECSLSPES/MARCH 5 WVTTPELVERSVVVVSVCSLSPES/MARCH 5 WVTTPELVERSVVVVSVSVSVSVVVVSVSVSVSVVVSVSVSVSVSVS | SmUGT25 | |
| SmUGT27 SmUGT28 SmUGT28 SmUGT29 SmUGT30 SmUGT31 SmUGT31 SmUGT32 SmUGT33 SmUGT33 SmUGT34 SmUGT35 SmUGT36 SmUGT36 SmUGT36 SmUGT37 SmUGT37 SmUGT38 SmUGT37 SmUGT38 SmUGT39 SmUGT40 SmUGT41 SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1 UKMIDSQPESSIVYVS FCSEGLISPEQMEEIANGLE 1 UKMIDSQPESSIVYVS FCSEGLISPEQMEEIANGLE 1 UKMIDSQPESSIVYVS FCSEGLISPEQMEEIANGLE 1 UKMIDSQPESSIVYVS FCSEGLISPEQMEEIANGLE 3 WVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACOGLINPELLAREXASKGFK VENVTFIPACIALAREXASKGFK VENVTFIPACIALAREXASKGFK VENVTFIPACIALAREXASKGFK VENVTFIPACIALAREXASKGFK VENVTFIPACIALAREXASKGFK VENTORULAREXASKGFK SENVTFIPACIALAREXASKGFK VENTORULAREXASKGFK VENTORULAREXASKGFFF VENTORULAREXASKGFFF VENTORULAREXASKGFFF VENTORULAREXASKGFFF VENTORULAREXASKGFFF VENTORULAREXASKGFFF VENTORULAREXASKGFFF VENTORULAREXASKGFFF VENTORULAREXASKGFFFF VENTORULAREXASKGFFFF VENTORULAREXASKGFFFFF VENTORULARE | SmUGT26 | |
| SmUGT28 SmUGT29 SmUGT30 SmUGT31 SmUGT32 SmUGT33 SmUGT33 SmUGT34 SmUGT35 SmUGT36 SmUGT36 SmUGT37 SmUGT36 SmUGT37 SmUGT38 SmUGT37 SmUGT38 SmUGT39 SmUGT40 SmUGT41 SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1 UKMIDSQPESSIVYVS FCSEGLSPEQMEEIANGLE 1 UKMIDSQPESSIVYVS FCSEGLSPEST 1 UKMIDSQPESSICALOFTER UKMIDSQPESSI 1 UKMIDSQPESSICALOFTER USANGLE 1 UKMIDSQPESSICALOFTER USANGLESSIN 1 UKKESVENAVEREV 2 UKMIDSQPESSICALOFTER USANGLESSIN 2 UKMIDSQPESSICALOFTER USANGLESSIN 2 UKMIDSQPESSICALOFTER USANGLESSIN 3 UKMID | SmUGT27 | |
| SmUGT29 SmUGT30 SmUGT31 SmUGT32 SmUGT33 SmUGT33 SmUGT34 SmUGT35 SmUGT36 SmUGT36 SmUGT37 SmUGT36 SmUGT37 SmUGT38 SmUGT39 SmUGT40 SmUGT41 SmUGT42 Motif Consensus 1 UKRAPCOLVLNFSVCFLTHCONNSTLESJSAG 2 UKRLDSQPESSVLYVSVSCSEGSLSVEQMEEIANGLE 1 UKRAPCOLVLNFSVCFLTHCONNSTLESJSAG 2 UKRLDSQPESSVLYVSVSCSEGSLSVEQMEEIANGLE 1 UKRAPCOLVLNFSVCFLTHCONNSTLESJSAG 2 UKRLDSQPESSVLYVSVSCSEGSLSVEQMEEIANGLE 3 UKRAPCOLVLNFSVCFLTHCONNSTLESJSAG 2 UKRLDSQPESSVLYVSVSCSEGSLSVEQMEEIANGLE 3 UKRAPCOLVLNFSVCFLTHCONNSTLESJSAG 2 UKRLDSQPESSVLYVSVSCSEGSLSVEQMEEIANGLE 3 UKRAPCOLVLNFSVCFLTHCONNSTLESJSAG 2 UKRLDSQPESSVLYVSVSCSEGSLSVEQMEEIANGLE 3 UKRAPCOLVLNFSVCVVSVSCSEGSLSVEQMEEIANGLE 3 UKRAPCOLVVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSVSV | SmUGT28 | |
| SmUGT30 SmUGT31 SmUGT32 SmUGT33 SmUGT33 SmUGT34 SmUGT35 SmUGT36 SmUGT36 SmUGT37 SmUGT37 SmUGT38 SmUGT37 SmUGT38 SmUGT39 SmUGT40 SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1 UGMAPCDRVLAFESVGFLTHCGWNSTLESJSAG CLKRLDSQPESSWYVSGCSEGLSPEQMEEIAWGLE 1 UGMAPCDRVLAFESVGFLTHCGWNSTLESJSAG CLKRLDSQPESSWYVSGCSEGLSPEQMEEIAWGLE 1 UGMAPCDRVLAFESVGFLTHCGWNSTLESJSAG CLKRLDSQPESSWYVSGCSEGLSPEQMEEIAWGLE 1 UGMAPCDRVLAFESUGLSPEQMEEIAWGLE 3 WVFTWVFRELEPKY 8 TLPEGFLERTSGRGL 9 PFYTCITLSULAFECALBOSKRDCDLVMADEIARAVRCVMDGDSEIRKKVKEM 6 SNVPFJWVVRSESK 7 OCGVVVVFYELEPKY 8 TLPEGFLERTSGRGL 9 PFYTCITLSULAFECALBOSKRDCDLVMADEIARAVRCVMDGDSEIRKKVKEM 1 TYCSFPLUEDFTWNTSLAFCASSOF 1 TYCSFPLUEDFTWVVFWLF2 8 TLPEGFLERTSGRGL 9 SDSVRTGULDFFCVSULPSALBOSKRDCDLVMADEIAFSJILKK 1 TYCSFPLUEDFTWVVFWLF2 1 TYCSFPLUEPFYV 1 TYCSFPLUEPFYV 1 TYCSFPLUEPFYV 1 TYCSFPLUEPFYV 1 TYCSFPLUEPFYV 2 SDSULTOFULPFFVVSULPSALFUNEFDAMAPCFFLAQRFKD 2 SDSVRTGULDFFCVSULPFVSULFSALFALSSSN 1 TYCSFPLUEPFWVFWFWLF2 3 SDSVRTGULDFFCVSULFFSVSTFVKLAQRFKD 2 SDSVRTGULDFFCVSULFSVSULFSALFST 3 SDSVRTGULDFFCVSULFSSKRAGLES 5 SDSVRTGULDFFCVSULFSVSULFSALFST 3 SDSVRTGULDFFCVSULFSVSULFSALFST 3 SDSVRTGULDFFCVSULFSVSULFSALFST 3 SDSVRTGULDFFCVSULFSVSULFSALFST 3 SDSVRTGULDFFCVSULFSVSULFSALFST 3 SDSVRTGULDFFCVSULFSVSULFSALFST 3 SDSVRTGULDFFCVSULFSSKRAGLEFSSCENLDQLFSALFST 3 SDSVRTGULDFFCVSULFSVSULFSALFST 3 SDSVRTGULFFCVSULFSVSULFSSKRAGLEFSSCENLDQLFSALFST 3 SDSVRTGULFFCVSULFSVSULFSSKRAGLEFSSCENLDQLFSALFST 3 SDSVRTGULFFCVSULFSVSULFSSKRAGLEFSSCENLDQLFSALFST 3 SDSVRTGULFFCVSULFSSKRAGLEFSSCENLDQLFSALFST 3 SDSVRTGULFFCVSULFSVSULFSSKRAGLEFSSCENLDQLFSALFST 3 SDSVRTGULFFCVSULFSVSULFSSKRAGLEFSSCENLDQLFSALFST 3 SDSVRTGULFFCVSULFSVSULFSSKRAGLEFSSCENLDQLFSALFST 3 SDSVRTGULFFCVSULFSVSULFSSKRAGLEFSSCENLDQLFSALFST 3 SDSVRTGULFFCVSULFSVSULFSSKRAGLEFSSCENLDQLFSALFST 3 SDSVRTGULFFCVSULFSVSULFSSKRAGLEFSSCENLDQLFSALFST 3 SDSVRTGULFSKSFRAGLEFSSCENLDQLFSKRAGLEFSCENLDQLFSALFST 3 SDSVRTGULFSKSFRAGLEFSSCENLDQLFSKRAGLEFSSCENLDQLFSALFSKF 3 SDSVRTGULFSKSFRAGLEFSSCENLDQLFSKRAG | SmUGT29 | |
| SmUGT31 SmUGT32 SmUGT33 SmUGT33 SmUGT34 SmUGT35 SmUGT36 SmUGT36 SmUGT37 SmUGT37 SmUGT38 SmUGT39 SmUGT40 SmUGT41 SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1 UGMAPCDRVLAFBSVCFLTHCGWNSTLESJSAG 2 UCKRLDSQPESSWYVSSCSEGSLSPEQMEEIAWGLE 1 UGMAPCDRVLAFBSVCFLTHCGWNSTLESJSAG 2 UCKRLDSQPESSWYVSSCSEGSLSPEQMEEIAWGLE 3 WVFTWAVELEPKY 4 UVSTPAPCORTNALLARENASKGFK VVMYTWFLVRSQPLNA 5 SmVPFJWVVRSEK 6 SSNVPFJWVVRSEK 7 UCGRVCUPFYCESPKY 8 TLPEGFLRFTSGRGL 9 PFVTCITUSULPFX 8 TLPEGFLRFTSGRGL 9 SSNVPFJWVVRSESK 1 UFGRSGRSSDLMJDAFT 1 SDSVRTVTUSULPFYV 8 TLPEGFLRFTSGRGL 9 SSNVPFJWVVRSSK 1 UFGRSGRSSDLMJDAFT 1 SDSVRTVSULPFYVSUESPKY 8 SSNVPFJWVVRSSK 1 UFGRSGRSL 9 SSNVPFJWVVRSL 1 UFGRSGRSL 9 SSNVPFJWVVRSL 1 UFGRSGRSL 9 SSNVPFJWVVRSL 1 UFGRSGRSL 9 SSNVFFJULPFYV 1 SSSL 1 UFGRSGRSL 9 SSNVPFJWVVRSL 1 UFGRSGRSL 9 SSNVPFJWVRSL 1 UFGRSSTPLCEPT 1 SSSL 1 UFGRSGRSL 1 UFGRSSTPLCEPT 1 SSSL 1 UFGRSSTPLCEPT 1 SSSL 1 UFGRSSTPLCEPT 1 SSSL 1 UFGRSSS 1 UFGRSSS | SmUGT30 | |
| SmUGT32 SmUGT33 SmUGT34 SmUGT35 SmUGT36 SmUGT36 SmUGT37 SmUGT38 SmUGT37 SmUGT38 SmUGT39 SmUGT40 SmUGT40 SmUGT41 SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1. UCGRAPOLEVLARESVCFLTHCGWNSTLESJSAG CLKKLDSQPESSVUYVSGCSEGSLSPEQMEELMAGLE 1. UCGRAPOLEVLARESVCFLTHCGWNSTLESJSAG CLKKLDSQPESSVUYVSGCSEGSLSPEQMEELMAGLE 3. UTGRAPOLEVLARESVCFLTHCGWNSTLESJSAG CLKKLDSQPESSVUYVSGCSEGSLSPEQMEELMAGLE 3. UTGRAPOLEVLARESVGCFLTHCGWNSTLESJSAG CLKKLDSQPESSVUYVSGCSEGSLSPEQMEELMAGLE 3. UTGRAPOLEVLARESVGCFLTHCGWNSTLESJSAG CLKKLDSQPESSVUYVSGCSEGSLSPEQMEELMAGLE 3. UTGRAPOLEVLARESUS 5. UTGRAPOLEVLARESVGCFLTHCGWNSTLESJSAG 6. SSNVPFJWVVRSSK 7. QGUVJWTPFLZEPKY 8. TLPSGFLGAVELKLDSSRDGDLWAADEJARAVRCVMDGDSEIRKKVKEM 6. SSNVPFJWVVRSSK 7. QGUVJWTPFLZEPKY 8. TLPSGFLGAVELKLDSSRDGDLWAADEJARAVRCVMDGDSEIRKKVKEM 5. SDSVRTVGULDFFCVSUBUDVANEGLGSVSTFVKLAQRFKD 9. PPYTCITUSULPFN 10. EJAKKAVAEGSSDLMJDAFT 11. GEKCKEMKKAKLKL 7. SDSSLTJCGVLDFFCVSVVVVSUFSALFWNGGYSTFVKLAQRFKD 11. GEKCKEMKKAKLKL 12. SFRQDLPFNYCTVUPFLCSVSUBJVDAFT 13. UTGRSTUPTIELSPTNDEJSTFVKLAQRFKD 14. KPVYTVGFLLP 15. UTGFVNVVPVSUFSALFNAGSTFVKLAQRFKD 15. UTGFVNVVPVSUFSALFNAGSTFVKLAQRFKD 16. DVASZLGIFVULVFFCVSSTRAAGL 7. SDSSLTAFGVVPFVVVEVSUFSALFNAGSTFVKLAQRFKD 17. SDSSLTAFGVVPFVVSUFSALFNAGSTFVYTT 20. MILFETIPDGLFPGKSADAL 3. WVLKKLEDKKNGLPFSGCSNLDQLFSSATAHKET 3. SDSVRVFUVFICEBSTVKLAPCFCPC 3. MVLDRLKEPSSGCAMSFFVYTT 3. SDSVRVFUVFICEBSTVKDVLFFASSTYTT 3. SDSVRVFUVFICEBSTVKLAPCFCPC 3. MVLKLEDKKNGLPFSGLMAK 4. WEVEFCPCVVVVVVSUFSALFNAGSTFVYLLQLPLTQFLQ 3. MVLKKEPSVEGFGCFAL 3. MVLKKEPSVEGFGCFAL 3. WVLKKEPSVEGFGCFAL 3. WVLKKEPSVEGFGCFAL | SmUGT31 | |
| SmUGT33 SmUGT34 SmUGT35 SmUGT35 SmUGT36 SmUGT37 SmUGT37 SmUGT38 SmUGT39 SmUGT40 SmUGT41 SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1. UCGRAPCIDEVLARESVCFLTHCGNNSTLESJSAG CLKKLDSQPESSVVVVSFCSEGSLSPEQMEEIANGLE HVVFIFYAGCHINPHLIZLARELASKGFK 4. VVMVTMPLVADQFESS SMVFFJWVVELSEK 5. FENVKELAVELKLDSKRDCDLVMADEIARAVRCVMDGDSEIRKKVKEM 6. SSNVPFJWVVESSEK 7. QCUVNTYFLEEPKY 8. TLPEGFLERTSGRGL 9. PPYTCITUSULPFN 10. EJAKKAVAEGGSSDLNJDAFT 11. GEKKGEMRKAKALKL 12. SEHQDLPPHYHCTNITPHLMNTLLAAFDASKPAFSDILKK 13. SDSVRTVTUSULPFCVSWIDVAAELGLPSVIFLTSNAGFLGLMLVLSTHH 14. KPYVTVGFLLP 15. DVASZLGIFVULPFVVVEVSUPSALASKISSN 16. DVASZLGIFVULWTASAAGL 7. SDSSLIFJCGVVPTVFVUESSK 17. SDSSLIFJCGVVFVVVVSVLPSALFNNGVSTFVKLAQRFKD 18. DVASZLGIFVULWTASAAGL 7. SDSSLIFJCGVVPTVFVUESSK 17. SDSSLIFJCGVVPTVVVDFFGKSSTRADLFSVIFTT 18. DVASZLGIFVULVPFCVSKSK 17. SDSSLIFJCGVVFVVVVVSVLPSALFNNGVSTFVKLAQRFKD 18. DVASZLGIFVULVPFCVSKSTVA 19. DVASZLGIFVULVPFCVSKSK 10. DVASZLGIFVULVPFCVSKSTVALFD 11. GEKCKEMPKSKRALKL 12. SFHQDLPPHYHCKLSFSK 13. SDSVRTVDIVPFFGKSSTRADLFSVIFTT 14. KPVYTVGFLLP 15. JVFCSTFLEPFQLJKLM 16. DVASZLGIFVULVPFCVSVFVVESLEPSYTFT 17. SDSSLIFJCLEPTRNHFDDHHNKLSSSN 16. DVASZLGIFVULVPFCVSVFVVESLEPSYTFT 26. MVLNRHLKEPSKEGLFSCNLDQLFSSATAHKST 27. SUSYLVHVRGLLPFLSGPNISVPCLPKP 28. MPLVFGLIFUEDEVFTSVFLFVSFLFVF 29. MEHDFOLFLEPSKFT 20. MEHDFOLFLEPSKFT 20. MHVKGLEPSKFT 21. SDSVFVFUEDEPSVFVFVE 22. MMLKEPSKEGLEPSSTRADLFSVTFT 23. MMLKKEFSKEGLEPSSTRADLFSVTFT 24. MEHDFOLFLEPSKFT 25. MVLFKELDOKYGTINLFFF 26. MVLVKFTGVFGTFAL | SmUGT32 | |
