Available online at www.sciencedirect.com

ScienceDirect

Chinese Journal of Natural Medicines 2019, **17**(2): 0131–0144

Chinese Journal of Natural Medicines

Transcriptomic analysis in *Anemone flaccida* rhizomes reveals ancillary pathway for triterpene saponins biosynthesis and differential responsiveness to phytohormones

MO Guo-Yan [1], HUANG Fang [1], FANG Yin [1], HAN Lin-Tao [1*], Kayla K. Pennerman [2], BU Li-Jing [3], DU Xiao-Wei [4], Joan W. Bennett [2], YIN Guo-Hua [2*]

[1] *China Key Laboratory of TCM Resource and Prescription, Ministry of Education, Hubei University of Chinese Medicine, Wuhan 430065, China;*

[2] *Department of Plant Biology and Pathology, Rutgers, The State University of New Jersey, New Jersey 08901, USA;*

[3] *Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA;*

[4] *Department of Pharmacy, Heilongjiang University of Chinese Medicine, Harbin 150040, China*

Available online 20 Feb., 2019

[ABSTRACT] *Anemone flaccida* Fr. Schmidt is a perennial medicinal herb that contains pentacyclic triterpenoid saponins as the major bioactive constituents. In China, the rhizomes are used as treatments for a variety of ailments including arthritis. However, yields of the saponins are low, and little is known about the plant's genetic background or phytohormonal responsiveness. Using one-quarter of the 454 pyrosequencing information from the Roche GS FLX Titanium platform, we performed a transcriptomic analysis to identify 157 genes putatively encoding 26 enzymes involved in the synthesis of the bioactive compounds. It was revealed that there are two biosynthetic pathways of triterpene saponins in *A. flaccida*. One pathway depends on *β*-amyrin synthase and is similar to that found in other plants. The second, subsidiary ("backburner") pathway is catalyzed by camelliol C synthase and yields *β*-amyrin as minor byproduct. Both pathways used cytochrome P450-dependent monooxygenases (*CYPs*) and family 1 uridine diphosphate glycosyltransferases (*UGTs*) to modify the triterpenoid backbone. The expression of *CYPs* and *UGTs* were quite different in roots treated with the phytohormones methyl jasmonate, salicylic acid and indole-3-acetic acid. This study provides the first large-scale transcriptional dataset for the biosynthetic pathways of triterpene saponins and their phytohormonal responsiveness in the genus *Anemone*.

[KEY WORDS] *Anemone flaccida* Fr. Schmidt; Triterpenoid saponins; Biosynthetic pathways; Transcriptomic analysis; Phytohormonal responsiveness

[CLC Number] R965 **[Document code]** A **[Article ID]** 2095-6975(2019)02-0131-14

Introduction

Anemone flaccida Fr. Schmidt, belonging to the Ranunculaceae family, is a perennial medicinal plant. The genus consists of about 150 species distributed in the temperate areas of both Northern and Southern hemispheres [1]. China has 53 species, 9 subspecies and 36 varieties which are naturally distributed in all provinces except Guangdong and Hainan. A field survey performed by Xiao PG *et al*. indicates that there are at least 38 species/varieties being used ethnopharmacologically in China, including *Anemone raddeana*, *Anemone rivularis*, *Anemone davidii*, and *Anemone begonifolia* [1]. Traditionally, *A. flaccida*'s dried roots, or rhizomes, (called 'é zhǎng cǎo', 'dì wū', 'wú gōng sān qī' or 'èr lún qī' in Chinese) are used as painkillers, antipyretics, detoxifiers and immunosuppressors, and the plant also is considered as a valuable Chinese medicine for treating punch injury and rheumatoid arthritis [2-3]. Natural products identified from *Anemone* spp. include triterpenoids, saponins, steroids, saccharides, flavonoids, aetheroleas, and organic acids [4-6]. Previous phytochemical studies show that the pentacyclic triterpenoid saponins, especially oleanane-type oleanolic acid glycosides and glycoside ivy ligands, are very abundant in this genus [7-10].

[Received on] 29-Sep.-2018

[Research funding] This work was supported by the National Natural Science Foundation of China (No. 31670334), the Science and Technology Innovation Team Project of Hubei Provincial Department of Education for Young and Middle-aged Scientists (No. T201608).

[*Corresponding author] E-mails: hanlt711.metabonomics@gmail.com (HAN Lin-Tao); guohuayin1997@gmail.com (YIN Guo-Hua)

These authors have no conflict of interest to declare.

Five triterpenoid saponins have been characterized from the rhizomes of *A*. *flaccida* [11]. These saponins share a common oleanane core that is conjugated with one or more sugar chains [12]. Similar glycoconjugated oleanane-type triterpene saponins have been isolated from more than one thousand plant species, including important food and medicinal plants such as *Panax ginseng*, *Panax quinquefolius*, *Glycyrrhiza uralensis*, *Polygonum cuspidatum* and *Bupleurum chinense*. The oleanane-type triterpene saponins extracted from *A*. *flaccida* exhibit multiple desirable biological activities, including anti-tumor [11, 13-14], anti-inflammatory [15], and anti-rheumatic [3, 16-19]. Moreover, total saponins of the rhizomes of *A*. *flaccida* are undergoing phase III clinical trials in China for treatment of rheumatoid arthritis (RA) [15]. However, the saponins accumulate at very low concentrations in *Anemone* rhizomes, posing a serious obstacle to obtaining sufficient amounts for medical application [20].

In order to improve the yields of triterpene saponins in medicinal plants, it is necessary to understand the relevant biosynthetic pathways. In this study, we extracted total RNAs of *A*. *flaccida* to perform 454 pyrosequencing and transcript annotation. By *de novo* transcriptome assembly, we identified a series of putative genes involved in traditional and back-burner biosynthetic pathways of triterpene saponins. Through qRT-PCR detection, we observed the different expression levels of the *CYPs* and *UGTs* involved in the triterpene saponins biosynthetic pathways in roots of *A*. *flaccida* treated with methyl jasmonate, salicyclic acid and indole-3-acetic acid. These data will be very useful for future transcriptional regulation and functional studies, and thereby provide mechanistic guidance of increasing yields of these important bioactive natural products.

Materials and Methods

Plant material

Five-year-old *A*. *flaccida* roots were harvested from plants cultivated for medical purposes from the fields in Enshi, Hubei Province, China. After washing and cleaning, samples of the root tissues were cut into small pieces, immediately frozen in liquid nitrogen and stored at −80°C until further use.

RNA preparation and cDNA synthesis

Total RNA was extracted from the root samples using Trizol reagent according to the manufacturer's instructions and then treated with RQ1 RNase-free DNase to reduce residual genomic DNA. The concentration and purity of total RNA were evaluated by electrophoresis on 1% agarose gels and measured using a NanoDrop 2000/2000c spectrophotometer. The cDNA was synthesized using PrimeScript RT Reagent Kit (Takara, Japan) according to the manufacturer's instructions. For reverse transcription, the reaction consisted of the following ingredients: 0.5 μg of total RNA, 0.5 μL of 50 μmol·L^{-1} oligo dT, 0.5 μL of 100 μmol·L^{-1} random hexamers, 2 μL of 5X PrimeScript Buffer, 0.5 μL of PrimeScript RT Enzyme Mix I. Nuclease-free water was added to a final volume of 10 μL. Reactions were performed in a GeneAmp PCR System 9700 at 37 °C for 15 min, followed by heat inactivation of reverse transcriptase at 85 °C for 5 min. Nuclease-free water was added to the finished reaction mix to a final volume of 100 μL and stored at −20 °C.

Library construction

Five microliters of reverse transcription products were subjected to electrophoresis on 1% agarose gels to determine the reaction efficiency and quality. The ds cDNAs were purified using the PureLink PCR Purification Kit (Life Technologies, USA). The purified cDNAs were then digested with *Bsg*I overnight at 37 °C and recovered with QIAquick PCR Purification Kit (Qiagen, Germany). Finally, about 10 μg of ds cDNAs was used for pyrosequencing with GS FLX Titanium Rapid Library Preparation Kit (Roche, Germany) following the manufacturer's instructions.

The 454-EST assembly and isotigs annotation

The 454 raw reads sequences were subjected to quality control, which included trimming the adapters and poly (A/T) tails, removing short sequences (< 50 bp) and low quality bases were trimmed (quality score threshold of 20) using SeqClean (version Lastest 86_64) and Lucy (version 1.20p). The trimmed reads were assembled into unique sequences (including isotigs and singlets) using Newbler 2.6 software with the default parameters. Functional annotations were initially performed using NCBI blastx algorithm. The assembled unique sequences then were searched for sequence similarities against the NCBI non-redundant nucleotide database with the blastn algorithm with an *e*-value cut-off of 10^{-5} so that the ribosomal sequences were found and removed [21]. The remaining sequences were searched against the following public databases: UniProt, SwissProt and trEMBL (http://uniprot.org/uniprot/), KEGG (http://genome.jp/kegg/), NCBI non-redundant protein (http://ncbi.nlm.nih.gov), Clusters of Orthologous Groups (http://ncbi.nlm.nih.gov/COG/), Pfam (http://pfam.xfam.org/) and Conserved Domains Database (http://ncbi.nlm.nih.gov/cdd). The GO database (http://geneontology.org) was used to analyze the functional categories of unique sequences. The assembled unique genes were categorized according to GO terms based on similarities with the protein sequences of *Arabidopsis thaliana* using the *Arabidopsis* Information Resource (http: //arabidopsis.org/).

The KEGG database, containing metabolic pathways that represent molecular interactions and reaction networks [22], was used to annotate pathways in which unique genes are involved [23]. Unique sequences were assigned EC numbers after blastx searches against the KEGG database. Biochemical pathways for unique sequences were identified according to the corresponding EC distribution in the KEGG database.

Candidate genes involved in triterpene saponins biosynthesis

Based on known biosynthesis pathways of triterpene saponins, the candidate genes could be classed as *HGMR*, *IPPI*, *FPPS*, *SS*, *SE*, *OSC*, *P450* and *UGT*. According to their gene names and synonyms, they were searched in the annotated unique genes. Repeated items from different annotation

databases were manually erased.

Primer design and qRT-PCR

To evaluate our transcriptomic analysis and screen for *CYPs* and *UGTs* that may be involved in biosynthesis of triterpene saponins, the qRT-PCR was performed to analyze all isotigs annotated as *CYPs* and *UGTs* in *A*. *flaccida* roots treated with one of three concentrations of three plant hormones (methyl jasmonate, MeJA; salicylic acid, SA; indole-3-acetic acid, IAA): 0.05 mmol·L^{-1} (MeJA1, SA1, and IAA1), 0.5 mmol·L^{-1} (MeJA2, SA2, and IAA2) and 5 mmol·L^{-1} (MeJA3, SA3, and IAA3). The primers were designed using Roche LCPDS2 software (http://lightcycler-probe-design-software.software.informer.com/2.0/) according to the transcriptome sequences in this study and synthesized at Shanghai Generay Biotech (Shanghai, China). The primers for qRT-PCR were listed in Table 2. Quantitative PCR was performed using LightCycler 480 SYBR Green I Master kit and a LightCycler 480 System (Roche, Germany). The reaction mixture was as follows: 5 μL of 2X LightCycler 480 SYBR Green I Master, 0.2 μL of 10 μmol·L^{-1} forward/reverse primers each, 1 μL of cDNA and 3.6 μL of nuclease-free water. Reactions were incubated in a 384-well optical plate at 95 °C for 10 min, followed by 40 cycles of 95 °C for 10 s and 60 °C for 30 s. Each sample was assayed in triplicate for analyses. At the end of the PCR cycles, melting curve analyses were performed to evaluate the specific PCR products. The melting curve was obtained by increasing the temperature slowly from 60 °C to 97 °C and continuously performing five acquisitions per degree Celsius. The expression levels of the mRNAs were normalized to the 18S rRNA reference gene and were calculated using the $2^{-\Delta\Delta CT}$ method [24-26].

Results

Transcriptome sequencing, sequence assembly and transcript annotation

A cDNA library of five-year-old *A*. *flaccida* roots with a total of 234,551 high-quality reads with an average length of 335 bp was produced. After trimming adaptor sequences and removing reads shorter than 50 bp, 20 3146 reads were obtained, which were assembled into 3 0494 unique sequences. The lengths of these unique sequences were sufficient to enable annotations with high accuracy [27]. Unique sequences of *A*. *flaccida* were annotated using public databases. Only half (50.28%) of *A*. *flaccida* unique sequences could be annotated the remaining (49.72%) unique sequences had no match to any sequences in the databases.

GO and KEGG annotation

Based on sequence homology, 3 0494 isotig sequences were categorized into 56 functional groups: 27 in biological processes, 16 in cellular components, and 13 in molecular functions (Fig. 1). Using KEGG annotation to identify the