| SmUGT34 SmUGT35 SmUGT36 SmUGT37 SmUGT37 SmUGT38 SmUGT39 SmUGT40 SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1. UCGRAPCIDEVLARESVGFLTHCGWNSTLESJSAG CLKKLDSQPESSVVVVSFCSEGSLSPEQMEEIAKGLE HVVFIFYDQELLARELASKGFK 4. VPMVTMPLVADQTENA 5. FRWVSELGAVELALDSKRDGDLVMADEIARAVRCVMDGDSEIRKKVKEM 6. SSNVPFJWVVRSEKK 7. QCUJVMPTFLEEPKY 8. TLPEGFLERTSGRGL 9. PPYTCITUDULPWA 10. EJAKKAVAEGGSSDLMJDAFT 11. GEKKGEMRKKAKLK 12. SFHQDLPPHYHCTNNIPPHLMNTLLAAFDASKPAFSDILKK 13. SDSVRTVTULDFFCVSNUFVLEQAVELFSNAGFLGLMLVLSTHH 14. KPYVTVGFLLP 15. DVASZLGITVULPFWA 16. DVASZLGITVULPFWA 17. SDSSLTJCEVTENTPHLSPKV 18. DVASZLGITVULPFWA 19. DVASZLGITVULPFWA 10. EJAKKAVAEGGSSDLMJDAFT 11. GEKKGEMRKKAKLK 12. SFHQDLPPHYHCTNNIPPHLMNTLLAAFDASKPAFSDILKK 13. SDSVRTVDULPFCVSVLPSALFANDGYSTFVKLAQRFKD 14. KPVYTVGFLLP 15. DVASZLGITVULPFCVSVLPSALFANDGYSTFVKLAQRFKD 16. DVASZLGITVULPFCVSVRVPUEPAALFNKDGYSTFVKLAQRFKD 17. SDSSLTJCFVNVPVEVEPAALFNKDGYSTFVKLAQRFKD 18. DVASZLGITVULPFCVSVRVPUEPAALFNKDGYSTFVKLAQRFKD 19. SHLDTVUPFFCKSSTRADLFSVIRKTD 20. MILNEFTIPDGLPFGGSDLAF 21. SHLDTVUPFFCKSSTRADLFSVIRKTD 22. YIESLERTKGLAPEFQLJKLKM 23. MVLDRILEDDKYGTINLFRF 24. MENEFGELEVENTSDILFRSPTYTT 25. YIESLERTKGLAPEFQLJKKLM 26. SMVLDRILEDDKYGTGINLFKF 27. SITYHVHRGLIPLESCPNISVPVEPKPLKAYDLQLPLTQFLQ 28. MMLWGCHERM | SmUGT33 | |
| SmUGT35 SmUGT36 SmUGT37 SmUGT38 SmUGT39 SmUGT39 SmUGT39 SmUGT40 SmUGT41 SmUGT42 Motif Consensus 1 UVGRAPOLEVLAHESVGEFLTHCGWNSTLESJSAG 2 CLKMLDSQPESSVUYVSFCSEGSLSPEQMEE LANGLE 3 HVVFIFYAGCHJNPHLLAKRLSKGFK 4 VPMVTMPLYADOPLNA 5 FFMVKBLGLAVEELKDSKRDGDLVMADE LARAVRCVMDGDSEIRKKVKEM 6 SSNVPFJWVVESESK 7 QGUJVNFYELEPKY 8 TLPEGFLERTSGRGL 9 PPYTCITUVLIPRA 10 EJAKKAVAEGGSBLNJDAFT 11 GEKCKBMKKAKELK 12 SFHQDLPPHYKCTKNIFPHLMNTLLAAFDASKPAFSDILKK 13 SDSVRTGULUPFCVSUPVEVUEPALAPKNDGYSTFVKLAQRFKD 14 KPYVTVGPLLP 15 UVRESUFGVNEVPVEVUEPALSENLOGLESVIFLTSNAGFLGLMLVLSTHH 16 DVASZLGIFVVLFWTASAGL 17 SDSSLTGVULPFYCVSUPSLEALSSSN 18 DULATEVVLEPVLVDEVEVLEALAFSSEYTT 21 GLALIXTHKLEPSKEGGLFSSNILDQLERDSATAHKET 22 YESLERTK | SmUGT34 | |
| SmUGT36 SmUGT37 SmUGT38 SmUGT39 SmUGT39 SmUGT40 SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1 UKWAPOLEVLANESVGETLHCGWNSTLESJSAG 2 CLKWLDSQPESSVVYVSGSERGSLSPEGMEELMKGLE 3 HVVFIFYAGGHJNPHLLAKRLSAKGFK 4 VPMVTMPLVAEOFLNA 5 FFMVKBLGLAVEELKDSKROGDLVMADE LARAVRCVMDGDSEIRKKVKEM 6 SSNVPFJWVVESEKK 7 QGUJVNFYELEPKY 8 ILPEGFLERTSGRGL 9 PPYTCITJVLLPRA 10 EJAKKAVAEGGSSDLMJDAFT 11 GEKGKEMKKAKELK 12 SFHQDLPPHYKCTKNITPHLMNTLLAAFDASKPAFSDILKK 13 SDSVRTGULUPFCVSUPVEVVEVSUEPAALFUNDGYSTFVKLAQRFKD 14 KPYVTYGPLLP 15 UYRCSPULEPTVNEVVEVVEVESALLDQLPSADLLAKFFKALSMLQQPFEQ 16 DVASZLGIPVULPFVSUEPASTRADLPSVIRKTD 17 SUSSILTADLPFOGSZDAF 18 EHLDTVVDFIFGMSSIRADLPSVIRKTD 19 EHLDTVVDFIFGMSSIRADLPSVIRKTD 20 RNMLPRELEDDKDYGINLFER | SmUGT35 | |
| SmUGT37 SmUGT38 SmUGT39 SmUGT40 SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1 | SmUGT36 | |
| SmUGT38 SmUGT39 SmUGT40 SmUGT41 SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1. 1. UKWAPOLEVLAHESVGEFLTHCGWNSTLESJSAG 2. CLKWLDSQPESSVVYVSGSEGSLSPEQMEELAWGLE 3. HVVFIFYRQCHJRPHLLAKRLSKEFK 4. VPMVTMPLYABOTINA 5. FRMVKELGLAVELKLDSKRDCDLVMADE LARAVRCVMDGDSEIRKKVKEM 6. SSNVPFJWVVESERK 7. QGUJVNFYELEPRY 8. ILPEGFLERTSGRGL 9. PPYTCITULULPR 10. EJAKKAVABGGSSDLNJDAFT 11. GEKGEMRKKAKELK 12. SFHQDLPPHYKCTKNIPPHLMNTLLAFDASKPAFSDILKK 13. SDSSUTVULDFYCVSMIDVANELGLESVIFLTSNAGFLGLMLYLSTHH 14. KPYTYGFLLP 15. UYRGSFLGUPVIPEVSKIDADAFT 16. DVASZLGIPVULPYTSVESHDANGYSTFVKLAQRFKD 17. SDSSLIFSCHAPTENPENDESSTLADLESSIN 18. DVASZLGIPVULPYTVSVESHENDATEN 19. BHLDTVVDIFLEPKSCENAA 20. RUKEPEDAVKNULVULPULULENDSATAHKET | SmUGT37 | |
| SmUGT39 SmUGT40 SmUGT41 SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1. UKWAPQLEVLARPSVGFLTHCGWNSTLESJSAG 2. CLKWLDSQPESSVVYVSGSEGSLSPEQMEELAWGLE 3. HVVFIFYAQGHJMPHLLARRIJASKGFK 4. VPWYTMPLYABOTINA 5. FRMVKELGLAVELRLDSKRDCDLVMADEIARAVRCVMDGDSEIRKKVKEM 6. SSNVPFJWVVRSESK 7. QGUJVNFYELEPRY 8. ILPEGFLERTSGRGL 9. PPYTCITULLPRN 10. EJAKKAVABGGSSDLNJDAFT 11. GEKGKEMKKAKELK 12. SFHQDLPHYHCTNN IPHLMNTLLAAFDASKPAFSDILKK 13. SDSVRTGULDFTCVSMIDVANELGLESVIFLTSNAGFLGLMLYLSTHH 14. KPYTTGULUPTCVSMIDVANELGLESVIFLTSNAGFLGLMLYLSTHH 15. JVRCSTFLGLPTVINFDDHHNKLSSSN 16. DVASZLGIPVULPTVSVEJESALFNDGYSTFVKLAQRFKD 17. SDSSLIFPENDERVERNE 18. EHLDTVVDIFLEPKSCENADE 19. BHLDTVVDIFLEPKSCENADE 20. RUKEFDEADVKINHULDGLERDSATAHKET 21. TIESLRTKGLAPERQULINKLN 22. | SmUGT38 | |
| SmUGT40 SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1. UKWAPCLEVLANESVGSFLTHCGWNSTLESJSAG 2. CLKWLDSQPESSVVYVSGSEGSLSPEQMEELAWGLE HVVFIFYDQGHJNPHLLARRIJSKGFK 4. VPWVTMPLYADQTINA 5. FPMVKBLGAVEELADSKROCDLVMADE JARAVRCVMDGDSEIRKKVKEM 6. SSNVPFJWVVRSSK 7. QGUJVNFYFELEPKY 8. ILPEGGLERTSGRGL 9. PPYTCITUJULJWA 10. EJAKKAVAEGGSSDLNJDAFT 11. GEKGKEMKKAKELK 12. SFRQDLPHYHCTNNIPPHLMNTLLAAFDASKPAFSDILKK 13. SDSVRTGULDFTCNNIPPHLMNTLLAAFDASKPAFSDILKK 14. KPYYTGULDFULVEYVENLONGYSTFVKLAQRFKD 15. DVASZIGIPVULWTYVSUPSALFNKDGYSTFVKLAQRFKD 16. DVASZIGIPVULWTYVSUPSALFNKDGYSTFVKLAQRFKD 17. SUBSLITPECLEPGCKSDAE 18. VKRESVERANENNE 29. GHILKVHKLKFPSKEAGLPESCENLDQLESSTFVKLAQRFKD 21. GHILKVHKLKEPSKEAGLPESCENLDQLESSTFVTT 22. TIFTTFVLIFPECLEPGOLJKKLM 23. OSGIQLIELPZVUPPEVLELKESSESTYTT 24. | SmUGT39 | |
| SmUGT41 SmUGT42 Motif Symbol Motif Consensus 1. UVGNAPQLEVLANESVGGFLTHCGWNSTLESJSAG 2. CLKNLDSQPESSVVYVSGSEGSLSPEQMEEIAWGLE HVVFIFYDQGHJNPHLLARKLASKGFK VPWVTMPLYADQTINA 5. FRWYKBLGLAVELRLDSKRDGDLVMADEIARAVRCVMDGDSEIRKKVKEM 6. SSMVPFJWVVRSSK 7. QGUVJWFYFLEPKY 8. ILPEGGLERTSGRGL 9. PPYTCITIVLUEWA 10. EJAKKAVAEGGSSDLNJDAFI 11. GEKGEKMKRAKELK 12. SFRQDLPPHYHCTNNIPPHLWNTLLAAFDASKPAFSDILKK 13. DSSRVTGULDFTCVSMIDVANELGLESITIFLTSNAGGFLGLMLYLSTHH 14. KPVYTUGULDEFTCVSMIDVANELGLESITEVKLAQRFKD 15. DVASZIGIPVULWTYSVLSSI 16. DVASZIGIPVULWTYSVLSSI 17 SDSSLITPDGLPEGGKSDAE 18. WHLDTWOPIPONSI 19. FILALESSN 10. GLHKVHKLKKPSEAGLPEGCENLDQLPSADLLAKFFKALSMLQQPFQ 11. GLHKVHKLKKPSEAGLPEGCENLDQLERSATALSKET 22. TIFTTFVHIRLESKENKADPEGULJKKIM 23. OSGIQLIELPZVDPEPLLLKESFSTYTT 24. EMEEFDEALVKULMUUDLIDLERDSATAHKET </td <td>SmUGT40</td> <td></td> | SmUGT40 | |
| SmUG142 Motif Symbol Motif Consensus 1. UVGMAPQLEVLAHPSVGGFLTHCGWNSTLESJSAG 2. CLKWLDSQPESSVVVVSGSEGSLSPEQMEEIAWGLE 3. HVVF1FYRAQGHJNPHLELAKRLSASGFK 4. VPWVTWFLYAQGHJNPHLELAKRLSASGFK 5. FRWYKELGLAVELELDBSRDGDLVMADEIARAVRCVMDGDSEIRKKVKEM 6. SSNVPFJWVVRSSEK 7. QGVJVMFYELEPRY 8. ILPEGELERTSGRGL 9. PPYTCITIVULIPM 10. EJAKKAVAEGSSULMJDAFI 11. GEKGKEMRKAKEL 12. SENGDLPFHYHCTNNIPFHLWNTLLAAFDASKPAFSDILKK 13. DSSWTGVLDUPCYCWSHIDVANELGLESTIFLTSNAGFLGLMLYLSTHH 14. KPYYTGULDFTVLSSAGL 15. JYCSFTUFULFTMISSAGL 16. DJASJGJEFFWPFYWJLPSALFNKDGYSTFVKLAQRFKD 17. GLHLKVHKLKFPSKEAGLPEGCENLDQLESADLLAKFFKALSMLQQFFEQ 18. HEDFTWDFIPKWSLESSNLDUFSSAGL 20. PNIAFETIPEDCLPEGCKSDAE 21. GLHLKVHKLKFPSKEAGLPEGCENLDQLESADLLAKFFKALSMLQQFFEQ 22. TTFTTFPUGLPEGULFWENST 23. OSGIQLIELPZVDPFPLELLKESFEYYTT 24. ENEEFDEAVKLWLUUDLENDANAMEKT 25. YIESLEPKOLJKKLM 26. SMYLDRKLBDOKYGINLFKP 27. SITYHVHKGLEPLESOPNISVPCLERST 28. | SmUGT41 | |