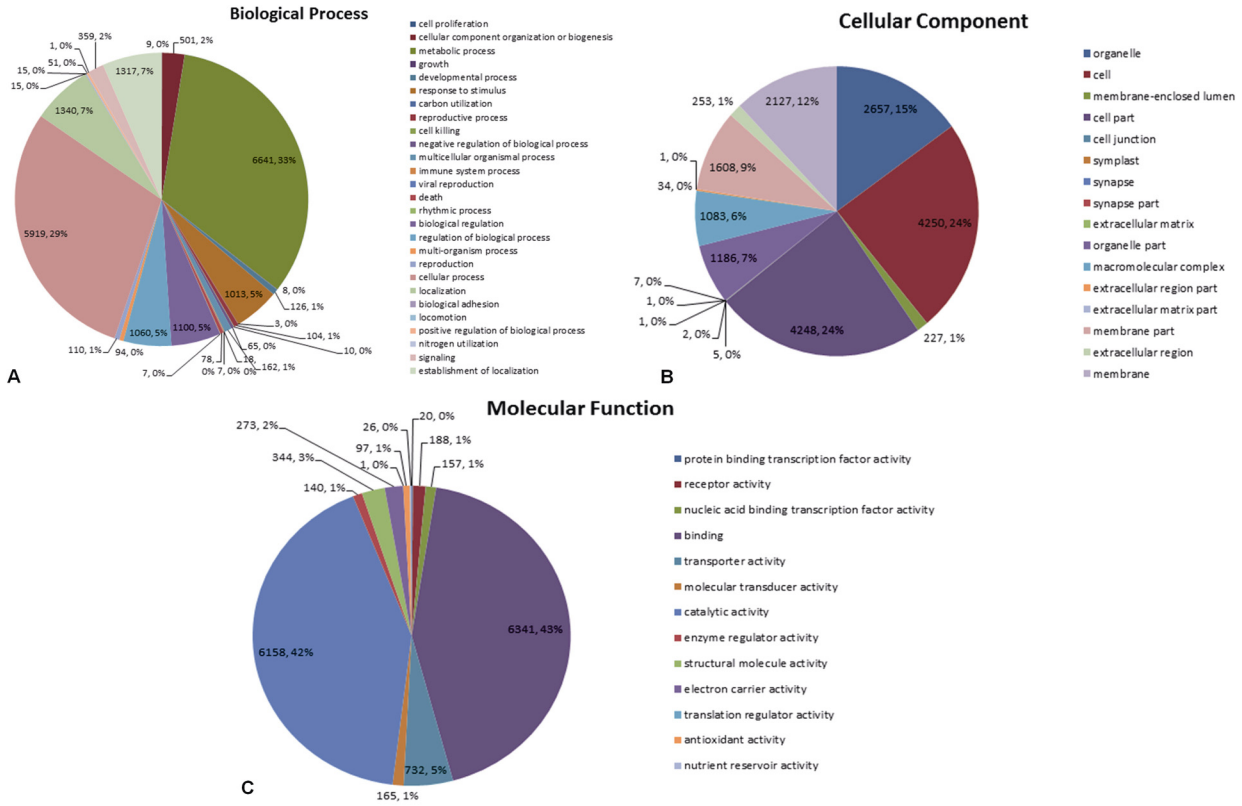

Fig. 1 GO analysis of unique sequences based on (A) biological process, (B) cellular component and (C) molecular function. The labels indicate the number and percentage of genes in each category

most active biological pathways, 1 1572 unique sequences were annotated to 299 KEGG pathways. There were 1 8970 unique sequences not assigned to any known biochemical pathway. The most represented pathways were "metabolic pathways" (5084 isotigs), "genetic information processing" (1803 isotigs), and "cellular processes" (1153 isotigs). The top three were phenylpropanoid biosynthesis (108), stilbenoid, diarylheptanoid and gingerol biosynthesis (72), and flavonoid biosynthesis (26).

Candidate genes related to triterpene saponins backbone biosynthesis

Based on the current knowledge of biosynthesis of all terpenoids found in nature, isopentenyl pyrophosphate (IPP) and its allylic isomer dimethylallyl diphosphate (DMAPP) are universal donors of C5-units for triterpene saponins backbone biosynthesis [28]. There are two independent routes for IPP synthesis in a plant cell: one is the mevalonate (MVA) pathway operating in the cytosol, the other is the non-mevalonate pathway in plastids, also called the 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway [29-30]. In our analysis, almost all of the known enzymes of triterpene saponins biosynthesis were discovered in the *A*. *flaccida* 454-EST dataset (Table 1), indicating that both the MVA pathway and the MEP pathway are active in the roots.

In the MVA pathway, the conversion of HMG-CoA into MVA is an irreversible, two-step biochemical reaction that is catalyzed by hydroxymethylglutaryl-CoA reductase (*HMGR*) (EC 1.1.1.34). Usually, *HMGR* is considered as an important rate-limiting enzyme in the MVA pathway [31-33]. We found transcripts of seven homologous *HMGR* genes, indicating that *HMGR* might be encoded by multiple genes in *A*. *flaccida*. Interestingly, only one isotig encoding PMK involved in the MVA pathway was found in this dataset.

The MEP pathway is commonly considered to synthesize monoterpenes and diterpenes [28]. In the transcriptome of *A*. *flaccida*, 16 isotigs were found to encode enzymes involved in this pathway. Among these enzymes, DXSs (EC 2.2.1.7) are responsible for the formation of 1-deoxy-D-xylulose 5-phosphate through the condensation of pyruvate and glyceraldehydes-3-phosphate.

In the putative triterpene biosynthetic pathways in *A*. *flaccida*, the C5 unit IPP produced by MVA and MEP pathways are assembled into geranyl pyrophosphate (GPP, C10), farnesyl diphosphate (FPP, C15), and geranylgeranyl diphosphate (GGPP) by a series of prenyl transferases. Subsequently, two units of FPP join in a "tail-to-tail" fashion to form squalene (C30) catalyzed by squalene synthase (SS). The squalene is oxidized by squalene epoxidase (SE) to yield 2,3-oxidosqualene (C30), which is considered as another important precursor of triterpenoid saponins, phytosterols, and steroidal saponins [34-36]. The IPP is often transformed into its isomer, dimethylallyl diphosphate (DMAPP, C5), by isopentenyl diphosphate isomerase (IPPI) before it is assembled. In the *A*. *flaccida* transcriptome, there were 1 IPPI, 2 GPPS, 2 FPPS, 3 GGPPS, 2 SS, and 4 SE isotigs.

Candidate genes related to biosynthesis of triterpenoid aglycones

The pentacyclic triterpenoid saponins, especially oleanane-type oleanolic acid glycosides, are characteristic of *Anemone*. The typical structure of an oleanolic acid glycoside is shown in Fig. 2. To date, over 60 triterpenoid saponins have been isolated from *Anemone* spp. As previously reported, the 2,3-oxidosqualene is cyclized to various tetra- or pentacyclic triterpene skeletons, such as phytosterol, dammarane, lupane or oleane (*β*-amyrin), *via* protonation and epoxide ring opening by a family of 2, 3-oxidosqualene cyclases [37-38]. This cyclization of 2,3-oxidosqualene is the first committed step and is a critical branch-point for the biosynthesis of phytosterol and triterpenoid biosynthesis.

Enzymes involved in the cyclization process are called oxidosqualene cyclases (OSCs, EC 5.4.99.x). Many different kinds of OSCs, such as cycloartenol synthase, lanosterol synthase, and *β*-amyrin synthase, have been cloned from various plant species [34] and their mechanisms of action also have been well-documented [29, 39]. Nearly 200 different triterpene skeletons have been found from natural sources or enzymatic reactions. The molecular diversity and the proposed mechanisms of formation of triterpene skeletons have been comprehensively reviewed [38]. However, only a limited number of possible cyclization products are involved in saponins biosynthesis [36]. Among the OSCs, cycloartenol synthase (CAS, EC 5.4.99.8) catalyzes the formation of the tetracyclic plant sterol precursor cycloartenol from 2,3-oxidosqualene [37, 40] and *β*-amyrin synthase (BAS, EC 5.4.99.39) catalyzes 2,3-oxidosqualene to yield the pentacyclic oleanane-type triterpenoid backbone *β*-amyrin [37, 41]. In this study, 3 lanosterol synthase isotigs, 4 CAS isotigs, 1 CAMS1 isotig and 7 BAS isotigs were identified (Table 1). BAS is generally regarded as the key enzyme for the formation of *β*-amyrin, which CAMS1 only generates with a yield of 0.2% [42]. It seems that there are two pathways to form oleanane-type triterpenoid backbone *β*-amyrin in *A*. *flaccida*. The traditional pathway, depending on BAS, is same as other plants. The second, ancillary pathway makes *β*-amyrin as minor byproduct and thus functions as "a backburner pathway".

After the basal triterpenoid aglycones backbone structure is formed by OSCs, these cyclization products undergo various modifications prior to glycosylation. Such modifications, usually designated as "decorating" or "tailoring" steps, introduce various functional groups at different positions of the aglycone skeletons, thereby greatly increasing the structural diversity of saponins. The three types of sapogenin decoration are: substitution, bridging, and unsaturation. The substituents are usually small functional groups, including hydroxyl- (-OH), keto- (=O), aldehyde- (-CHO), methyl- ($-CH_3$), hydroxymethyl- ($-CH_2OH$), and carboxyl- (-COOH) moieties. For the oleanane-type triterpenoid backbone, multiple

Table 1 The expressed genes involved in triterpene saponins biosynthesis in *A. flaccida* roots

Pathway	Gene name	Abbreviation	EC	No.	Unique sequences
MVA	Acetyl-CoA acetyltransferase	AACT	2.3.1.9	5	isotig01627|isotig10165|isotig10619 |isotig11282|HSAPGDX01CPQCB
	Hydroxymethylglutaryl-CoA synthase	HMGS	2.3.3.10	1	HSAPGDX01C33HZ
	HMG-CoA reductase	HMGR	1.1.1.34	7	isotig00332|isotig02401|isotig03220|isotig04960|isotig06119|HSAPGDX01CDZHD|HSAPGDX01CZJU4
	Mevalonate kinase	MK	2.7.1.36	2	isotig01599|isotig10206
	Phosphomevalonate kinase	PMK	2.7.4.2	1	CL30343Contig1
	Mevalonate diphosphate decarboxylase	MVD	4.1.1.33	2	isotig08075|isotig00642
MEP	1-deoxy-D-xylulose-5-phosphate synthase	DXS	2.2.1.7	4	isotig00051|isotig02395|isotig04358|isotig10331
	1-deoxy-D-xylulose-5-phosphate reductoisomerase	DXR	1.1.1.267	1	isotig00433
	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	MCT	2.7.7.60	2	isotig03481|HSAPGDX01A8J5Q
	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	CMK	2.7.1.148	2	isotig03769|isotig14582
	2-C-methyl-D-erythritol 2, 4-cyclodiphosphate synthase	MDS	4.6.1.12	2	isotig04021|isotig06696
	4-hydroxy-3-methylbut-2-enyl diphosphate synthase	HDS	1.17.7.1	4	isotig00386|isotig04667|isotig15675|HSAPGDX01D30J5
	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	HDR	1.17.1.2	1	isotig00104
	Isopentenyl diphosphate isomerase	IPPI	5.3.3.2	1	isotig03333
	Squalene synthase	SS	2.5.1.21	2	isotig01999|HSAPGDX01DLCQ9
	Squalene epoxidase	SE	1.14.13.132	4	isotig04487|isotig08220|isotig13034|HSAPGDX01D36DA
Monoterpene	Geranyl diphosphate synthase	GPPS	2.5.1.1	2	HSAPGDX01BYSVI|isotig10187
Sesquiterpene	Farnesyl diphosphate synthase	FPPS	2.5.1.10	2	isotig08759|HSAPGDX01A2PLF
Diterpene	Geranylgeranyl diphosphate synthase	GGPPS	2.5.1.29	3	isotig08873|isotig09209|isotig00421
	Lanosterol synthase		5.4.99.7	3	HSAPGDX01DYPSC/HSAPGDX01CBKH8|HSAPGDX01BW8FQ
	Cucurbitadienol synthase	CAS	5.4.99.8	4	isotig05594|isotig13352|HSAPGDX01AK7X2|HSAPGDX01D3UOR
	beta-amyrin synthase	BAS	5.4.99.39	7	CL20139Contig1|CL28422Contig1|CL34399Contig1|CL41511Contig1|CL5336Contig1|CL7082Contig1|CL7877Contig1
	Camelliol C synthase	CAMS1	5.4.99.38	1	isotig14954
	Cytochrome P450	P450	1.14.x.x	40	isotig02871|isotig03344|isotig04974|isotig05047|isotig05706|isotig06020|isotig06965|isotig07049|isotig08321|isotig08843|isotig09353|isotig09684|isotig09723|isotig10101|isotig11233|isotig11595|isotig11847|isotig11862|isotig12183|isotig12820|isotig12824|isotig12981|isotig15430|HSAPGDX01DEGI8|HSAPGDX01AV38C|HSAPGDX01BU52G|HSAPGDX01D3WVD|HSAPGDX01DLAL8|HSAPGDX01CNHZ2|HSAPGDX01EAEL1|HSAPGDX01CLXCM|HSAPGDX01EOTHB|HSAPGDX01ELSC1|HSAPGDX01ELR98|HSAPGDX01A8CC4|HSAPGDX01BQ3UJ|HSAPGDX01BB98P|HSAPGDX01AX7JH|HSAPGDX01B3O1K|HSAPGDX01BK7D9
	Glycosyltransferases	GT	2.4.1.x	58	isotig00021|isotig00192|isotig00569|isotig00626|isotig00723|isotig00779|isotig00974|isotig01355|isotig01770|isotig02200|isotig02323|isotig02391|isotig03013|isotig03161|isotig03657|isotig03905|isotig04308|isotig04324|isotig04625|isotig04637|isotig04876|isotig05225|isotig06483|isotig06722|isotig07307|isotig07731|isotig09627|isotig09875|isotig10670|isotig10675|isotig10706|isotig11128|isotig11478|isotig12070|isotig12873|isotig13150|isotig13264|isotig13334|isotig14347|isotig15280|isotig16016|HSAPGDX01DVS7I|HSAPGDX01C65EF|HSAPGDX01C4E7C|HSAPGDX01DG8S4|HSAPGDX01CTQUJ|HSAPGDX01C2MJK|HSAPGDX01DQYRK|HSAPGDX01BBMIJ|HSAPGDX01B4J7O|HSAPGDX01BYVLC|HSAPGDX01C4E7C|HSAPGDX01EID9G|HSAPGDX01BHU0E|HSAPGDX01EF1RQ|HSAPGDX01CK9O7|HSAPGDX01CBAB2|HSAPGDX01C6OA2|HSAPGDX01CI6EX
	UDP-glucosyl transferase	UGT		17	isotig00569|isotig00626|isotig00779|isotig01355|isotig02200|isotig03161|isotig04308|isotig07731|isotig09627|isotig09875|isotig10675|HSAPGDX01DVS7I|HSAPGDX01C65EF|HSAPGDX01C4E7C|HSAPGDX01BYVLC|HSAPGDX01B4J7O|HSAPGDX01C6OA2

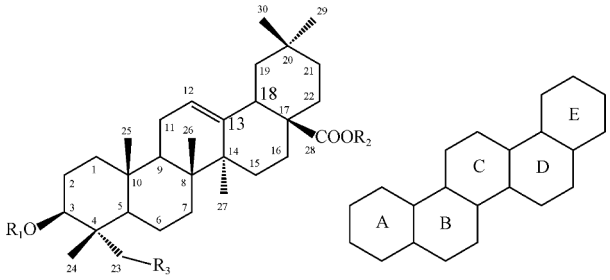

Fig. 2 The chemical structure and formulas of oleanolic acid glycosides

decorations at various positions create a large structural diversity [36]. The type and extent of modifications greatly vary among different saponins-containing plant species [43]. Furthermore, decoration of similar cyclization products also differs between organs and tissues of an individual plant [44], as well as between intact plant and derived cell cultures [45]. Therefore, regulation of sapogenin decoration is a possible mechanism to maintain specialized pools of tailored saponins.

Most sapogenin decorations are catalyzed by cytochrome P450-dependent monooxygenases (CYPs, EC 1.14.x.x) [46]. As one of the largest and most diverse enzyme families in plants, many CYPs capable of decorating the *β*-amyrin skeleton have been identified in dicotyledonous plants. Among them, only CYP51H10 has been identified in monocots. Other CYPs have multiple catalytic activities in dicots for different C position of the *β*-amyrin backbone. In this study, 40 isotigs were annotated as CYPs (Table 1).