| 1. UVGN POLICULARPSVGGFLTHCGWNSTLESJSAG 2. CLKWLDSQPESSUVYVSGSEGSLSPEQMEEIAWGLE 3. HVVFIFYAQGHJWHLLARRLSSGFK 4. VPWTMFLYADQTINA 5. FRMVKSLGLAVELRLDSKROGDLVMADEIARAVRCVMDGDSEIRKKVKEM 6. SSMVPFJWVVRSSEGSGSLSPEQMEEIAWGLE 7. QGUVINTFILEPRY 8. ILPEGELRFISGRGL 9. PPTVCITIDULIVA 10. EGEKGEMMKRAKELK 11. GEKGEMMKRAKELK 12. SENGDLPPHHCTNNIPPHLANTLLAAFDASKPAFSDILKK 13. DSVRTOFLUDIDFCVSMIDVANELGGLSYTIFLTSNAGFLGLMLYLSTHH 14. SDSVRTOFLUDIDFCVSMIDVANELGGLSYTFVKLAQRFKD 15. DVASSLGIVULPPTPASAGL 16. DVASSLGIVULPPTPASAGL 17. SDSSLLTCFVNPVYSVLFSALFNKDGYSTFVKLAQRFKD 18. VKREBVERVESVME 19. BHLDTVVDFIROSSTRLADLFSVIRKTD 20. PNIAFFTIPDGLPFGCKSDAE 21. GGIGLIELPZVVLPVPSVLFSALFNAKET 22. TIFTTFVINIKLKKFSEGNAA 23. OSGIGLIELPZVVLVULPULRDSATAKET 24. EMEEFDEAVVKLWLUUDLELRDSATAKET 25. YIESLRFKGLAPFEQULJKKLM 26. </td <td>SmUGT42</td> <td>Motif Conconsue</td> | SmUGT42 | Motif Conconsue |
| 2. CLINDLSQFESS VITYSGSESSISFEQMEELAWGLE 1. UVVFIFYDQGHINAULARKLASGFK 4. VPWTMFLTADQFINA 5. FRWYKELGAVEELABSRDGGDLVMADEIARAVRCVMDGDSEIRKKVKEM 6. SSMVPFJWVVESSK 7. QGVJWFTFELEFKY 8. LIPSGELERTSGRGL 9. PFTCITIDULINM 10. EJRKAVARGSSIDNJDAFT 10. EJRKAVARGSSIDNJDAFT 11. GEKGGEMEKKAKELK 12. SSSVPTGJVLDFVSWIDVANELGLFSYIFLTSNAGFLGLMLYLSTHH 14. KPYTYGVLDFVSWIDVANELGLFSYIFLTSNAGFLGLMLYLSTHH 15. LIPGSTPUCHERSAAGL 17. SDSSLLIPGFNPVPVSVLFSALFNKDGYSTFVKLAQRFKD 18. VKRESVERVERVEN 19. BHLDTVVDFIFGSADEL 10. JUNASLGFVSUTASAAGL 17. SDSSLLIPGFNPVPVSVLFSALFNKDGYSTFVKLAQRFKD 18. VKRESVERVERVEN 19. BHLDTVVDFIFGSADEL 21. GLAIKVKHKKFPSKEAGLPSGCENLDQLFSADLLAKFFKALSMLQQFFEQ 22. ITFITTPIDCLFPGGKSDRE 23. OSGIQLIELPZVVDFVTALSASFYYTT 24. ENEEFDGAVKVKUHUQLERDSATAMKET 25. YILSLRTKGLAPFEQLJKKLM 26. SWYLDKRLBDKDYGINLFKP 27. SJTYHVHRGLIPLESGNISVPGLPKP 28. MPHVGGLIELWEXDFDVKVYLKTAYDLLQLPLTQFLQ 29. EWGGWEIERH | 1 | IVGWAPQLEVLAHPSVGGFLTHCGWNSTLESJSAG |
| 1 FRMV XELGIAUSIANCE 5 FRMV XELGIAUSIANCE 6 SSMVEFXWVRESSEK 7 SSMVEFXWVRESSEK 8 IPEGETSSERI 9 IPEVTCIIYDIVJENE 9 IPEVTCIIYDIVJENE 9 IPEVTCIIYDIVJENE 9 SKAN FENDANCOSS DUNDAFI 10 GEKCKEMKKAKELK 11 GEKCKEMKKAKELK 12 SPROLPPHYNCTKN IPENLMNTLLAAFDASKPAFSDILKK 13 SDSVRVTGUVLDFECVSMIDVANELGLPSYIFLTSNAGFLGLMLYLSTHH 14 KEVYTVQELLP 15 IYFCSTPICLEPINNNFDOHHKLSSSN 16 DVASZLGIEVVLPWTASAAGL 17 SDSSLLIFGEVNPVVSVLPSALFNKDGYSTFVKLAQRFKD 18 VKREEVERAVREVME 19 BHLDTVVDFIFGSSIRLADLPSVIKKTD 20 FNIAFETIPOLIPEGCKSDAE 21 GLHIKVHKLKPPSKEAGLPEGCENLDQLPSADLLAKFFKALSMLQQPFEQ 22 ITFITTPULSESKENA 23 OSGIQLIELPZVVPEVLOKLOSKADA 24 ENEEFDEAVVKLWALUQUELGNSATANKET 25 YIESLENDKOYGINLFKP 26 SMYLDKRLEDDKOYGINLFKP 27 SIYYHVHRGLEDLEVEATSDEVDKVPLKTAYDLLQLPLTQFLQ 28 MEWVGECHEVERTSNE 29 EWGVGEVERGIRAL | 3. | CLEWIDSQFESSVVIVSFGSEGSISFEQMEEIAWGLE HVVFIPYPAQGHJNPMLZLAKRLASKGFK UDMUTMDIVAEOFINA |
| 7. OGVIVNPYELEPRY 8. ILPEGFLERTSGRGL 9. PPYTCIIYDU/JPAN 10. EJARKANARGGSSDUNDAFI 11. GEKCKEMKKAKELK 12. SFROLPPHYECTKNIPPHLMNTLLAAFDASKPAFSDILKK 13. SDSVRVTGLVLDFFCVSMIDVANELGLPSYIFLTSNAGFLGLMLYLSTHH 14. KPVTYGVPLIP 15. IYFCSTPICLEPINNNFDOHHKLDSSN 16. DVASZLGIEVVLFWTASAAGL 17. SDSELLIGEFUNPVYVSULPSALFNKDGYSTFVKLAQRFKD 18. VKREEVERVNEW 19. BHLDTVVDFIPGNSSIRLADLPSVIKKTD 20. FNIAFETIPGLIPRGCKSDAE 21. GLHIKVHKLKFPSKEAGLPEGCENLDQLPSADLLAKFFKALSMLQQPFEQ 22. ITFITTPHOLPRGCKSDAE 23. OSGIQLIELPZVVPPPLELLKESFSYYIT 24. ENREFDEGVVLFWTASLERDANAKET 25. YIESLRTKGLAPFEQLJKKLM 26. SMYLDKRLEDDKDYGINLFKP 27. SIYYHVHRGLEPLFLSGPNISVPCLPKP 28. MPHVGCLPLEVEATSDVDKVPLKTAYDLLQLPLTQFLQ 29. EWGCGMEIREN 30. KYVEEVGFGGTRAL | 5. | FRWVKELGLAVELRLDSKRDGDLVMADEIARAVRCVMDGDSEIRKKVKEM SSNVDFIWVUPFSFK |
| 9. PPVTCITYDLVLPWA 10. EJAKKAVARGGSSDLNJDAFI 11. GEKGEMKKKAKELK 12. SFHQDLPPHYLCTKNIFPHLMNTLLAAFDASKPAFSDILKK 13. SDSVRVTGLVLDFFCVSMIDVANELGLPSYIFLTSNAGFLGLMLYLSTHH 14. KPVYTVGVLD 15. IYFCSTPICLEPINNNFPDHHNKLSSSN 16. DVASZLGIPVLPWASAAGL 17. SDSBLLIPGFVNPVVSUPSALFNKDGYSTFVKLAQRFKD 18. VKREEVERAVREVME 19. EHLDTVVDFIPGSSIRLADLFSVIKKTD 20. PNIAFETIPDGLPFGGKSDAB 21. GLHIKVHKLKFPSKEAGLPEGCENLDQLPSADLLAKFFKALSMLQQPFEQ 22. ITFITTPULRIERSFSVITT 23. QSGIQLIELPZVVPPPLELKRSFSYYIT 24. ENREFDEAVVKLWHLUDLRDSATAHKET 25. YIESLRTKGLAPFEQLJKKLM 26. SMYLDKRLBDDDKYGINLFKP 27. SIYYHVHRGLLFJPLSGPNISVPCLPKP 28. MPHVGCLPLUEVEATSDEVDKVPLKTAYDLLQLPLTQFLQ 29. EWGGWEIERH 30. KYVEEVGFGTRAL | 7. 8. | QGVJVNTFYELEPKY LLPEGFLERTSGRGL |
| 11. GEKGKEMRKKAKELK 12. SFHQDEPHYRCTKNIPPHLMINTLLAAFDASKPAFSDILKK 13. SDSVRVTGLVLDFFCVSMIDVANELGLPSYIFLTSNAGFLGLMLYLSTHH 14. KPVITVGPLIP 15. IYFCSTPICLEPIRNNHFDDHNKLSSSN 16. DVASZLGFLVULFWTASAGL 17. SDSELLIFGFVNPVVSVSLFSALFNKDGYSTFVKLAQRFKD 18. VKREEVEKAVREVME 19. EHLDTVVDFIFGSSIRLADLPSVIKTD 20. FNIAFETIPDGLPFGGKSADE 21. GLHIKVHKLKFPSKEAGLPEGCENLDQLPSADLLAKFFKALSMLQQPFEQ 22. ITFITTPVIRLEVFSKEAGLPEGCENLDQLPSADLLAKFFKALSMLQQPFEQ 23. QSGIQLIELPZVDPPLELLKESFEYYIT 24. ENEEFDEAVVKLVHLVDLERDSATAHKET 25. YIESLRTKGLAPFEQLJKKLM 26. SMYLDRKLBDDKVGINLFKP 27. SIYYHVHRGLLPLPLSGPNISVPCDKPL 28. MPHVGCLPLEVEATSDEVDKVPLKTAYDLLQLPLTQFLQ 29. EWGGMEIREN 30. KYVEEVGFGGRAL | 9. 10. | PPVTCIIYDLVLPWA EJAKKAVAEGGSSDLNJDAFI |
| 3. SDSVRVTGLVLDFFCVSMIDVANELGLPSYIFLTSNAGFLGLMLYLSTHH 14. KPVTTVGPLP 15. IYFCSTPICLEFIRNNHFDDHHNKLSSSN 16. DVASZLGFUVLFWTASAGL 17. SDSELLFOFVNPVVSULPSALFNKDGYSTFVKLAQFFKD 18. VKRESUEKAVREVME 19. EHLDTVVDFIFGKSIRLADLPSVIRKTD 20. PNIAFETIPDGLPFGCKSNAE 21. GLHIKVHKLKFPSKEAGLPEGCENLDQLPSADLLAKFFKALSMLQQPFEQ 22. ITFITTPVI NIRLFKSFONA 23. QSGIQLIELPZVDPPLELLKESFSVIIT 24. ENREFDEAVVKLWHLUQLERDSATAHKET 25. YIESLRTKGLAPFEQLJKKLM 26. SMYLDRRLEDDOKYGINLFKP 27. SIYYHVHRGLLPLPLSGPNISVPCDFKP 28. MPHVGCLPLEVEATSDEVDKVPLKTAYDLLQLPLTQFLQ 29. EWGVGMEIREN 30. KYVEEVWGTGRTAL | 11. | GEKGKEMRKKAKELK SFHQDLPPHYHCTKNIPPHLMNTLLAAFDASKPAFSDILKK |
| IIICSTFILLOFTALMONHUUMHNKLSSSN DVASZLGTVVLFYALARNHUUMHNKLSSSN DVASZLGTVVLFYALARNHUUMHNKLSSSN DVASZLGTVVLFYALARNHUUMHNKLSSSN SUSSLLTPGVNPVPVSVLFSALFNNDGYSTFVKLAQRFKD VKRESVEKAVREVME HILDTVVDFIFGMSSIRLADLPSVIRKTD FILLTTPIDCLPFGCKSNAB GLHIKVHKLKFPSKEAGLPEGCENLDQLPSADLLAKFFKALSMLQQPFEQ TFTITPYN IRRLFKSFONAA GSGIGLIELPZVDPPLELLKESFSYITT SIYUNHKGLPPLSGENISVEGPKF SIYUNHKELDOKYGINLFKP SIYUHVHGLLPLLSGENISVEGLPKF SIYUNHKELDOKVGINLFKP SIYUHVHGLLPLSGENISVEGLPKF SIYUNHKGLPLEVENTSDEVOKVPLKTAYDLLQLPLTQFLQ EWGYGMEIREN KYVEEVWGTGIRAL | 13. 14. | SUSVKVTGLVLDFFCVSMIDVANELGLPSYIFLTSNAGFLGLMLYLSTHH REVYTVGPLLP |
| SUSSLIL/FOFWE/FY/SUFSALENADIOI/FYKLAQKFKD 18. VKREEVEKAVREVME 19. EHLDTVVDFIFGKSSIRLADLPSVIRKTD 20. PNIAFETIFDGLPFGKSKDAE 21. GLHIKVHKLKFPSKEAGLPEGCENLDQLPSADLLAKFFKALSMLQQPFEQ 22. ITFITTPVI NIRLPKSPGNAA 23. QSGIQLIELPZVDFPLELLKESFSYIIT 24. ENREFDEAVVKLVHLVQLELROSATAHKET 25. YIESLRTKGLAPFPQLJKKLN 26. SMYLDKRLEDDKVGGINFKP 27. SIYYHVHRGLLPLPLSGENISVEGLPKP 28. MPHVGCLPLEVEATSDEVOKVPLKTAYDLLQLPLTQFLQ 29. EWGVGMEIBRH 30. KYVEEVWGFGIRAL | 16. | IIICSIFICEEFIKNNHFDDHHNKLSSSN DVASZLGIPVVLFWTASAAGL SNGFLIIDGENHDVLVLU DOALENDOVGTEURIAODEKD |
| 20. PNIAFTIPPOLIPSGKSDAE 21. GLHIXVHKLKPPSGKSDAE 22. ITFITTPVIKLKPSGKADAE 23. QSGQLIELPZVDPPLELLKESFGNAA 24. ENREFDEAVVKLVHLVQLEROSATAMAKET 25. YIESLRTKGLAPFSQLJKKLN 26. SMYLDRKLEDORVGLNFKP 27. SIYYHVHRGLPLPLSGPNISVPGLPKF 28. MPHVGCLPLPLSGPNISVPGLPKP 29. EWGVGMEIBRH 30. KYVEEVWGTGIRAL | 18. | VKREEVEKAVREVME RHLDTVUDFIDGMSSIRLADLPSVIRKTD |
| 22. TTFTTTPYNIKRLEKSPGNAA 23. OSGIQLELPZVOPPELLIKESFEYYIT 24. ENEEFDEAVVKLVHLVQELRDSATAHKET 25. YIESLKTKGLAPFEQLJKKLM 26. SMYLDRKLEDDKDYGINLFKP 27. SIYTHVHRGLEPLEDGKNGINLFKP 28. MPHVGELPLEVGATSDEPDKVPKVPKLKTAYDLLQLPLTQFLQ 29. EWGVGMEIERH 30. KYVEEVWGTGIRAL | 20. 21. | PNIAFETIPDCLPPGCKSDAE GLHIKVHKLKFPSKEAGLPEGCENLDQLPSADLLAKFFKALSMLOOPFEO |
| 24. ENERFDEAVVKLVHLVQELROSATAHKET 25. YIESLERKCLAPEQULYKLINKLM 26. SMYLDKRLEDDKDYGINLFKP 27. SITYHVHROLPFLSGENISVPGLPKP 28. MPHVEGLPLEVEATSDEVDKVPYLKTAYDLLQLPLTQFLQ 29. EWGVGMEIERH 30. KYVEEVWGTGIRAL | 22. 23. | ITFITTPYNIKRLPKSPGNAA QSGIQLIELPZVDPPPLELLKESFEYYIT |
| 20. SMYLDKRLEDDKDYGINLFKP 27. SIYYHWIRGLEDEDLSGPNISVPGLPKP 28. MPHVEGLPLEVEATSDLPVDKVPYLKTAYDLLQLPLTQFLQ 29. EWGVGMEIERH 30. KYVEEVWGTGIRAL | 24. 25. | ENEEFDEAVVKLVHLVQELRDSATAHKET YIESLRTKGLAPFEQLJKKLN |
| 20. EWGVGMEIERH 30. KYVEEVWGTGIRAL | 20. 27. | SMYLDRKLEDDRDYGINLFKP SIYYHVHRGLDPLSGPNISVPGLPKP MBUUCCI DI EUGATE DI DUDUUNY MDAVDI I OLDI TOTOLO |
| | 29. 30. | EWGVGMEIERH KVVEEVWGTGIRAL |