Candidate genes related to the formation of sugar chains

After being decorated with hydroxy- and carboxy-moieties by CYPs, the decorated sapogenin backbone usually undergoes glycosylation. Glycosylation reactions of natural products are catalyzed by family 1 uridine diphosphate glycosyltransferases (UGTs, EC 2.4.1.x) [47], which introduces saccharide side chains to the decorated sapogenin skeleton for modulation of saponins stability, biological activity, solubility, and signaling for storage or intra- and intercellular transport [48]. Glycosylation of saponins is considered as one of the major factors in determining the bioactivity and bioavailability of natural plant products, such as flavonoids and terpenoids [49]. The number and type of saccharide chains vary enormously from oligomeric sugar chains to 2−5 monosaccharide units, or to 1−2 monosaccharide units. Moreover, different saccharide chains have their own preference to the position of glycosylation on the skeleton [48]. In the oleanane-type saponins, the saccharide chains commonly involve glucose, arabinose, rhamnose, xylose, and glucuronic acid. In addition, there are some rare monosaccharide residues including apiose, fucose, quinovose and ribose [36].

UGTs comprise the largest glycosyltransferase (GT) superfamily in plants with 92 known families and also some non-classified sequences at the superfamily level (www.cazy.org/GlycosylTransferases.html) [50]. *UGTs* specifically utilize UDP-sugars as donors and then transfer various sugars to plant metabolites [47, 51]. Generally, they are localized in the cytosol and are involved in the biosynthesis of plant natural products such as flavonoids, phenylpropanoids, terpenoids and steroids [52-53]. A few reported *UGTs* are involved in glycosylation of triterpenoid sapogenins, such as *UGT71G1*, *UGT73K1*, *UGT74M1*, *UGT73F3*, *UGT73P2*, and *UGT91H4* [54-57]. In our 454 transcript dataset, 58 unique sequences were found to encode GTs.

The qRT-PCR analysis of CYPs and UGTs in phytohormone-treated root of A. flaccida

Methyl jasmonate (MeJA) is an important inducer of plant secondary metabolites [58-59]. MeJA, IAA, and SA also are key phytohormones regulating multiple physiological processes of plants [59-62]. Using the qRT-PCR assay, we checked if MeJA, SA and IAA could induce or enhance the expression of *CYPs* and *UGTs* involved in the biosynthesis of triterpene saponins in root of *A. flaccida*. The phytohormonal responsiveness of the *CYPs* and *UGTs* expression appeared greatly different depending on the identity and concentrations of the hormone (Fig. 3). In the control group, 12 *CYPs* (*CYP12*, *CYP18*, *CYP20*, *CYP21*, *CYP22*, *CYP24*, *CYP25*, *CYP26*, *CYP27*, *CYP34*, *CYP36*, *CYP38*, and *CYP39*) and 3 *UGTs* (*UGT4*, *UGT10*, and *UGT15*) could not be detected. However, after treatment with any concentration of MeJA, with the exception of *CYP21*, all other *CYPs* were detected. Expression levels of 13 *CYPs* (*CYP1*, *CYP2*, *CYP6*, *CYP8*, *CYP10*, *CYP15*, *CYP19*, *CYP23*, *CYP28*, *CYP32*, *CYP33*, *CYP35*, and *CYP40*) increased at least two-fold, and those of 5 *CYPs* (*CYP1*, *CYP6*, *CYP8*, *CYP15*, and *CYP28*) increased at least four-fold. *UGTs* were also not dose-dependent after MeJA treatment: 7 *UGTs* (*UGT1*, *UGT2*, *UGT3*, *UGT5*, *UGT8*, *UGT11*, and *UGT12*) increased two-fold, and 4 *UGTs* (*UGT2*, *UGT8*, *UGT11*, and *UGT1*) increased four-fold.

In the SA-treated groups, in addition to *CYPs* and *UGTs* were not detected in the control group, *CYP7* also could not be detected. Except for 9 *CYPs* (*CYP8*, *CYP16*, *CYP17*, *CYP19*, *CYP23*, *CYP28*, *CYP33*, *CYP37*, and *CYP40*) and 4 *UGTs* (*UGT1*, *UGT2*, *UGT11*, and *UGT12*), all others inspected genes increased less than two-fold. The IAA-treated group had the same undetectable *CYPs* and *UGTs* as the SA-treated group. Six *CYPs* (*CYP14*, *CYP19*, *CYP29*, *CYP30*, *CYP31*, and *CYP32*) and four *UGTs* (*UGT6*, *UGT7*, *UGT9*, and *UGT14*) increased less than two-fold; all others increased more than two-fold.

Some of *CYPs* and *UGTs* genes were highly induced in roots of *A. flaccida* treated with MeJA, SA and IAA. The top five highly induced gene candidates had more than two-fold changes in expression (Table 3). Three *CYPs* were highly induced by two hormone treatments. *CYP28* increased more than 36 and 16 times by 0.05 mmol·L^{-1} MeJA and SA respectively. *CYP17* increased 6 and 49 times by 0.50 mmol·L^{-1} SA and 0.05 mmol·L^{-1} IAA. *CYP1* increased 7 and 26 times by 0.50 mmol·L^{-1} MeJA and 0.05 mmol·L^{-1} IAA. Five *UGT* genes were highly induced by one hormone treatment. For *UGTs*, three were highly induced by all three hormone treat-

ments: *UGT2* increased 25, 16 and 340 times by 5 mmol·L^{-1} MeJA, 0.50 mmol·L^{-1} SA and 0.05 mmol·L^{-1} IAA. *UGT11* increased 6, 8 and 13 times by 5 mmol·L^{-1} MeJA, 0.05 mmol·L^{-1} SA and 5 mmol·L^{-1} IAA; and *UGT12* increased 5, 13 and 20 times by 0.05 mmol·L^{-1} MeJA, 0.05 mmol·L^{-1} SA and 5 mmol·L^{-1} IAA. The expression of two genes was highly increased under two hormone treatments: *UGT5* increased 3 and 24 times by 5 mmol·L^{-1} MeJA and 0.05 mmol·L^{-1} IAA, while *UGT1* increased 2.5 and 34 times by 0.05 mmol·L^{-1} SA and IAA.

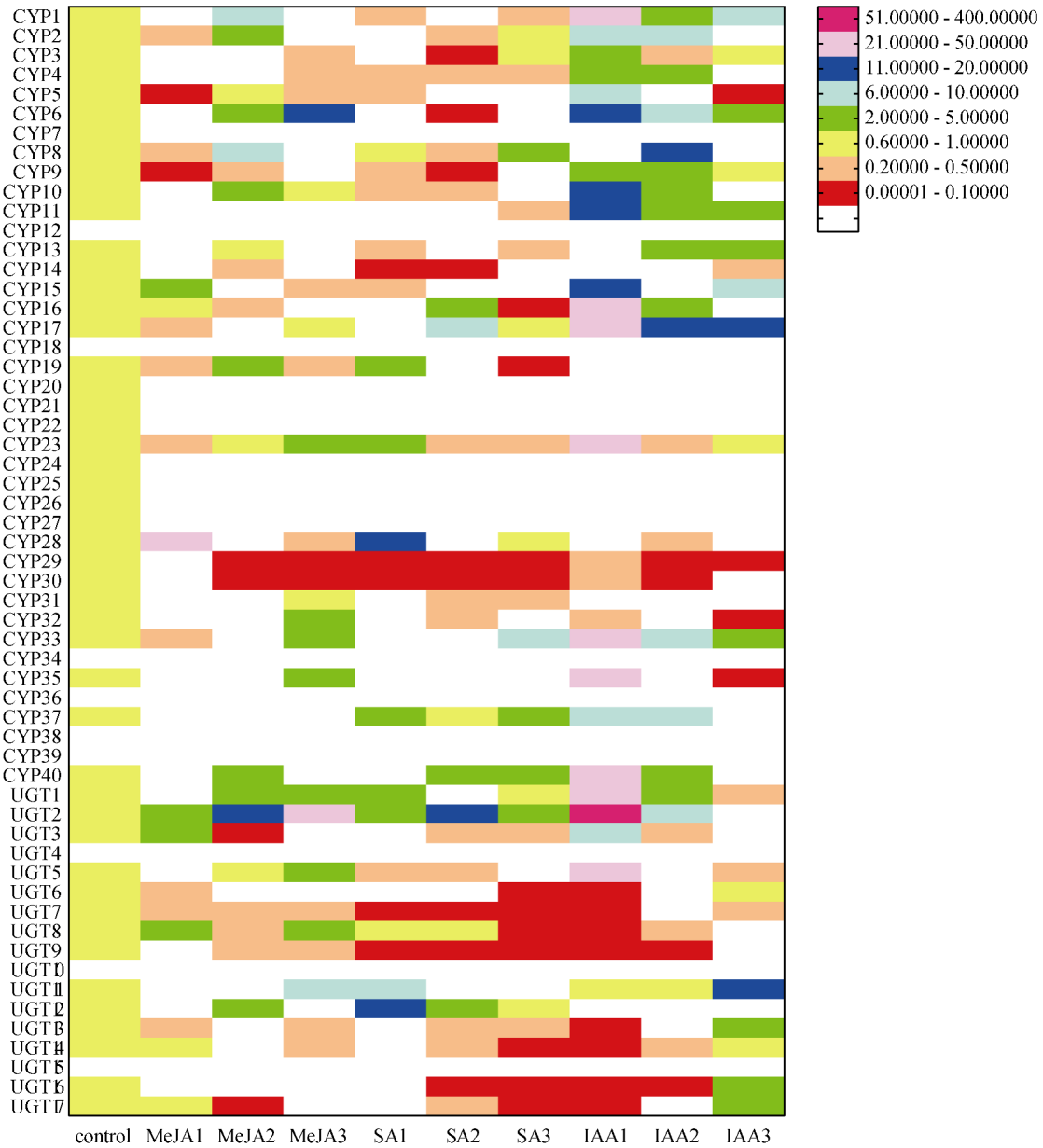

Fig. 3 The qRT-PCR analysis of *CYPs* and *UGTs* in MeJA-, SA-, and IAA-treated roots of A. flaccida. MeJA1/SA1/IAA1, 0.05 mmol·L^{-1}; MeJA2/SA2/IAA2, 0.5 mmol·L^{-1}; MeJA3/SA3/IAA3, 5 mmol·L^{-1}. *CYPs* are cytochrome P450-dependent monooxygenases (EC 1.14.x.x); *UGTs* are family 1 uridine diphosphate glycosyltransferases (EC 2.4.1.x). The corresponding unique sequence names represented by *CYPs* and *UGTs* are listed in Table 2

Discussion and Conclusion

A. flaccida is a medicinal plant that biosynthesizes bioactive ingredients classed as oleanane-type triterpene saponins [1]. Its phytochemical and ethnopharmacological properties have been studies mostly in the context of traditional Chinese medicine. Contemporary research has shown that triterpene saponins have important pharmacological effects, but their low content in *A. flaccida* and other plants limits their potential in the drug market [20].

Up-regulating biosynthesis is the key to improve the content of triterpene saponins in medicinal plants. Therefore, it is necessary to understand the biosynthetic pathways of triterpene saponins.

Table 2 Primers for qRT-PCR analysis of *CYPs* and *UGTs*

Unique sequence name	Primers for qRT-PCR (5′→3′)	Product size (bp)
isotig02871	*CYP1*-F: ATTGTTTCTTCTTTGGCATAGC *CYP1*-R: ATCAGAAAGAGATTGTAGAGACCG	139
isotig03344	*CYP2*-F: ATCTGTTATTGTCGCCTCGTC *CYP2*-R: CGCAAGTAAGGTCCATAGTCTC	155
isotig04974	*CYP3*-F: TCCACACTCACAACATCATCAG *CYP3*-R: CGTATTCCAAGATGGGTGTTC	105
isotig05047	*CYP4*-F: TAAGCCTTGGAAACGAGTCTG *CYP4*-R: TATATGCAGAGTCGCTTTTGG	128
isotig05706	*CYP5*-F: AACTTACCGAACCAGATTTGC *CYP5*-R: CTTGAAGGAGTCTAGCTAACAACATA	114
isotig06020	*CYP6*-F: CATGTCCACAGGTGTCCAATC *CYP6*-R: TGAAGCGTATGAAGGCAGTAAG	110
isotig06965	*CYP7*-F: AAAACGATGATTAGCTCTGGG *CYP7*-R: CACAGTTGGTCCATTGTTTCA	110
isotig07049	*CYP8*-F: TGTTACACATGGTTCGGTACTATTC *CYP8*-R: TTTCTAGTCTTCCCAAAGTGCC	104
isotig08321	*CYP9*-F: TAAGGCTTCACCCTCCTCTAAT *CYP9*-R: GGGGATGTAGCGACTATATGC	118
isotig08843	*CYP10*-F: ATGATTTGAGAATGGTGTGTGAC *CYP10*-R: CACAACTCTTTTGCACACGAA	133
isotig09353	*CYP11*-F: TTTGAACATTGAGTAGAACATCCA *CYP11*-R: TGGCGAAATAATCGACGAG	92
isotig09684	*CYP12*-F: GCTCTTTTCTCTTCATCGGC *CYP12*-R: CGGACTTCATACCCTGGCT	139
isotig09723	*CYP13*-F: CCGACTTTGACATTGGCAT *CYP13*-R: CCGATTTCTCAAGCCTGC	113
isotig10101	*CYP14*-F: GCACCTCCGACGACCTGT *CYP14*-R: ATTGGACGGGACCCTGAT	132
isotig11233	*CYP15*-F: GAAGAAGATCAGCCACTCGG *CYP15*-R: CCATTGCCCATTCCACAG	115
isotig11595	*CYP16*-F: CAACAAGCAACAAAATCTCGTC *CYP16*-R: GGAACACATATGCCCTATTTACTT	105
isotig11847	*CYP17*-F: GACAGGGCTCAGAAGAAAATG *CYP17*-R: GGCGTCCATCTTTGTATTGTT	128
isotig11862	*CYP18*-F: TGATATTAGCAGGAATGAAAACAAC *CYP18*-R: GAGCCCCACAATTTGGTCTA	119
isotig12183	*CYP19*-F: AGCCCGGAAGTGACAAAA *CYP19*-R: GACACAGTTTAAGTAGTCCAAATGAT	109
isotig12820	*CYP20*-F: GACTTTTTACAATAATGTCTCCTACCA *CYP20*-R: CGGTCCCATAGAGTGAAGAAAG	97
isotig12824	*CYP21*-F: CTTTTGTTGAATTGCCTTAGGA *CYP21*-R: CAGCATTAGGGGGCCATAC	150
isotig12981	*CYP22*-F: TGAACGACCGCCTGAAAC *CYP22*-R: GACGGAGAGAACGCCAAA	98