Fig. 2 Conserved motifs of SmUGT proteins. The SmUGT sequences were visually analyzed; the top 30 motifs were selected and designated as motif 1 to motif 30

15

12



Fig. 3 *cis*-Acting elements in a 3000 bp region upstream of *SmUGT* genes. Numbers represent the numbers of *cis*-acting elements; colors from white to red represent numbers from less to more

closely related to *Medicago truncatula* than to *Arabidopsis thaliana*.

Structural features of SmUGT proteins

The secondary structures of the SmUGTs consisted of an α -helix, extension strand, β -turn, and random coil (Table 2). SmUGTs harbored a large proportion of α -helices (41.70% average), followed by random coils (36.78%) and extended strands and β -turns (15.18% and 6.34%, respectively).

Images of the predicted 3D structural models were colored according to the $N \rightarrow C$ terminus. Several SmUGTs with the same predicted proteins were in the same subfamily (Fig. 6 and Table S1).

SmUGT gene expression in tissues and at different developmental stages

The expression patterns of *SmUGTs* were explored based on transcriptome data obtained from fruit at eight developmental stages and from roots, leaves, and flowers; the results are shown as a heat map in Fig. 7.

According to the expression patterns during different stages of fruit development (Fig. 7a), *SmUGTs* were divided into two categories: those with expression levels higher from S1 to S4 than from S5 to S8 and those whose levels were higher from S5 to S8 than from S1 to S4. Most *SmUGTs* (93%) had low expression levels in S8, but those of *SmUGT22, SmUGT31,* and *SmUGT32* were significantly higher in S8. Few genes with high expression in S5 were found. Most genes were expressed at high levels in the reproductive organs, especially in female flowers and fruits early in development (Fig. 7b).



Fig. 4 Phylogenetic analysis of SmUGTs. Sapindus mukorossi (42, red stars), Barbarea vulgaris (10, green circles), Arabidopsis thaliana (9, brown circles), Panax ginseng (7, yellow circles), Medicago truncatula (5, cyan circles), Glycine max (4, gray circles), Panax quinquefolius (2, orange circles), Centella asiatica (2, purple circles), Saponaria vaccaria (1, blue circle), and Siraitia grosvenorii (1, black circle)



Fig. 5 Synteny analysis of the *SmUGT* genes of soapberry and two other plant species. Gray lines in the background and red lines represent the collinear blocks and syntenic *UGT* gene pairs in the genomes of soapberry and other plant species, respectively

| Protoin name | Sequence length | Amino acid proportion(%) | | | | | | | | | | | |
|--------------|-----------------|--------------------------|---------------------|---------------|------------------|--|--|--|--|--|--|--|--|
| 1 Totem name | (aa) | Alpha helix(Hh) | Extended strand(Ee) | Beta turn(Tt) | Random coil (Cc) | | | | | | | | |
| SmUGT1 | 486 | 43.83 | 14.81 | 6.79 | 34.57 | | | | | | | | |
| SmUGT2 | 545 | 41.47 | 14.5 | 7.34 | 36.7 | | | | | | | | |
| SmUGT3 | 437 | 40.73 | 14.42 | 6.86 | 37.99 | | | | | | | | |
| SmUGT4 | 532 | 38.35 | 14.1 | 7.33 | 40.23 | | | | | | | | |
| SmUGT5 | 461 | 41.21 | 14.1 | 7.59 | 37.09 | | | | | | | | |
| SmUGT6 | 461 | 42.08 | 14.32 | 6.94 | 36.66 | | | | | | | | |
| SmUGT7 | 453 | 40.18 | 13.25 | 6.62 | 39.96 | | | | | | | | |
| SmUGT8 | 453 | 41.06 | 13.47 | 5.96 | 39.51 | | | | | | | | |
| SmUGT9 | 459 | 42.92 | 15.03 | 4.58 | 37.47 | | | | | | | | |
| SmUGT10 | 459 | 40.96 | 16.56 | 6.1 | 36.38 | | | | | | | | |
| SmUGT11 | 484 | 41.32 | 16.74 | 7.23 | 34.71 | | | | | | | | |
| SmUGT12 | 477 | 44.03 | 15.3 | 5.24 | 35.43 | | | | | | | | |
| SmUGT13 | 477 | 45.28 | 14.88 | 6.08 | 33.75 | | | | | | | | |
| SmUGT14 | 149 | 53.02 | 15.44 | 8.72 | 22.82 | | | | | | | | |
| SmUGT15 | 473 | 42.71 | 15.64 | 5.07 | 36.58 | | | | | | | | |
| SmUGT16 | 394 | 43.65 | 14.47 | 6.09 | 35.79 | | | | | | | | |
| SmUGT17 | 431 | 43.16 | 14.39 | 5.1 | 37.35 | | | | | | | | |
| SmUGT18 | 466 | 44.42 | 15.45 | 5.79 | 34.33 | | | | | | | | |
| SmUGT19 | 469 | 43.5 | 15.35 | 5.54 | 35.61 | | | | | | | | |
| SmUGT20 | 472 | 43.01 | 15.47 | 6.36 | 35.17 | | | | | | | | |
| SmUGT21 | 470 | 43.62 | 14.26 | 4.89 | 37.22 | | | | | | | | |
| SmUGT22 | 456 | 42.11 | 15.57 | 4.61 | 37.72 | | | | | | | | |
| SmUGT23 | 471 | 36.94 | 14.23 | 7.01 | 41.83 | | | | | | | | |
| SmUGT24 | 455 | 40.66 | 15.82 | 5.93 | 37.58 | | | | | | | | |
| SmUGT25 | 457 | 48.36 | 14.66 | 5.03 | 31.95 | | | | | | | | |
| SmUGT26 | 449 | 43.43 | 16.48 | 5.12 | 34.97 | | | | | | | | |
| SmUGT27 | 458 | 44.98 | 15.72 | 6.33 | 32.97 | | | | | | | | |
| SmUGT28 | 496 | 41.33 | 17.34 | 7.86 | 33.47 | | | | | | | | |
| SmUGT29 | 489 | 43.76 | 14.93 | 7.16 | 34.15 | | | | | | | | |
| SmUGT30 | 481 | 43.66 | 14.76 | 5.82 | 35.76 | | | | | | | | |
| SmUGT31 | 457 | 39.17 | 16.41 | 8.32 | 36.11 | | | | | | | | |
| SmUGT32 | 459 | 38.56 | 16.56 | 7.41 | 37.47 | | | | | | | | |
| SmUGT33 | 470 | 39.36 | 15.32 | 6.17 | 39.15 | | | | | | | | |
| SmUGT34 | 460 | 39.57 | 15.43 | 5.65 | 39.35 | | | | | | | | |
| SmUGT35 | 538 | 32.53 | 13.94 | 6.88 | 46.65 | | | | | | | | |
| SmUGT36 | 490 | 42.45 | 14.49 | 5.31 | 37.76 | | | | | | | | |
| SmUGT37 | 450 | 42.44 | 17.11 | 4.67 | 35.78 | | | | | | | | |
| SmUGT38 | 463 | 36.5 | 15.33 | 6.7 | 41.47 | | | | | | | | |
| SmUGT39 | 468 | 35.47 | 15.6 | 6.41 | 42.52 | | | | | | | | |
| SmUGT40 | 452 | 42.48 | 15.49 | 6.86 | 35.18 | | | | | | | | |
| SmUGT41 | 477 | 42.35 | 16.14 | 6.71 | 34.8 | | | | | | | | |
| SmUGT42 | 486 | 34.57 | 14.4 | 8.02 | 43 | | | | | | | | |

Table 2. The secondary structure of the SmUGT proteins

In the phylogenetic tree, the members of UGT71 had high expression levels from S1 to S4 and were downregulated from S5 to S8 (except *SmUGT19* and *SmUGT25*, which showed the opposite trend). Other than *SmUGT28* and *SmUGT29*, the members of UGT73 showed down-regulation from S5 to S8. Three members of UGT74 were



Fig. 6 Three-dimensional model templates of several SmUGT proteins. **a** Prediction model templates for SmUGT15-SmUGT21 and SmUGT25 of UGT71 are both 2acv.1.A. **b** Prediction model templates for SmUGT12-SmUGT14 and SmUGT27 of UGT73 are 7c2x.1.A. **c** Prediction model templates for SmUGT28-SmUGT30 of UGT73 are all 6lzx.1.A. **d** Prediction model templates for SmUGT1, SmUGT2 and SmUGT11 of UGT74 are all 2pq6.1.A. **e** Prediction model templates for SmUGT24, SmUGT31-SmUGT34 are all 6l90.1.A. **f** Prediction model templates for SmUGT23, SmUGT39 and SmUGT42 of UGT74 are all 7ery.1.A

upregulated from S5 to S8. Four members of UGT94 were upregulated from S5 to S8. The expression levels of the 42 *SmUGT*s were high in female flowers at the S1 and S3 stages of fruit development but low in mature fruit (S8).

Validation by RT-qPCR showed that *SmUGT6*, *SmUGT7*, *SmUGT8*, *SmUGT12*, and *SmUGT13* were highly expressed from S1 to S4, whereas *SmUGT28* was highly expressed from S5 to S8 (Fig. 8). The results of RT-qPCR were essentially consistent with those of the transcriptomic analysis (Figure S3).

As shown in Fig. 9, SmUGT1, SmUGT5, SmUGT9, SmUGT10, SmUGT27, SmUGT38, and SmUGT39 were strongly correlated with other genes in the triterpenoid saponin synthesis pathway. SmUGT1 and SmUGT16 were strongly correlated with SmERFs, SmGATAs and SmMYBs, while SmUGT31 was strongly correlated with SmbZIPs, SmERFs and SmMYBs.

Relations between SmUGTs and saponins

We performed a correlation analysis of the contents of total saponins and monomeric saponins [32] (obtained previously, Table S3) with the changes in *SmUGT*

s expression at the different stages of development (Fig. 10). Correlation clustering was also conducted.

There were strong positive correlations between Saponin 2 content and the expression levels of *SmUGT9*, *SmUGT16*, and *SmUGT17* during all stages. The content of Saponin 4 was positively correlated with the expression levels of *SmUGT1*, *SmUGT11*, *SmUGT16*, and *SmUGT17*; Saponin 9 and Saponin 54 with *SmUGT31*; Saponin 24 with *SmUGT1*, *SmUGT16*, *SmUGT17*, and *SmUGT30*; and Saponin 27 and Saponin 51 with *SmUGT1* and *SmUGT9*, respectively. In addition, the expression levels of *SmUGT9*, *smUGT32*, *SmUGT36*, and *SmUGT39* showed strong negative correlations with the contents of several saponins, particularly those of Saponin 30 and Saponin 32.

Discussion

Gene structural and conserved motifs of SmUGTs

Plant UGTs belong to the GT1 family and function in the synthesis of triterpenoid saponins. The members of this family have the conserved domain UDPGT (PF00201), whose C-terminal region contains a highly conserved sequence of 44 amino acids (PSPG-box). This sequence, together with the UDPGT domain, is a marker of UGTs.



Fig. 7 Expression levels of SmUGT genes during the indicated periods (a) and in the indicated tissues (b). The heat map showed red, black and green representing high, medium and low expression, respectively, and clustering related expression levels



Fig. 8 Validation of differential expression results obtained by RNA-seq. Relative expression levels determined by RT-qPCR and RNA-seq analyses of six selected genes at eight stages of fruit development



Fig. 9 Coexpression network of *SmUGTs* and other genes (**a**) and transcription factors (b). **a** Colors range from orange to dark blue, and shapes vary from small to large representing the spectrum of gene expression strength. **b** Hexagons represent *SmUGTs* and circles represent transcription factors

The sequence of the N-terminal region of UGTs is highly diverse, which explains their ability to recognize a variety of substrates [35]. The C-termini of the 42 SmUGTs harbored UDPGT domains (Figure S1) and PSPGboxes (Fig. 2), indicating that the locations of the typical domains of this protein family were conserved among species. The specificity of the N-terminus of the SmUGTs protein was also noted. In addition, 97.6% of the SmUGTs were acidic and 78.6% were hydrophilic, which are similar to those of *Dimocarpus longan* [36], suggesting that SmUGTs are highly conserved proteins.

The position, loss, and gain of introns can be used to provide insight into the evolution of gene families within phylogenetic groups. Among the 42 genes, 45.2% had intron insertions (Fig. 1), a value lower than those of *Hippophae rhamnoides* (50.6%) [37], maize (60%) [20], and *Arabidopsis thaliana* (58%) [18], and similar to that of the Sapindaceae *Dimocarpus longan* (47.3%) [36]. Therefore, we speculate that intron insertion was gradually lost during the evolution of *SmUGTs*.

Phylogenetic analysis of SmUGTs

UGTs are present in several gene families in plants, which are classified into 14 groups (A to N) in *Arabidopsis thaliana* [18]. The 42 SmUGTs in this study belonged to the subfamilies UGT71, UGT73, UGT74, and UGT94; they were divided into groups D, E, F and G. We evaluated the members of these groups in other species, among which group E is typically the largest [38]. In this study, 20% of SmUGTs belonged to group E (Fig. 4). Members of different groups exhibit different functional characteristics. Members of group D and group E can recognize a series of substrates, including terpenes [39]. Members of group F and group G are abundant in plants that produce triterpenoid saponins (e.g., *Panax ginseng* and *Panax quinquefolium* [23, 24]), and catalyze the glycosylation of substrates such as proginsenediol into ginsenosides. Therefore, it is speculated that these SmUGTs are related to glycosylation in the synthesis of triterpenoid saponins.