Continued

Unique sequence name	Primers for qRT-PCR (5′→3′)	Product size (bp)
isotig15430	*CYP23*-F: CTAAGTTTAGTAAAGATGGGCAGG *CYP23*-R: TCTTAGGAGAGAAAAGCTCAATAGT	102
HSAPGDX01DEGI8	*CYP24*-F: GGGTCCAAGGATCTGTTTAGG *CYP24*-R: CAACATGCCCTTTTGCTACTG	107
HSAPGDX01AV38C	*CYP25*-F: GGCTGTTCTCTCCGAGTCAC *CYP25*-R: GTGGGAACCGAGACGATG	129
HSAPGDX01BU52G	*CYP26*-F: TATTTCCTATAAACAAGAGACGACC *CYP26*-R: CACCCCTTTGCATATCCATT	110
HSAPGDX01D3WVD	*CYP27*-F: ACAACAGCGAACTTGCTAACAT *CYP27*-R: AATGGTTAAGATTTTGAATACTAGGC	132
HSAPGDX01DLAL8	*CYP28*-F: CGACTTCTTCCCATCTGTATTTAG *CYP28*-R: CAGGACAAGAGCATGGAGTATC	144
HSAPGDX01CNHZ2	*CYP29*-F: CTACTTCTTCTTTGTTGTGGGC *CYP29*-R: CATTTCTCGCAGTTGCTCAG	137
HSAPGDX01EAEL1	*CYP30*-F: GATTATCATCCACATTTACAACCAT *CYP30*-R: CGGTAGTGTGGTTTTCCAAGAT	126
HSAPGDX01CLXCM	*CYP31*-F: ATATGACCATGTGGGCTACTGT *CYP31*-R: TGGTGGTATGTTCTTTACAATCTCT	98
HSAPGDX01EOTHB	*CYP32*-F: CTTCTTCCTTGGTCATTTTGG *CYP32*-R: TCATTGGAGTTCTGTTTGCTG	145
HSAPGDX01ELSC1	*CYP33*-F: CGAAATAGATGTCTCGGAATGG *CYP33*-R: CAGCCTTCCCATCTTCATAGC	92
HSAPGDX01ELR98	*CYP34*-F: GTTCCGGCAACAAACAAGTC *CYP34*-R: CACACACGAAAATCTTAGAGCAAT	131
HSAPGDX01A8CC4	*CYP35*-F: TGAACAAGACGACCAATAGTGAT *CYP35*-R: TGACGCAAGCGGTAATGAC	114
HSAPGDX01BQ3UJ	*CYP36*-F: GGACGGCGAAGATGATTTAT *CYP36*-R: ATGGCGGGAGTGGAATGT	144
HSAPGDX01BB98P	*CYP37*-F: CACCTCAGCCACTGGGATG *CYP37*-R: TTATGGAACTCTGGCTTCAATG	128
HSAPGDX01AX7JH	*CYP38*-F: TTACAACATATTAAGTCATGGGAGA *CYP38*-R: CGAGTTTTCGAGCTCAAATAAG	134
HSAPGDX01B3O1K	*CYP39*-F: CATTTTCCTTTTCAAACCTCG *CYP39*-R: CTATTCCGGGCTCGCACT	112
HSAPGDX01BK7D9	*CYP40*-F: GACGCATTGAACGATCTTATTG *CYP40*-R: ATACTAGGATCTTGCTCATTCATGT	98
isotig00569	*UGT1*-F: TTCCCACTTCGTTTTACCCT *UGT1*-R: GGTTAGAACAGAGTGAACAAAGATT	130
isotig00626	*UGT2*-F: CCAAGTCTGTTGCCGATGA *UGT2*-R: CGTCCCGTTCTGGTTCTTT	138
isotig00779	*UGT3*-F: CAACCAACTAATGCAAAGTATGTG *UGT3*-R: CTTGACTACCCTCCCTTGTTCT	144
isotig01355	*UGT4*-F: AGCAAGATGAAGAGTGTCTGAAAT *UGT4*-R: GTGTCCTGAATTCGCAAGC	137

Continued

Unique sequence name	Primers for qRT-PCR (5′→3′)	Product size (bp)
isotig02200	*UGT5*-F: AGAACAAGTAAGTGTCACGGTCA *UGT5*-R: CACCTTCTTTTACAGCCATCTT	137
isotig03161	*UGT6*-F: GTGAGCAGCAAGTGCAAGAG *UGT6*-R: CTTCTGGTAACACTTCCTCGG	138
isotig04308	*UGT7*-F: GTGGATGGAACTCTACTGTTGAAG *UGT7*-R: CCAACTTCCCACACATCCTG	116
isotig07731	*UGT8*-F: GCAGAGCTGTCAGAGGCAAC *UGT8*-R: CCATCCACTATGTGTCTAAAAACC	112
isotig09627	*UGT9*-F: ACAAATCTGTGCCTTCTCCTG *UGT9*-R: GTTGAGACATACATAAGAAGGCTAAT	102
isotig09875	*UGT10*-F: TTGATAAGCCCGTTGAAAGAG *UGT10*-R: ATTCATTTGCTTGTTGGTCAGT	140
isotig10675	*UGT11*-F: CAGATGGAAGAACTAATCACGG *UGT11*-R: TGGCAGCAACAGTCCTTTATC	132
HSAPGDX01DVS7I	*UGT12*-F: GAAGATTTATGCAATGTCCCAG *UGT12*-R: CCTGCTTTGCTGCCATAAC	141
HSAPGDX01C65EF	*UGT13*-F: CACGAACAAGAACAACACCG *UGT13*-R: TCTCCGATGGCTTACCTCC	99
HSAPGDX01C4E7C	*UGT14*-F: AGTCCTTTCGCTCGCCTT *UGT14*-R: GCTGGTAATGAAGTCTGGAATG	136
HSAPGDX01BYVLC	*UGT15*-F: CTCATGCAACTCCATCGCT *UGT15*-R: GTTAGAGTCTGTGAAGGTGTTGATAC	117
HSAPGDX01B4J7O	*UGT16*-F: CCGCAATGGCTCAAGAAC *UGT16*-R: TCAGTTTTTATGGATAGTGAGGTCT	153
HSAPGDX01C6OA2	*UGT17*-F: GCCTCTATTGACGAACACCAC *UGT17*-R: ATCAAGCCCGAATGCGAG	132
18s ribosomal RNA	*18sRNA*-F: CCTGTTATTGCCTCAAACTTCC *18sRNA*-R: GTGGAGCGATTTGTCTGGTT	140

Table 3 Top 5 genes highly increased fold changes of expression in root of *A. flaccida* treated with MeJA, SA and IAA

Gene No.	Fold changes	MeJA (mmol·L^{-1})	Gene No.	Fold changes	SA (mmol·L^{-1})	Gene No.	Fold changes	IAA (mmol·L^{-1})
CYP28	36.00187	0.05	*CYP28*	16.87326	0.05	*CYP17*	49.86653	0.05
CYP6	12.81712	5.00	*CYP17*	6.10504	0.50	*CYP40*	37.18546	0.05
CYP1	7.63873	0.50	*CYP33*	6.07689	5.00	*CYP35*	26.90869	0.05
CYP8	7.51618	5.00	*CYP16*	4.70219	0.50	*CYP1*	26.05220	0.05
CYP15	5.80247	0.50	*CYP37*	4.56305	0.05	*CYP23*	24.19550	0.05
UGT2	25.33980	5.00	*UGT2*	16.60255	0.50	*UGT2*	340.93038	0.05
UGT11	6.75833	5.00	*UGT12*	13.23849	0.05	*UGT1*	34.13863	0.05
UGT12	5.51489	0.05	*UGT11*	8.71399	0.05	*UGT5*	24.19550	0.05
UGT8	4.52108	0.05	*UGT1*	2.56093	0.05	*UGT12*	20.15874	5.00
UGT5	3.25653	5.00				*UGT11*	13.42329	5.00

In this study, 454 pyrosequencing was employed to perform a large-scale 454-EST investigation of *A. flaccida* roots. This technology generated a high-quality sequence dataset for understanding this non-model medicinal plant. Given the paucity of data for non-model plants, *de novo* assembly was the only option for sequence assembly [63]. The resulting 454 dataset was used for the discovery of novel genes involved in saikosaponin biosynthesis and other secondary natural

products of *Anemone* and perhaps other Ranunculaceae plants.