Phylogenetic analysis can provide insight into the evolutionary relationships among species and enable the identification of putative direct and collateral homologues. Homologous genes typically cluster into the same subgroup or subbranch and have similar functions [40]. The 10 UGTs associated with triterpenoid saponin synthesis in Barbarea vulgaris were members of UGT73 and catalyzed the 3-O-glucosylation of the sapogenins oleanolic acid and hederagenin [25]. Therefore, the genes of this subfamily (e.g., SmUGT12, SmUGT13) in soapberry may have similar functions. In Medicago truncatula, the function of glycosylation at C28/C3 of members of UGT71 and UGT73 subfamilies has been reported [26], so we performed a similar functional prediction for SmUGTs. The potential functions of SmUGTs will be further elucidated in future studies. As a woody plant, soapberry is characterized by a long growth cycle and the production of multiple types of saponin. The complex structure and low content of triterpenoid saponins make

| | | _ | | | | | | ſ | | | _ | | | | | | | | _ | | 1 | - | | _ | | | | | | 1 . | | _ | | _ | | | | | | | | | |
|------|------|-------|--------|---------|--------|-------|-----------|---------|---------|--------|--------|---------|---------|---------|-------|---------|---------|--------|--------------------|--------|----------|-------|--------|---------|---------|----------|---------|----------|---------|-------------|--------------|----------|---------|---------|----------|--------------------|-------------|--------|--------|--------|--------|-------|---------------|
| | | | Г | | | | | | | | | - | | - | | | | _ | | | <u> </u> | Г | 5 | | | | _ | _ | | | Г | | _ | | _ | | _ | _ | | | | 7 | |
| | | | | | | Ē | _ | | | ⊥ | ⊥ | | | | | | | | | | | | | | | | | | | | | | | | _ | | | | | | | | |
| ſ | | 0.52 | 0.69 | 9 0.67 | 0.63 | 0.63 | 0.63 | 0.54 | 0.57 | 0.31 | 0.46 | 0.22 | 0.32 | 0.15 | 0.15 | 0.23 | 0.49 | 0.42 | 0.10 | 0.60 | 0.49 | 0.64 | 0.33 | 0.17 | 0.15 | 0.07 | -0.11 - | -0.12 - | 0.23 -0 | .46 0.0 | 7 -0.46 | -0.26 | -0.26 | 0.07 | 0.11 | 0.04 - | 0.05 | -0.13 | -0.33 | -0.37 | -0.56 | -0.89 | Saponin 23 |
| | | 0.63 | 0.45 | 5 0.49 | 0.45 | 0.45 | 0.47 | 0.37 | 0.60 | 0.14 | 0.41 | 0.23 | 0.18 | 0.18 | 0.20 | -0.03 | 0.41 | 0.23 | -0.02 | 0.32 | 0.36 | 0.28 | 0.17 | -0.13 | 0.02 | •0.09 | 0.17 | -0.20 -1 | 0.24 -0 | .35 0.3 | 8 +0.21 | 0.03 | 0.02 | 0.08 | 0.14 | 0.04 | 0.07 | -0.08 | -0.36 | -0.44 | -0.30 | -0.68 | Saponin 41 |
| | ſĹ | 0.69 | 0.53 | 3 0.56 | 0.53 | 0.53 | 0.56 | 0.47 | 0.66 | 0.26 | 0.49 | 0.27 | 0.28 | 0.27 | 0.28 | 0.06 | 0.48 | 0.30 | 0.03 | 0.38 | 0.45 | 0.33 | 0.21 | -0.10 | -0.00 | -0.10 | -0.23 | -0.22 -1 | 0.35 -0 | .42 0.3 | 5 -0.16 | 0.07 | 0.05 | 0.16 | 0.05 | 0.12 | 0.04 | -0.09 | -0.38 | -0.46 | -0.35 | -0.67 | Saponin 52 |
| l | 1_ | 0.61 | 0.50 | 0 0 54 | 0.46 | 0.48 | 0.53 | 0.40 | 0.48 | 0.42 | 0.44 | 0.24 | 0.24 | 0.17 | 0.26 | 0.12 | 0.44 | 0.28 | 0.27 | 0.44 | 0.64 | 0.25 | 0.11 | 0.12 | 0.06 | 0.01 | -0.21 | -0.12 -1 | 0.44 -0 | 44 0.1 | 2 -0.25 | -0.01 | -0.01 | -0.03 | 0.10 | 0.06 | 0.12 | -0.24 | -0.34 | -0.39 | -0.62 | -0.74 | Sanonin 17 |
| | Ц | 0.01 | 0.00 | | | 0.40 | 0.00 | 0.10 | | | 0.44 | 0.24 | 0.24 | | 0.20 | 0.12 | | 0.20 | 0.27 | 0.44 | 0.01 | 0.20 | | 0.12 | 0.00 | | | | | | | | -0.01 | -0.05 | | | | | | | 0.02 | | Saponin 17 |
| Г | ſ | 0.52 | 0.54 | • 0.52 | 0.49 | 0.48 | 0.51 | 0.44 | 0.54 | 0.32 | 0.40 | 0.20 | 0.20 | 0.11 | 0.13 | 0.04 | 0.37 | 0.23 | 0.29 | 0.51 | 0.57 | 0.39 | 0.20 | 0.03 | 0.09 | 0.04 | -0.16 | -0.14 - | 0.31 -0 | .45 0.0 | 5 -0.38 | +0.17 | +0.20 | 0.00 | 0.18 | -0.04 - | 0.12 | -0.25 | -0.32 | -0.41 | -0.57 | 0.81 | Saponin 34 |
| | | 0.55 | 0.58 | 8 0.55 | 0.54 | 0.52 | 0.54 | 0,50 | 0.60 | 0.33 | 0.43 | 0.20 | 0.25 | 0.16 | 0.15 | 0.07 | 0.39 | 0.25 | 0.26 | 0,54 | 0.58 | 0.45 | 0.26 | 0.01 | 0.08 | 0.03 | -0.19 | -0.15 -1 | 0.33 -0 | .47 0.0 | 4 -0.35 | -0.15 | -0.20 | 0.08 | 0.13 | 0.00 - | 0.09 | -0.24 | -0.33 | -0.41 | -0.52 | -0,78 | Saponin 47 |
| | | 0.56 | 0.55 | 5 0.62 | 0.49 | 0.54 | 0.60 | 0.41 | 0.37 | 0.56 | 0.64 | 0.56 | 0.34 | 0.24 | 0.39 | 0.28 | 0.55 | 0.47 | 0.24 | 0.30 | 0.44 | 0.31 | -0.02 | 0.13 | •0.10 | -0.16 | -0.22 | -0.35 -1 | 0.33 -0 | .69 0.1 | 0 -0.16 | 0.07 | 0.11 | 0.01 | 0.07 | 0.17 | 0.11 | 0.23 | -0.03 | -0.49 | -0.69 | -0.77 | Saponin 37 |
| 14 | | 0.86 | 0.67 | 7 0.73 | 0.68 | 0.70 | 0.75 | 0.63 | 0.70 | 0.56 | 0.72 | 0.44 | 0.54 | 0.53 | 0.60 | 0.34 | 0.71 | 0.55 | 0.02 | 0.33 | 0.55 | 0.35 | 0.21 | -0.01 | -0.15 | -0.25 | 0.41 | -0,36 - | 0.60 +0 | .56 0.4 | 1 0.10 | 0.34 | 0.33 | 0.33 | 0.27 | 0.42 | 0.08 | 0.09 | -0.35 | -0.53 | -0.43 | -0.55 | Saponin 51 |
| | ſ | 0.70 | 0.60 | 0.61 | 0.58 | 0.58 | 0.61 | 0.54 | 0.65 | 0.41 | 0.55 | 0.29 | 0.35 | 0.31 | 0.34 | 0.14 | 0.52 | 0.35 | 0.16 | 0.43 | 0.57 | 0.36 | 0.22 | -0.04 | -0.03 | -0.11 | -0.29 | -0.25 -1 | 0.46 -0 | .50 0.2 | 3 -0.15 | 0.07 | 0.05 | 0.17 | 0.01 | 0.16 | 0.04 | -0.12 | -0.36 | -0.48 | -0.47 | -0.68 | Saponin 39 |
| | [[L | 0.71 | 0.66 | 5 0.67 | 0.64 | 0.63 | 0.66 | 0.58 | 0.69 | 0.37 | 0.55 | 0.25 | 0.38 | 0.33 | 0.34 | 0.19 | 0.56 | 0.41 | 0.07 | 0.44 | 0.52 | 0.46 | 0.31 | -0.04 | -0.03 | -0.12 | 0.30 | -0.26 | 0.44 -0 | .46 0.2 | 7 -0.18 | 0.04 | 0.02 | 0.21 | 0.03 | 0.18 | 0.06 | -0.11 | -0.42 | -0.49 | -0.43 | -0.70 | Saponin 49 |
| | ٦_ | 0.78 | 0.61 | 1 0.65 | 0.62 | 0.62 | 0.65 | 0.57 | 0.71 | 0.33 | 0.56 | 0.26 | 0.40 | 0.40 | 0.41 | 0.18 | 0.59 | 0.41 | -0.05 | 0.32 | 0.44 | 0.37 | 0.31 | -0.15 | -0.12 | -0.22 | 0.37 | -0.33 -1 | 0.48 -0 | 41 0.4 | 2 -0.04 | 0.18 | 0.17 | 0.27 | 0.09 | 0.25 | 0.05 | -0.05 | -0.46 | -0.54 | -0.30 | -0.58 | Saponin 46 |
| ſ | 4 | 0.76 | 0.57 | 7 0.64 | 0.57 | 0.50 | 0.63 | 0.50 | 0.61 | 0.29 | 0.56 | 0.21 | 0.36 | 0.36 | 0.42 | 0.20 | 0.50 | 0.42 | 0.00 | 0.28 | 0.45 | 0.30 | 0.21 | 0.08 | 0.12 | 0.22 | 0.34 | 0.32 | 0.48 .0 | 42 0.3 | 0.007 | 0.18 | 0.19 | 0.16 | 0.06 | 0.22 | 0.06 | 0.02 | 0.41 | 0.54 | 0.41 | 0.64 | Samania 62 |
| | | 0.70 | 0.57 | | 0.57 | 0.07 | 0.00 | 0.00 | 0.01 | | 0,00 | 0.51 | 0.50 | | 0.42 | 0.20 | 0.57 | 0.42 | 0.00 | 0.20 | 0.45 | | 0.51 | -0.08 | -0.12 | -0.22 | -0,04 | -0.52 | 0.48 -0 | | -0.07 | 0.10 | 0.15 | 0.10 | | 0.25 | | 0.02 | | -0,54 | -0.41 | | Saponin 33 |
| | | 0.06 | -0.22 | 2 =0.32 | -0.21 | -0.24 | •0.19 | -0.03 | 0.00 | 0.41 | 0.09 | 0.28 | 0.02 | 0.06 | 0.08 | -0.15 | -0.21 | -0.25 | 0.62 | 0.27 | 0.53 | -0.38 | -0.49 | 0.31 | 0.28 | 0.35 | 0.20 | 0.35 | 0.12 -0 | .34 -0 | 8 0.38 | 0.26 | 0.15 | 0.09 | 0.20 | 0.11 | 0.41 | -0.07 | 0.50 | 0.49 | -0.21 | 0.35 | Saponin 44 |
| 11 | | 0.49 | -0.15 | 5 -0.02 | -0.03 | -0.01 | 0.02 | -0.05 | 0.38 | -0.17 | 0.21 | 0,34 | 0.05 | 0.29 | 0.31 | -0.22 | 0.16 | -0.01 | -0.48 | -0.10 | -0.04 | -0,36 | -0.30 | -0.11 | 0,12 | -0.03 | 0.13 | 0,11 0 | 0.03 -0 | .06 0.8 | 3 0,51 | 0.63 | 0.64 | 0.24 | 0.16 | 0.31 | 0,37 | 0.31 | 0.04 | 0.13 | 0.39 | 0.21 | Saponin 1 |
| | 1 | 0.07 | -0.05 | 7 -0.14 | -0.04 | -0.08 | -0.02 | 0.13 | 0.06 | 0.52 | 0.26 | 0.29 | 0.19 | 0.29 | 0.30 | 0.02 | -0.02 | -0.06 | 0.50 | -0.06 | 0.31 | -0.30 | -0.26 | -0.11 | -0.23 | -0.15 | 0.24 | -0.17 - | 0.37 -0 | .36 -0.3 | 7 0.56 | 0.44 | 0.33 | 0.29 | 0.41 | 0.27 | 0.29 | 0.14 | 0.33 | 0.02 | -0.08 | 0.44 | Saponin 38 |
| | Ίг | 0.32 | 0.15 | 5 0.04 | 0.25 | 0.16 | 0.17 | 0.41 | 0.51 | 0.31 | 0.31 | 0.16 | 0.40 | 0.51 | 0.36 | 0.09 | 0.16 | 0.08 | 0.06 | 0.19 | 0.32 | 0.03 | 0.09 | -0.12 | -0.03 | -0.06 | 0.19 | -0.01 | 0.35 +0 | .22 0.1 | 0 0.61 | 0.50 | 0.35 | 0.66 | 0.52 | 0.47 | 0.38 | 0.06 | 0.04 | 0.15 | 0.28 | 0.52 | Saponin 28 |
| | 1 | 0.23 | 0.12 | 2 -0.02 | 0.19 | 0.10 | 0.10 | 0.36 | 0.45 | 0.29 | 0.22 | 0.08 | 0.32 | 0.40 | 0.25 | 0.03 | 0.06 | -0.01 | 0.19 | 0.25 | 0.38 | 0.02 | 0.08 | -0.08 | 0.03 | 0.03 | -0.15 | 0.06 | 0.33 -0 | .19 -0.0 | 0.49 | 0.36 | 0.20 | 0.55 | 0.41 | 0.35 | 0.29 | -0.08 | 0.03 | 0.18 | 0.19 | 0.45 | Saponin 36 |
| | _ | 0.85 | 0.31 | 1 0.44 | 0.42 | 0.44 | 0.51 | 0.42 | 0.59 | 0.41 | 0.66 | 0.49 | 0.49 | 0.69 | 0.76 | 0.23 | 0.61 | 0.43 | -0.27 | -0.14 | 0.20 | -0.10 | -0.00 | -0.29 | -0.38 | -0.50 | 0.47 | -0.46 -1 | 0.57 -0 | .38 0.7 | 4 0.67 | 0.83 | 0.83 | 0.51 | 0.53 | 0.63 | 0.26 | 0.40 | -0.20 | -0.43 | 0.07 | 0.09 | Saponin 2 |
| r | - | 0.56 | 0.21 | 1 0.21 | 0.34 | 0.31 | 0.34 | 0.47 | 0.53 | 0.48 | 0.58 | 0.46 | 0.60 | 0.78 | 0.73 | 0.31 | 0.44 | 0.36 | -0.18 | -0.08 | 0.17 | -0.06 | -0.01 | -0.19 | -0.30 | -0.37 | 0.37 | -0.29 -1 | 0.49 -0 | 36 0.4 | 4 0.94 | 0.89 | 0.80 | 0.79 | .82 | 0.78 | 0.54 | 0.48 | 0.10 | -0.05 | 0.31 | 0.61 | Saponin 24 |
| | 4 | 0.62 | 0.27 | | 0.27 | 0.22 | 0.26 | 0.62 | 0.00 | 0.62 | 0.00 | 0.17 | 0.50 | 0.24 | 0.02 | 0.27 | 0.41 | 0.22 | 0.02 | 0.10 | 0.22 | 0.00 | 0.01 | 0.11 | 0.18 | 0.26 | 0.22 | 0.20 | 0.50 0 | 46 0.2 | | 0.70 | 0.02 | 0.77 | 26 | 0.72 | 0.66 | 0.22 | 0.11 | 0.00 | 0.10 | 0.40 | Saponin 24 |
| 4 | | 0.37 | 0.27 | 0.23 | 0.37 | 0.33 | 0.36 | 0.32 | 0.00 | 0,33 | 0.00 | 0,47 | 0.39 | 0,74 | 0.07 | 0.27 | 0.41 | 0.32 | -0.03 | 0.10 | 0.33 | -0.00 | -0.01 | -0.11 | -0.18 | -0.23 | -0.52 | -0.20 | 0,30 -0 | .45 0.3 | 2 0.84 | 0,79 | 0.67 | 0,77 | | 0.72 | 0,33 | 0.37 | 0.11 | -0,00 | 0.19 | 0,49 | Saponin 27 |
| | ſ | 0.39 | -0.06 | 6 0.14 | 0.03 | 0.07 | 0.10 | +0.09 | 0.09 | -0.27 | 0.06 | -0.03 | -0.03 | 0.18 | 0.25 | +0.05 | 0.24 | 0.11 | -0.59 | +0.53 | +0.38 | -0.24 | 0.09 | -0.48 | -0.41 | -0.50 | -0.32 - | -0.39 -1 | 0.15 0. | 26 0.7 | 6 0.23 | 0.39 | 0.49 | 0.02 | 0.01 | 0.13 | 0.24 | 0.17 | -0.42 | -0.47 | 0.35 | 0.03 | Saponin 5 |
| [| 1 | 0.39 | 0.12 | 2 0.21 | 0.20 | 0.17 | 0.17 | 0.12 | 0.29 | -0.26 | +0.03 | +0.36 | 0.01 | 0.17 | 0.13 | -0.07 | 0.20 | 0.03 | -0.41 | +0.32 | +0.16 | -0.00 | 0.44 | •0.60 | •0.40 | -0.45 | -0.48 · | -0.38 -1 | 0.35 0. | 36 0.5 | 4 0.04 | 0.16 | 0.18 | 0.11 | 0.09 | 0.02 | 0.51 | -0.28 | -0.75 | -0.56 | 0.33 | -0.03 | Saponin 10 |
| | L | 0.16 | -0.22 | 2 -0.14 | -0.08 | -0.12 | -0.10 | -0.05 | 0.03 | -0.24 | -0.11 | -0.32 | 0.02 | 0.32 | 0.25 | -0.05 | 0.03 | -0.07 | -0.48 | -0.73 | -0.46 | -0.37 | 0.24 | -0,74 | -0.64 | -0.64 | 0.54 | -0.46 -1 | 0.33 0. | 48 0.5 | 2 0.60 | 0.56 | 0.55 | 0.32 | 0.28 | 0.25 | 0.28 | 0.04 | -0,43 | -0.36 | 0,73 | 0.72 | Saponin 3 |
| | L | 0.33 | -0.18 | 8 -0.08 | -0.02 | +0.03 | 0.00 | 0.04 | 0.15 | 0.00 | 0.18 | 0.11 | 0.21 | 0.52 | 0.51 | 0.06 | 0.18 | 0.09 | -0.47 | -0.61 | -0.32 | -0.42 | -0.03 | -0.52 | -0.52 | -0.57 | -0.42 · | -0.39 -1 | 0.33 0. | 15 0.6 | 4 0.89 | 0.85 | 0.84 | 0.50 | 0.55 | 0.52 | 0.16 | 0.40 | -0.10 | -0.18 | 0.65 | 0.79 | Saponin 4 |
| | _ | +0.31 | -0.07 | 7 -0.15 | -0.16 | -0.20 | -0.22 | -0.21 | -0.16 | -0.32 | -0.47 | -0.54 | -0.49 | -0.61 | -0.64 | -0.42 | -0.39 | -0.44 | 0.42 | 0.24 | 0.18 | 0.08 | 0.18 | -0.05 | 0.22 | 0.31 | 0.13 | 0.20 | 0.14 0. | 27 -0.4 | 7 -0.83 | -0.79 | -0.80 | -0.57 | 0.78 | -0.74 - | 0.65 | 0.80 | -0.35 | -0.11 | -0.32 | -0.59 | Saponin 14 |
| [| | -0.32 | -0.05 | 9 -0.18 | -0.17 | -0.22 | -0.23 | -0.19 | -0.19 | -0.27 | -0.54 | -0.66 | -0.44 | -0.56 | -0.60 | -0.34 | -0.40 | -0.44 | 0.43 | 0.28 | 0.29 | 0.04 | 0.21 | 0.07 | 0.28 | 0.39 | 0.12 | 0.33 | 0.02 0. | 39 -0.: | -0.73 | -0.72 | -0.75 | -0.53 | 0.69 | -0.65 - | 0.68 | -0.95 | -0.43 | 0.03 | -0.29 | -0.40 | Saponin 45 |
| | _ | -0.17 | 0.09 | 0.00 | -0.02 | -0.05 | -0.05 | -0.06 | -0.05 | -0.07 | -0.27 | -0.35 | -0.34 | -0.51 | -0.51 | -0.30 | -0.25 | -0.30 | 0.57 | 0.43 | 0.41 | 0.17 | 0.12 | 0.13 | 0.30 | 0.38 | 0.13 | 0.22 | 0.03 0 | 03 -0 : | 3 -0.87 | -0.74 | -0.77 | -0.52 | 0.69 | -0.63 | 0.51 | -0.75 | -0.27 | -0.10 | -0.56 | -0.73 | Sanonin 22 |
| l | | | | | 0.10 | 0.16 | 0.10 | 0.10 | 0.07 | 0.17 | 0.20 | | 0.00 | 0.00 | | | 0.00 | 0.41 | | | 0.20 | 0.00 | 0.10 | 0.10 | | | | | | | | | | | | | | | 0.16 | | | | Saponin 22 |
| | _ | -0,34 | 0.01 | 1 -0,14 | -0,10 | -0.16 | -0.18 | -0.10 | -0.07 | -0.17 | -0.39 | -0,43 | -0,38 | -0,56 | -0.65 | -0.36 | -0.39 | -0.41 | 0.56 | 0.52 | 0,38 | 0.20 | 0.13 | 0.19 | 0,44 | 0.55 | 0.28 | 0.39 | 0.18 0. | 10 -0.0 | 4 -0.80 | -0.81 | -0.88 | -0,46 | 0.67 | -0.65 | -0.41 | 0.78 | -0,16 | 0.14 | -0.41 | -0.56 | Saponin 50 |
| | ſ | 0.31 | 0.49 | 9 0.43 | 0.39 | 0.38 | 0.41 | 0.36 | 0.36 | 0.36 | 0.28 | 0.12 | 0.11 | -0.06 | -0.03 | 0.02 | 0.23 | 0.13 | 0.52 | 0.60 | 0.65 | 0.40 | 0.15 | 0.18 | 0.19 | 0.19 | -0.07 - | -0.02 -1 | 0.25 -0 | .45 -0.3 | 18 -0.55 | -0.38 | -0.42 | -0.16 | 0.29 | -0.19 - | 0.18 | -0.38 | -0.22 | -0.30 | -0.73 | -0.87 | Saponin 25 |
| | | 0.17 | 0.42 | 2 0.31 | 0.32 | 0.29 | 0.30 | 0.31 | 0.31 | 0.32 | 0.17 | 0.04 | 0.05 | -0.15 | -0.16 | -0.04 | 0.09 | 0.03 | 0.61 | 0.70 | 0.69 | 0.41 | 0.14 | 0.26 | 0.32 | 0.34 | 0.04 | 0.13 | 0.17 •0 | .40 -0.4 | 4 -0.60 | +0.50 | +0.56 | -0.18 | 0.34 | -0.27 - | 0.16 | 0.49 | -0.15 | -0.13 | -0.72 | -0.81 | Saponin 31 |
| IIIr | 5 | 0.08 | 0.32 | 2 0.20 | 0.22 | 0.18 | 0.19 | 0.22 | 0.24 | 0.17 | 0.00 | -0.15 | -0.07 | -0.24 | -0.27 | -0.14 | -0.02 | -0.10 | 0.58 | 0.63 | 0.63 | 0.33 | 0.18 | 0.19 | 0.32 | 0.36 | 0.05 | 0.18 - | 0.15 -0 | .21 -0.4 | 4 -0.64 | -0.55 | -0.62 | -0.25 | 0.44 | -0.37 - | 0.30 | 0.65 | -0.26 | -0,10 | -0.61 | -0.73 | Saponin 35 |
| | llr | 0.17 | 0.32 | 2 0.25 | 0.23 | 0.21 | 0.22 | 0.20 | 0.27 | 0.14 | 0.05 | -0.08 | -0.08 | -0.24 | -0.23 | -0.15 | 0.04 | -0.06 | 0.49 | 0.61 | 0.61 | 0.30 | 0.14 | 0.20 | 0.33 | 0.35 | 0.07 | 0.17 - | 0.12 -0 | .24 -0.3 | -0.65 | -0.51 | -0.55 | -0.29 · | 0.47 | -0.36 - | 0.29 | 0.58 | -0.27 | -0.14 | -0.63 | -0.82 | Saponin 43 |
| | 1_ | 0.15 | 0.42 | 2 0.37 | 0.30 | 0.30 | 0.31 | 0.23 | 0.21 | 0.20 | 0.13 | 0.02 | -0.02 | -0.23 | -0.19 | -0.02 | 0.13 | 0.08 | 0.46 | 0.58 | 0.52 | 0.42 | 0.16 | 0.24 | 0.26 | 0.27 | 0.05 | 0.06 | 0.08 +0 | .32 -0.3 | 3 -0.73 | -0.57 | -0.57 | -0.32 | 0.44 | -0.34 - | 0.26 | 0.42 | -0.22 | -0.24 | -0.73 | -0.95 | Total Saponin |
| | 4 | 0.20 | 0.34 | 4 0.27 | 0.26 | 0.23 | 0.24 | 0.22 | 0.31 | 0.14 | 0.11 | -0.02 | -0.07 | -0.20 | -0.21 | -0.17 | 0.06 | -0.05 | 0.47 | 0.53 | 0.52 | 0.32 | 0.16 | 0.05 | 0.22 | 0.23 | 0.00 | 0.03 | 0.11 -0 | 29 -0.3 | 6 -0.62 | -0.48 | -0.52 | -0.23 | 0.45 | -0.35 - | 0.27 | -0.48 | -0.26 | -0.26 | -0.58 | -0.83 | Saponin 13 |
| 11 | 4 | 0.26 | 0.37 | 7 0.33 | 0.29 | 0.28 | 0.30 | 0.24 | 0.29 | 0.21 | 0.10 | -0.07 | -0.01 | -0.16 | -0.12 | -0.06 | 0.13 | 0.02 | 0.45 | 0,56 | 0.64 | 0.30 | 0.17 | 0.20 | 0.26 | 0.27 | -0.02 | 0.11 - | 0.24 -0 | .23 -0.3 | 3 -0.59 | -0.43 | -0.45 | -0.27 | 0,40 | -0.28 - | 0.32 | -0.58 | -0.35 | -0.22 | -0.66 | -0.81 | Saponin 16 |
| | | 0.27 | 0.01 | 1 0.08 | 0.01 | 0.01 | 0.04 | 0.10 | 0.20 | 0.14 | 0.02 | 0.05 | 0.26 | 0.25 | 0.18 | 0.26 | 0.01 | 0.18 | 0.12 | 0.22 | 0.26 | 0.10 | 0.17 | 0.15 | 0.28 | 0.22 | 0.26 | 0.27 | 07 0 | 00 0.7 | 0 0.41 | 0.20 | 0.17 | 0.28 | 0.61 | 0.22 | 0.16 | 0.25 | 0.18 | 0.02 | 0.25 | 0.66 | Garage 22 |
| | | 0.27 | 0.01 | 1 0.08 | -0.01 | 0.01 | 0.04 | -0.10 | 0.20 | -0.14 | 10.02 | 0.00 | -0.25 | -0.23 | -0.18 | 10.30 | -0.01 | -0.18 | 0.15 | 0.33 | 0.30 | | -0.17 | 0.15 | 0.38 | 0.33 | 0.25 | 0.27 | | .09 0.2 | 0 10.41 | 10.20 | 10.17 | 10.38 | 0.51 | | 0.10 | .0.35 | -0.18 | -0.02 | 0.35 | 0.00 | Saponin 33 |
| [| 5 | 0.12 | 0.26 | 5 0.12 | 0.15 | 0.10 | 0.09 | 0.15 | 0.12 | 0.10 | -0.08 | -0.18 | -0.14 | -0,36 | -0.41 | -0.16 | -0.14 | -0.16 | 0.62 | 0.61 | 0.50 | 0.38 | 0.18 | 0.19 | 0.33 | 0.40 | 0.13 | 0.19 0 | 0.02 -0 | .20 -0.0 | 51 -0.75 | -0.71 | -0.78 | -0.30 | 0,50 | -0.48 - | 0.29 | 0.59 | -0.14 | -0.06 | -0.59 | -0.72 | Saponin 26 |
| | IIIr | -0.09 | 0.19 | 9 0.06 | 0.09 | 0.04 | 0.03 | 0.09 | 0.13 | 0.03 | •0.16 | -0.25 | -0.20 | -0.39 | -0.43 | -0.24 | -0.18 | -0.23 | 0.58 | 0.64 | 0.57 | 0.28 | 0.13 | 0.26 | 0.44 | 0.50 | 0.20 | 0.34 | 0.01 -0 | .09 -0.: | 3 -0.73 | -0.68 | -0.74 | -0.35 | 0.55 | -0.49 - | 0.32 | 0.72 | -0.20 | 0.06 | -0.55 | -0.68 | Saponin 40 |
| | 4 | -0.06 | 0.19 | 9 0.11 | 0.08 | 0.06 | 0.05 | 0.03 | 0.09 | -0.03 | -0.14 | -0.20 | -0.25 | -0.44 | -0.44 | -0.25 | -0.15 | -0.20 | 0.51 | 0.57 | 0.47 | 0.28 | 0.11 | 0.21 | 0.39 | 0.43 | 0.20 | 0.25 | 0.07 -0 | -0.4 | 4 -0.82 | +0.71 | -0.74 | -0.45 | 0.64 | -0.55 - | 0.36 | 0.64 | -0.21 | -0.06 | -0.60 | -0.83 | Saponin 48 |
| 4 | Г | 0.01 | 0.11 | 1 0.15 | 0.01 | 0.04 | 0.06 | -0.11 | -0.09 | -0.08 | -0.16 | -0.23 | -0.31 | -0.42 | -0.32 | -0.20 | -0.05 | -0.13 | 0.31 | 0.18 | 0.24 | 0.08 | 0.08 | 0.04 | 0.09 | 0.14 | 0.00 | 0.01 - | 0.03 0. | 05 -0.1 | 9 -0.73 | -0.55 | -0.49 | -0.58 | 0.63 | -0.52 - | 0.56 | -0.53 | -0.37 | -0.33 | -0.55 | -0.81 | Saponin 21 |
| | ΎΓ | 0.21 | 0.27 | 7 0.27 | 0.19 | 0.19 | 0.21 | 0.12 | 0.20 | 0.05 | -0.00 | +0.16 | -0.13 | -0.23 | -0.19 | -0.15 | 0.07 | -0.05 | 0.34 | 0.40 | 0.47 | 0.20 | 0.17 | 0.07 | 0.18 | 0.19 | -0.04 | 0.05 | 0.18 +0 | .09 •0. | 4 -0.62 | -0.44 | -0.44 | -0.35 | 0.50 | -0.36 | 0.44 | 0.59 | -0.43 | -0.29 | -0.55 | -0.79 | Saponin 18 |
| | Ч | 0.15 | 0.18 | 8 0.13 | 0.11 | 0.08 | 0.10 | 0.08 | 0.22 | 0.06 | -0.00 | -0.08 | -0.20 | -0.28 | -0.26 | -0.31 | -0.06 | -0.20 | 0.51 | 0,45 | 0.52 | 0.11 | 0.04 | 0.02 | 0.24 | 0.26 | 0.04 | 0.10 | 0.10 -0 | 20 -0.3 | 4 -0.58 | -0.43 | -0.47 | -0.33 | 0.53 | -0.43 - | 0.32 | 0.56 | -0.25 | -0.20 | -0.54 | -0.74 | Saponin 19 |
| | | 0.11 | 0.10 | 0 0.07 | 0.04 | 0.01 | 0.04 | -0.01 | 0.14 | -0.02 | -0.11 | -0.19 | -0.26 | -0.33 | -0.30 | -0.33 | -0.11 | -0.25 | 0.46 | 0.37 | 0.47 | 0.04 | 0.04 | 0.02 | 0.23 | 0.27 | 0.04 | 0.13 -4 | 0.10 -0 | .06 -0.3 | 2 -0.59 | -0.45 | -0.47 | -0.41 | 0.58 | -0.47 - | 0.42 | -0.65 | -0.32 | -0.19 | -0.50 | -0.69 | Saponin 20 |
| , | | -0.65 | -0.46 | 6 -0.44 | -0.49 | -0.50 | -0.55 | -0.58 | -0.65 | -0.78 | -0.85 | -0.88 | -0.65 | -0.63 | -0.67 | -0.35 | -0.55 | -0.49 | -0.21 | -0.52 | -0.64 | -0.19 | 0.26 | -0.37 | -0.17 | -0.07 | 0.01 | 0.01 | 0 33 0 | 90 -0 | 8 -0.53 | -0.59 | -0.51 | -0.57 | 0.64 | -0.66 | 0.82 | -0.53 | -0.43 | -0.12 | 0.36 | 0.08 | Sanonin 9 |
| | | 0.62 | 0.44 | 5 0 47 | 0.50 | 0.62 | 0.67 | 0.00 | 0.50 | 0.78 | 0.82 | 0.72 | 0.77 | 0.82 | 0.86 | 0.60 | 0.66 | 0.64 | 0.06 | 0.12 | 0.26 | 0.16 | 0.08 | 0.16 | 0.21 | 0.20 | 0.21 | 0.28 | | <i>(</i> 0' | 2 0.90 | 0.81 | 0.77 | 0.72 | | 0.87 | 0.60 | 0.66 | 0.26 | 0.06 | 0.00 | 0.20 | Gaponin 22 |
| | Γ | -0.63 | -0.43 | 5 -0.47 | -0,50 | -0.53 | -0.57 | +0.60 | -0.50 | -0.78 | -0.82 | -0.73 | -0,77 | -0.82 | -0.86 | -0.39 | -0.66 | -0.64 | 0.06 | -0.13 | -0,35 | -0.15 | 0.08 | -0.16 | 0.21 | 0.29 | 0,31 | 0.28 | 0.51 0. | 66 -0 | 02 -0.80 | -0.81 | -0.77 | -0.72 | 0.90 | -0,87 | 0.69 | 0.66 | -0.26 | 0.06 | 0.09 | -0.29 | Saponin 32 |
| | 14 | -0.58 | -0.31 | 7 -0.41 | -0.43 | -0.47 | -0.51 | -0.52 | -0,44 | -0.73 | -0.77 | -0.74 | -0.72 | -0.77 | -0.82 | -0.56 | -0.60 | -0.61 | 0.09 | -0.13 | -0.31 | -0.10 | 0.16 | -0.24 | 0.13 | 0.22 | 0.21 | 0.19 (| 0.42 0. | 64 -0.3 | 13 -0.80 | -0.80 | -0.77 | -0.67 | 0.87 | -0.85 - | 0.74 | 0.69 | -0.33 | -0.04 | 0.07 | -0.31 | Saponin 12 |
| | 1 | -0.62 | -0.38 | 8 -0.43 | -0.44 | -0.48 | +0.52 | -0.51 | -0.47 | -0.69 | -0.81 | -0.81 | -0.71 | -0.77 | -0.83 | -0.52 | -0.63 | -0.62 | 0.14 | -0.09 | +0.25 | +0.09 | 0.18 | •0.17 | 0.16 | 0.27 | 0.21 | 0.25 | 0.38 0. | 68 -0.4 | -0.80 | -0.83 | -0.81 | -0.67 · | 0.86 | -0.85 - | 0.76 | -0.77 | -0.35 | 0.02 | 0.05 | -0.26 | Saponin 30 |
| | 5 | -0.43 | -0.21 | 1 -0.26 | -0.29 | -0.32 | -0.35 | -0.36 | -0,30 | -0.50 | -0.62 | -0.65 | -0.61 | -0,70 | -0,74 | -0.49 | -0.49 | -0.52 | 0.30 | 0.08 | -0.02 | -0.01 | 0.17 | -0.12 | 0.18 | 0.28 | 0.16 | 0.20 | 0.25 0. | 45 -0.4 | 3 -0.85 | -0.82 | -0,81 | -0.65 | 0,85 | -0.81 - | 0.73 | 0.79 | -0.37 | -0.09 | -0.19 | -0.50 | Saponin 11 |
| L | ll | -0.50 | -0.24 | 4 -0.31 | -0.31 | -0.36 | -0.39 | -0.37 | -0.34 | -0.56 | -0.69 | -0.76 | -0.61 | -0.69 | -0.76 | -0.47 | -0.53 | -0.54 | 0.23 | 0.00 | -0.12 | 0.00 | 0.25 | -0.19 | 0.12 | 0.23 | 0.12 | 0.18 | 0.25 0. | 56 -0.4 | 4 -0.81 | -0.81 | -0.82 | -0.60 | 0.82 | -0.80 - | 0.77 | -0.81 | -0.41 | -0.09 | -0.05 | -0.36 | Saponin 15 |
| | 1 | -0.43 | -0.24 | 4 -0.27 | -0.30 | +0.33 | +0.37 | •0.40 | -0.29 | -0.65 | +0.62 | -0.62 | -0.64 | -0.68 | -0.74 | -0.52 | -0.48 | -0.50 | 0.08 | •0.06 | +0.26 | 0.01 | 0.19 | -0.28 | 0.10 | 0.16 | 0.16 | 0.09 | 0.38 0. | 49 -0.3 | -0.81 | -0.77 | -0.75 | -0.60 | 0.85 | -0.80 | 0.69 | 0.61 | -0.35 | -0.16 | 0.00 | -0.47 | Saponin 29 |
| | Ľ | -0.32 | -0.26 | 6 -0.26 | -0.30 | -0.33 | -0.36 | -0.41 | -0.20 | -0.69 | -0.56 | -0.52 | -0.63 | -0.64 | -0.69 | -0.57 | -0.44 | -0.50 | -0.02 | -0.04 | -0.25 | -0.05 | 0.11 | -0.26 | 0.17 | 0.20 | 0.23 | 0.15 | 0.42 0 | 45 -0.0 | -0.74 | -0.67 | -0.64 | -0.57 | 0.84 | -0.75 | 0.59 | .0.53 | -0.32 | -0.11 | 0.05 | -0.48 | Saponin 42 |
| , | | -0.37 | -0.32 | 2 -0.34 | -0.26 | -0.32 | -0.36 | -0.24 | -0.32 | -0.53 | -0.58 | -0.81 | -0.25 | -0.11 | -0.25 | -0.12 | -0.32 | -0.30 | -0.35 | -0.64 | -0.59 | -0.17 | 0.45 | -0.62 | -0.46 | -0.38 | -0.35 | -0.22 -1 | 0.06 0 | 85 0.0 | 2 0.10 | -0.06 | -0.07 | 0.03 | 0.07 | -0.16 | 0.61 | -0.36 | -0.51 | -0.18 | 0.72 | 0.66 | Saponin 54 |
| ļ | _ | -0.04 | 0.01 | 1 0.05 | -0.01 | -0.07 | -0.04 | -0.11 | -0.12 | -0.43 | -0.45 | -0.75 | -0.27 | -0.27 | -0.30 | -0.11 | -0.09 | -0.15 | -0.19 | -0.17 | -0.13 | 0.07 | 0.50 | -0.78 | -0.15 | -0.11 | 0.24 | 0.06 | 0.15 | 66 0.0 | 7 .0.48 | -0.40 | -0.35 | -0.33 | 0.47 | -0.37 | 0.84 | -0.71 | -0.81 | -0.35 | 0.10 | -0.22 | Sanonin 8 |
| l | | 0.10 | .001 | 0 0.02 | .0.00 | .0.07 | .0.09 | 0.22 | 0.02 | .0.44 | .0.25 | .0.29 | .0.25 | 0.27 | 0.36 | | .0.07 | .0.19 | .0.25 | .0.16 | .0.20 | .0.07 | 0.12 | .0.21 | .0.00 | .0.05 | 0.02 | | 1.18 | 25 0 | 2 .0.40 | .0.32 | .0.16 | .0.22 | 0.52 | 0.36 | 0.44 | 0.25 | 0.44 | 0.20 | 0.14 | 0.41 | Caponii 6 |
| | 4 | 0.10 | -0.05 | 0.01 | -0.08 | -0,07 | -0.08 | -0.23 | 0.05 | -0,33 | -0.25 | -0.28 | -0,53 | -0.27 | -0.26 | -0.54 | -0.07 | -0,18 | -0.55 | -0,10 | -0.26 | -0.06 | 0.13 | -0.51 | -0.00 | -5,00 | 0.03 | 0.04 | | | 0,40 | -0.23 | -0.13 | -3.52 | 0,33 | -0,30 | | -0.25 | -1.44 | -0.29 | 0.14 | | saponin o |
| | | -0.08 | -0.31 | 1 -0.18 | -0.29 | -0.27 | -0.27 | -0.44 | -0.21 | -0.65 | +0.42 | -0.37 | -0.51 | -0.40 | -0.36 | -0.43 | -0.23 | -0.32 | -0.34 | -0.36 | -0.41 | -0.26 | 0.02 | -0.34 | -0.06 | -0.08 | 0.07 | -0.03 (| 0.26 0. | 49 0.3 | 5 -0.37 | -0.24 | -0.14 | -0.47 | 0.59 | -0.46 | -0.51 -6 | -0.25 | -0.37 | -0.24 | 0.21 | -0.26 | Saponin 7 |
| | | NOL Y | JOTI's | NGL32 | SCI 2ª | GT12 | all's all | Ser' NG | 513° 21 | GT38 N | GI10 N | GT21 mi | SOLO IR | 3720 AV | ST37 | Selo Al | ST24 NG | 573' N | 372 ₃₁₆ | 3TH AN | GTAN M | JOLAN | ST2 mi | JOT? AU | STA' NG | STA- mil | ST NG | Sin Sing | N' NGT | NOT! | mUGTI | JOT10 JU | 511' 15 | 572,200 | 30 | S ^{SU} NG | 11/ AUG | in all | STI JC | STE NG | 52 300 | 50" | |
| | ್ | - CSI | - 65 | · 611 | GDL | GIL | -32 | GU1. | SUL | GTI | CU. | 210 | STIL | GU1. | -92 | GU1. | GT1. | SUL | GTH | C.S. | | cn. | Sec | 65° | CS1. | 2 | CS3 | 5 | GP (| N. 6 | ⇒ <u>c</u> Ω | GTI | GTH | CT1 | 671 | C.D. | 681 | GTI | GIL | GIN | STI | | |

Fig. 10 Correlation analysis between gene expression levels and saponin contents in soapberry. Red, white, and blue in the heat map represent positive, no, and negative correlation, respectively

it difficult to obtain these molecules directly by chemical synthesis [5]. In the future, leveraging the function of SmUGTs, targeted and efficient production of soapberry triterpenoid saponins may be achieved by synthetic biology and bioreactor cultivation, leading to increased economic and practical value of SmUGTs.

The soapberry genes in each group were very similar and often clustered together. The *UGT* genes of herbs and

woody plants may differ markedly. The proteins encoded by the clustered genes, e.g., SmUGT12, SmUGT13, and SmUGT14, showed consistent predicted 3D structures (Fig. 6 and Table S1), and shared a template.

Expression patterns of *SmUGTs* at different stages of fruit development

The expression levels of genes at different developmental stages and in different organs can reflect their functions. The expression levels of most SmUGTs were high from S1 to S4 and low from S5 to S8 (Fig. 7). RT-qPCR validated the gene expression patterns during fruit development (Fig. 8). We speculate that most of the genes mediate glycosylation during the early stages of fruit development and promote the accumulation of saponins in fruit peel during the later stages. In other species, the expression of UGT genes changes as fruit development progresses. In Hippophae rhamnoides, most UGT genes are expressed at high levels in the early or middle stages of fruit development, and at lower levels in later stages [37]. In peach, 35% of UGT genes have the highest expression level in S1 (the first exponential growth phase, characterized by rapid cell division) [42]. This pattern is consistent with the expression patterns of most SmUGTs in our study. In Gardenia jasminoides, the expression levels of some UGT genes decrease with the development and maturation of fruits. However, the expressions of others increase with the continuous ripening of fruits [43].

We also analyzed the expression levels of *SmUGTs* in vegetative organs (roots and leaves) and reproductive organs (fruits, female and male flowers). Most of the genes were expressed at high levels in the reproductive organs, particularly in female flowers and in fruits in an early stage of development. Indeed, in Medicago truncat*ula*, the expression of *UGT71G1* in flowers is higher than in roots, stems, and leaves [26]; many UGTs in maize are expressed at high levels in reproductive tissues, such as the anther and style [20]. In peach, the expression levels of 32% of UGTs are highest in flowers, and the abundances of 60 UGTs are highest during fruit development and postharvest maturation [42]. Therefore, UGTs are active in the reproductive organs of soapberry and the aforementioned plants, suggesting that in soapberry triterpenoid saponins are synthesized in flowers and fruits, and mainly accumulate in fruits.

An analysis of *cis*-acting elements suggested that light, hormones, and other factors modulate the expression of *UGTs*. In a previous study on *Epimedium pubescens*, the expression levels of > 12 *UGT* genes of group E and group D were significantly induced by intense light [44]. In a study on ginseng, MeJA promoted the expression of *UGT* and upregulated upstream terpenoid biosynthesis genes [41, 45]. *UGT* expression is induced by jasmonic acid in *Litsea cubeba* [46] and is downregulated by ABA stress in the roots of *Cajanus cajan* [47]. Therefore, exogenous light or hormones modulate the expression of *UGT* genes.

The expression levels of *SmUGT1*, *SmUGT9*, *SmUGT16*, and *SmUGT17* showed strong positive correlations with the contents of a variety of monomer saponins (Fig. 10), implicating these genes in saponin synthesis. In addition, several of those genes showed negative correlations with the contents of Saponin 30 and Saponin 32, possibly because of the synthesis of more complex triterpenoid saponins by glycosylation using these two monomeric saponins as receptor substrates.

Conclusions

This study was the first to comprehensively analyze the SmUGTs related to triterpenoid saponin synthesis in Sapindus mukorossi, and 42 such genes were identified. These genes were unevenly localized on 12 chromosomes, their lengths ranged from 450 to 1638 bp, and their numbers of amino acids from 149 to 545. Most SmUGTs were acidic and unstable, hydrophilic proteins, and the secondary structures were mainly *a*-helices and random coils. The 42 SmUGTs were divided into four subfamilies, which showed sequence conservation and functional differences among them. Many types of *cis*-acting elements were involved in development, light response, and plant hormone response. The expression patterns of SmUGT s were significantly different among organs and developmental stages, and most were expressed during stages S1 to S4 of fruit development. In addition, we predicted the functions of SmUGTs based on their phylogenetic relationships and expression patterns. Our findings provide insight into the differences and genetic relationships among SmUGTs and lay a theoretical foundation for the functional study of key genes related to the synthesis of triterpenoid saponins in soapberry. In future studies, we will investigate the functions of SmUGTs to construct cellular factories for the targeted and efficient production of soapberry triterpenoid saponins.

Materials and methods

Plant materials

Three soapberry trees (6.5 m average height, 13.5 cm average diameter at breast height [DBH]) were cultivated at Soapberry National Forest Germplasm Banks in Jianning County, Fujian Province, China ($26^{\circ}49'$ N latitude, $116^{\circ}52'$ E longitude, 300 m above sea level) [48]. We collected the roots, leaves, flowers and fruits samples between June and November 2018. Fruit samples were collected at eight growth stages: S1, early ovary development; 15 days after pollination (DAP); S2, 30% of largest fruit size, 45 DAP; S3, 70% of largest fruit size, 75 DAP;

S4, 80% of largest fruit size, 90 DAP; S5, 90% of largest fruit size, 105 DAP; S6, beginning of maturity, 120 DAP; S7, great change in pericarp, 135 DAP; and S8, fully developed and mature, 150 DAP. Three biological replicates were taken at each stage, for a total of 24 samples. Fruits were randomly picked from the east, south, west, and north sides of the middle and upper parts of the crowns of trees at 10 a.m. on sunny days. After the fruit had been picked, the pericarp was immediately separated from the seed. A portion of each pericarp sample was transferred to liquid nitrogen and stored at – 80 °C for RNA extraction [32] The representative herbarium of *Sapindus mukurossi* in this study is deposited in Chinese Virtual Herbarium, the voucher number is PE 01391570.

Identification and gene structural analysis

SmUGT genes were obtained from the Sapindus mukorossi reference genome annotation file (the complete genome of Sapindus mukorossi has been deposited in the National Genome Sciences Data Center (NGDC) database (https://ngdc.cncb.ac.cn). The whole genome GWH number WGS069104 can be accessed at PRJCA019364, and the whole genome was compared with the protein annotation information of Glycine max (https://www.soybase.org/sbt/) and Medicago truncatula (http://www.medicagogenome.org/) in NCBI (BLASTX; $e \le 1 \times 10^{-5}$). In all, 42 predicted *SmUGTs* were annotated. Gene structure maps of SmUGTs were generated using GSDS v2.0 (http://gsds.gao-lab.org/), and SmUGT s were mapped to soapberry chromosomes using TBtools v2.042. TBtools v2.042 was also used to perform collinearity analysis among soapberry, Arabidopsis thaliana, and Medicago truncatula. Pfam (http://pfam-legacy. xfam.org/) was used to predict conserved domains.

Conserved motifs and phylogenetic analysis

ProParam (http://web.expasy.org/protparam) was used to analyze the physicochemical properties of SmUGT proteins and WoLF PSORT (https://wolfpsort.hgc.jp/) was used to predict their subcellular localizations. We used MEME v5.5.5 (http://meme-suite.org/) to predict motifs in candidate SmUGT amino acid sequences. We used MEGA7 software with the Muscle program to perform multiple sequence alignments and neighbor-joining (NJ) analyses, with the bootstrap value set to 1000 replicates.

cis-acting element and protein structural analysis

PlantCARE was used to predict *cis*-acting elements in the 3000-bp upstream region (https://bioinformatics.psb.ugent. be/webtools/plantcare/html/). The secondary structures of SmUGTs were predicted using SoPMA (https://npsa-prabi. ibcp.fr/cgi-bin/npsa_automat.pl?page_npsa%20_sopma.

<u>html</u>), and SWISS-MODEL (https://swissmodel.expasy. org/) was used for tertiary structure prediction.

Expression patterns of SmUGT genes

We collected the vegetative organs (roots and leaves), reproductive organs (flowers and fruits), and peels from plants at different developmental stages and obtained high-quality FPKM transcriptome data [32]. The transcriptome data are available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database. The accession number is PRJNA784159. Transcriptomic data were normalized using TBtools v2.042 to create a map of gene expression patterns. Plant materials were stored at - 80 °C. Samples were ground into powder with liquid nitrogen; total RNA was extracted using an RNA Rapid Extraction Kit, then reverse-transcribed into cDNA. RNA RT-qPCR was performed to verify the expression of six randomly selected SmUGTs at various developmental stages. The correlations between expression of SmUGTs and expression of other genes encoding triterpenoid saponin synthesis pathway/related transcription factors were analyzed. Results with Pearson correlation coefficient (r) > 0.6/0.9were selected for mapping using Cytoscape software.

PCR was performed using a PerfectStart[™] Green qPCR SuperMix Kit (Beijing Quantype Gold AQ602) on a LightCycler 480II Fluorescence Quantitative PCR Instrument (Roche, Basel, Switzerland). Primers were designed using Primer 3 (https://bioinfo.ut.ee/prime r3-0.4.0/). *SmACT* was used as an internal reference [49], and expression levels were calculated using the 2^{-ΔΔCt} method. Correlations among expression profiles of six selected genes measured by RT-qPCR and RNA-seq were determined using origin v2019. Using the Omic-Share cloud platform (https://www.omicshare.com/ tools/Home/Soft/ica2), we analyzed the correlations of *SmUGT* expression levels with saponin contents in fruit at eight developmental stages.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-024-05281-4.

Supplementary Material 1. Supplementary Material 2. Supplementary Material 3. Supplementary Material 4. Supplementary Material 5. Supplementary Material 6.

Acknowledgements Not applicable.

Authors' contributions

Conceptualization, M.Z. and Z.C.; original draft preparation, M.Z., J.F., Y.G., C.Z. and Y.X.; reviewing and editing, M.Z., L.J., X.A. and Z.C.; methodology, investigation, formal analysis, M.Z., J.F., Y.G., C.Z. and Y.X.; supervision, L.J., X.A. and Z.C.; funding acquisition, L.J. and Z.C. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (No. 32071793 and No. 32371857), the National Key Research and Development Program of China (No. 2023YFD2201103) and the Special Foundation for National Science and Technology Basic Research Program of China (No. 2019FY100803).

Availability of data and materials

The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information files. Should any raw data files be needed in another format they are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

The plant materials used in this study were cultivated by our research group in the Soapberry National Forest Germplasm Banks (affiliated to the National Forestry and Grassland Administration). The collection of plant materials from Soapberry National Forest Germplasm Banks has been approved by the National Forestry and Grassland Administration.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests. Supplementary Information.

Received: 8 February 2024 Accepted: 10 June 2024 Published online: 21 June 2024

References

- Zheng Y, Liu J, Shi S, Jia L, Weng X, Luo S, Sheng K. Maturation processes and the dynamics of oil and saponin in Sapindus mukorossi. J Nanjing For Univ (Nat Sci Ed). 2021;45:76–82 (in Chinese).
- Liu J, Chen Z, Sun C, Wang L, He Q, Dai T, Yao N, Gao S, Zhao G, Shi S, Jia L, Weng X. Variation in fruit and seed properties and comprehensive assessment of germplasm resources of the genus Sapindus. Scientia Silvae Sinicae. 2019;55:44–54 (in Chinese).
- Xu Y, Jia L, Chen Z, Gao Y. Advances on triterpenoid saponin of Sapindus mukorossi. Chemistry. 2018;81:1078–88 (in Chinese).
- Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A. Triterpene biosynthesis in plants. Annu Rev Plant Biol. 2014;65:225–57.
- Xu Y, Chen Z, Jia L, Weng X. Advances in understanding of the biosynthetic pathway and regulatory mechanism of triterpenoid saponins in plants. SciSinVitae. 2021;51:525–55 (in Chinese).
- Augustin JM, Kuzina V, Andersen SB, Bak S. Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry. 2011;72:435–57.
- Kavya NM. A Review on Saponin Biosynthesis and its Transcriptomic Resources in Medicinal Plants. Plant Mol Biol Report. 2021;39:833–40.
- 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.
- Zhao Y, Li C. Biosynthesis of Plant Triterpenoid Saponins in Microbial Cell Factories. J Agric Food Chem. 2018;66:12155–65.
- 10. Vincken J-P, Heng L, de Groot A, Gruppen H. Saponins, classification and occurrence in the plant kingdom. Phytochemistry. 2007;68:275–97.

- Lu J, Li J, Wang S, Yao L, Liang W, Wang J, et al. Advances in ginsenoside biosynthesis and metabolic regulation. Biotechnol Appl Biochem. 2018;65:514–22.
- 12. Nasrollahi V. The effect of drought stress on the expression of key genes involved in the biosynthesis of triterpenoid saponins in liquorice (Glycyrrhiza glabra). Phytochemistry. 2014;103:32–7.
- 13. 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.
- 14. Sawai S, Saito K. Triterpenoid biosynthesis and engineering in plants. Front Plant Sci. 2011;30:25.
- Cárdenas PD. Evolution of Structural Diversity of Triterpenoids. Front Plant Sci. 2019;10:1523.
- Kai G, Zhou W, Wang Y, Hua Q, Shi M. Research progress in terms of the biosynthesis and regulation of terpenoids from medicinal plants. Sci Sin-Vitae. 2018;48:352–64 (in Chinese).
- 17. Paquette S, Møller BL, Bak S. On the origin of family 1 plant glycosyltransferases. Phytochemistry. 2003;62:399–413.
- Li Y, Baldauf S, Lim E-K, Bowles DJ. Phylogenetic Analysis of the UDPglycosyltransferase Multigene Family of Arabidopsis thaliana. J Biol Chem. 2001;276:4338–43.
- Mamoon Rehman H, Amjad Nawaz M, Bao L, Hussain Shah Z, Lee J-M, Ahmad MQ, et al. Genome-wide analysis of Family-1 UDP-glycosyltransferases in soybean confirms their abundance and varied expression during seed development. J Plant Physiol. 2016;206:87–97.
- 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.
- Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, et al. The genome of the domesticated apple (Malus × domestica Borkh.). Nat Genet. 2010;42:833–9.
- Wei W, Wang P, Wei Y, Liu Q, Yang C, Zhao G, et al. Characterization of Panax ginseng UDP-Glycosyltransferases Catalyzing Protopanaxatriol and Biosyntheses of Bioactive Ginsenosides F1 and Rh1 in Metabolically Engineered Yeasts. Mol Plant. 2015;8:1412–24.
- Lu C, Zhao S-J, Wang X-S. Functional regulation of a UDP-glucosyltransferase gene (Pq3-O-UGT1) by RNA interference and overexpression in Panax quinquefolius. Plant Cell Tiss Organ Cult. 2017;129:445–56.
- Lu C, Zhao S, Wei G, Zhao H, Qu Q. Functional regulation of ginsenoside biosynthesis by RNA interferences of a UDP-glycosyltransferase gene in Panax ginseng and Panax quinquefolius. Plant Physiol Biochem. 2017;111:67–76.
- 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 in Saponin-Mediated Insect Resistance. Plant Physiol. 2012;160:1881–95.
- Achnine L, Huhman DV, Farag MA, Sumner LW, Blount JW, Dixon RA. Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume Medicago truncatula: Biosynthesis of saponins. Plant J. 2005;41:875–87.
- 27. de Costa F, Barber CJS, Kim Y, Reed DW, Zhang H, Fett-Neto AG, et al. Molecular cloning of an ester-forming triterpenoid: UDP-glucose 28- O -glucosyltransferase involved in saponin biosynthesis from the medicinal plant Centella asiatica. Plant Sci. 2017;262:9–17.
- Zhong Y, Xue X, Liu Z, Ma Y, Zeng K, Han L, et al. Developmentally Regulated Glucosylation of Bitter Triterpenoid in Cucumber by the UDP-Glucosyltransferase UGT73AM3. Mol Plant. 2017;10:1000–3.
- Zhang J, Jia L. Advances in pharmacological effects of Sapindus. Modern J Integr Traditional Chin Western Med. 2020;29:1922–5+1929 (in Chinese).
- Peng Y, Fu C, Wang Y, Wu Z. Advances in pharmacological effects of Sapindus, Modern Journal of Integrated Traditional Chinese and Western Medicine. Hubei Agric Sci. 2022;61:164–8 (in Chinese).
- Dai Z, Wang J, Ma X, Sun J, Tang F. Laboratory and Field Evaluation of the Phytotoxic Activity of Sapindus mukorossi Gaertn Pulp Extract and Identification of a Phytotoxic Substance. Molecules. 2021;26:1318.
- Xu Y, Zhao G, Ji X, Liu J, Zhao T, Gao Y, et al. Metabolome and Transcriptome Analysis Reveals the Transcriptional Regulatory Mechanism of Triterpenoid Saponin Biosynthesis in Soapberry (Sapindus mukorossi Gaertn.). J Agric Food Chem. 2022;70:7095–7109.

- Rahimi S, Kim J, Mijakovic I, Jung K-H, Choi G, Kim S-C, et al. Triterpenoidbiosynthetic UDP-glycosyltransferases from plants. Biotechnol Adv. 2019;37: 107394.
- Gao Y, Yang X, Yang X, Zhao T, An X, Chen Z. Characterization and expression pattern of the trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase gene families in Populus. Int J Biol Macromol. 2021;187:9–23.
- Paquette SM, Jensen K, Bak S. A web-based resource for the Arabidopsis P450, cytochromes b5, NADPH-cytochrome P450 reductases, and family 1 glycosyltransferases (http://www.P450.kvl.dk). Phytochemistry. 2009;70:1940–7.
- Liu P, Zhang S, Shen X, Li X, Zhao Y, Zhang Z, et al. Genome-wide identification, function prediction, and expression pattern analysis of the UGT family in Dimocarpus longan Lour. Chinese Journal Applied Environmental Biology. 2021;27:1626–35 (in Chinese).
- Lu Z, Liu H, Zhang G, Yu L, Luo H, He C. Genome-wide Identification, Characterization, and Expression Analysis of UGT Gene Family Members in Sea Buckthorn (Hippophae rhamnoides L.). Forest Reaearch. 2021;34:9–19 (in Chinese).
- Qin J, Sun C, Zhang M, Wang Y. Classification, Function and Evolution of Plant UDP-glycosyltransferase. Genomics and Applied Biology. 2018;37:440–50 (in Chinese).
- Caputi L, Lim E-K, Bowles DJ. Discovery of New Biocatalysts for the Glycosylation of Terpenoid Scaffolds. Chem Eur J. 2008;14:6656–62.
- Yang X, Li J, Guo T, Guo B, Chen Z, An X. Comprehensive analysis of the R2R3-MYB transcription factor gene family in Populus trichocarpa. Ind Crops Prod. 2021;168: 113614.
- Jung S-C, Kim W, Park SC, Jeong J, Park MK, Lim S, et al. Two Ginseng UDP-Glycosyltransferases Synthesize Ginsenoside Rg3 and Rd. Plant Cell Physiol. 2014;55:2177–88.
- Wu B, Gao L, Gao J, Xu Y, Liu H, Cao X, et al. Genome-Wide Identification, Expression Patterns, and Functional Analysis of UDP Glycosyltransferase Family in Peach (Prunus persica L. Batsch). Front Plant Sci. 2017;8:389.
- Liu Q, Gu Z, Fu C, Yang C. Cloning and expression pattern analysis of UGT gene family UGT86A1 and UGT85A2 in Gardenia jasminoides. J Central South University Forest Technol. 2021;41:173–82 (in Chinese).
- 44. Yao Y, Gu J, Luo Y, Wang Y, Pang Y, Shen G, et al. Genome-wide analysis of UGT gene family identified key gene for the biosynthesis of bioactive flavonol glycosides in Epimedium pubescens Maxim. Synth Syst Biotechnol. 2022;7:1095–107.
- Kang KB, Jayakodi M, Lee YS, Nguyen VB, Park H-S, Koo HJ, et al. Identification of candidate UDP-glycosyltransferases involved in protopanaxadioltype ginsenoside biosynthesis in Panax ginseng. Sci Rep. 2018;8:11744.
- 46. Wang M, Gao M, Zhao Y, Chen Y, Wu L, Yin H, et al. LcERF19, an AP2/ERF transcription factor from Litsea cubeba, positively regulates geranial and neral biosynthesis. Horticulture Research. 2022;9:uhac093.
- Song Z. Genome-wide identification and characterization of UGT family in pigeonpea (Cajanus cajan) and expression analysis in abiotic stress. Trees. 2019;33:987–1002.
- Xu Y, Gao Y, Chen Z, Zhao G, Liu J, Wang X, et al. Metabolomics analysis of the soapberry (Sapindus mukorossi Gaertn.) pericarp during fruit development and ripening based on UHPLC-HRMS. Sci Rep. 2021;11:11657.
- Xu Y, Zhao G, Hao Y, Weng X, Chen Z, Jia L. Reference Genes Selection and Validation for RT-qPCR in Sapindus mukorossi. Biotechnol Bull. 2022;38:80–9 (in Chinese).

Publisher's Note

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