In particular, we focused on the biosynthetic pathways of triterpene saponins in *A*. *flaccida* at the transcriptional level and showed that there are two possible biosynthetic pathways of triterpene saponins in *A*. *flaccida*. One is the well elucidated pathway for making *β*-amyrin using *β*-amyrin synthase (BAS), similar to that found in other plants. The other is an ancillary or "backburner" pathway catalyzed by a family 1 camelliol C synthase (CAMS1) which makes *β*-amyrin as a minor byproduct. Triterpenoids are usually formed in the cytoplasm *via* the mevalonate pathway [64]. Our dataset had 157 unique sequences annotated to 26 candidate genes encoding enzymes involved in triterpene saponins biosynthesis. This indicated that triterpene saponins backbones of this plant were biosynthesized not only *via* the MEP pathway in the plastids but also *via* the mevalonate pathway operating in the cytoplasm. *β*-amyrin is the basal pentacyclic oleanane-type triterpenoid aglycone backbone structure and is generally considered to be a product of 2,3-oxidosqualene cyclization catalyzed by BAS [37, 41]. In our dataset, seven unique sequences encoding BAS were discovered. In addition, one unique sequence was identified to encode CAMS1, which is the cyclase that generates predominantly a monocyclic triterpene alcohol and can catalyze 2, 3-oxidosqualene to produce *β*-amyrin (0.2%), achilleol A (2%), and camelliol C (98%) in *Arabidopsis* [42]. Thus, it could be inferred that there is a "backburner" pathway in *A*. *flaccida* to produce trace *β*-amyrin for maintaining the biosynthesis of triterpene saponins under dysfunctions of the BAS enzyme.

Using qRT-PCR analysis for all known isotigs annotated as *CYPs* and *UGTs* we showed that some isotigs had correspondingly increased expression more than four-fold respectively in response to MeJA, SA and IAA treatment. Several *CYPs* and *UGTs* were not detected in the negative control; we inferred that MeJA induced their expression. Those *CYP* and *UGT* isotig candidates with highly increased expression were likely to be functional genes, necessitating further investigations.

There are multiple copies of *HMGR* genes in *Anemone*. Based on predictions, they seemed to be fragmentized with at best one-quarter of the functional domain remaining. Two possible explanations for these dramatic differences of predicted functional domains are: a) *HMGR* genes contain new functional domain sequences that are not contained in current protein domain databases; b) *HMGR* genes do not have important functions in *Anemone* having lost their conserved sequences by mutations. Further experiments will be necessary to distinguish these possibilities.

It is noteworthy that some deduced protein sequences of unique sequences of *ABC* transporter genes, *HMGR*, *DXS*, *OSC* and *UGT,* were extremely short and lacking a structural domain. If the deduced sequences are real, they may represent new genes which were formed after the original genes underwent exon shuffles and fusions with near sequences or after they were inserted into new positions after being copied and fused with new sequences. This raises a few questions: What are the functions of extremely short isotigs? Why are there two enzymes to catalyze 2, 3-oxidosqualene to yield the pentacyclic oleanane-type triterpenoid backbone *β*-amyrin in the dataset of this study? When will CAMS1 be activated to make *β*-amyrin in *A*. *flaccida*? What will activate CAMS1 to catalyze 2,3-oxidosqualene to produce *β*-amyrin? And does CAMS1 have the same function as BAS? All these questions must be addressed through structured experiments.

In summary, an overview of transcriptomic information on the biosynthetic pathways of triterpene saponins for *A*. *flaccida* was obtained in this study. By investigating the expression of *CYPs* and *UGTs* involved in the biosynthetic pathways of triterpene saponins in phytohormone-treated roots, different isotigs annotated as *CYPs* and *UGTs* were found to have increased expressions. More importantly, these data should help to understand and guide future research on *Anemone* species and establish an important foundation for improving the content of triterpene saponins by up-regulating biosynthesis. The long-term aim of our research program is the sustainable development of drug resources. This transcriptomic dataset provides comprehensive information on gene discovery, transcriptome profiling, transcriptional regulation and molecular markers for *A*. *flaccida,* but thereby contributes to further improvements through marker-assisted breeding or genetic engineering on this species, as well as for other medicinal plants in the *Anemone* genus.

List of abbreviations:

AACT: acetyl-CoA acetyltransferase
ABC: ATP-binding cassette
BAS: *β*-amyrin synthase
BLAST: Basic Local Alignment Search Tool
CAS: cucurbitadienol synthase/cycloartenol synthase
CAMS1: family 1 camelliol C synthase
CDD: Conserved Domains Database
CMK: 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase
COG: Clusters of Orthologous Groups
cDNA: complementary DNA
CYPs: cytochrome P450-dependent monooxygenases
DMAPP: dimethylallyl diphosphate
DXR: 1-deoxy-D-xylulose-5-phosphate reductoisomerase
DXS: 1-deoxy-D-xylulose-5-phosphate synthase
EC: enzyme commission
EST: expressed sequence tag
FPP: farnesyl diphosphate
FPPS: farnesyl diphosphate synthase
GGPP: geranylgeranyl pyrophosphate
GGPPS: geranylgeranyl pyrophosphate synthase
GO: Gene Ontology
GPP: geranyl pyrophosphate
GT: glycosyltransferases
HDR: 4-hydroxy-3-methylbut-2-enyl diphosphate reductase

HDS: 4-hydroxy-3-methylbut-2-enyl diphosphate synthase
HMGR: hydroxymethylglutaryl-CoA reductase
HQ: high-quality
IAA: indole-3-acetic acid
ID: identity
IPP: isopentenyl pyrophosphate
IPPI: isopentenyl diphosphate isomerase
KEGG: Kyoto Encyclopedia of Genes and Genomes
MCT:2-C-methyl-D-erythritol-4-phosphate-cytidylyltransferase
MDR: multidrug resistance proteins
MDS:2-C-methyl-D-erythritol-2, 4-cyclodiphosphate synthase
MeJA: methyl jasmonate
MEP: 2-C-methyl-D-erythritol-4-phosphate
MRP: MDR-associated proteins
MVA: mevalonate
MVD: mevalonate diphosphate decarboxylase
MVK: mevalonate kinase
NCBI: National Center for Biotechnology Information
nt: non-redundant nucleotide
nr: non-redundant protein
OSC: oxidosqualene cyclase
P450: cytochrome P450
PDR: pleiotropic drug resistance
PMK: phosphomevalonate kinase
SA: salicylic acid
SE: squalene epoxidase
SNV: single nucleotide variation
SS: squalene synthase
SSR: Simple Sequence Repeat
TAIR: the Arabidopsis Information Resource
UGT: family 1 uridine diphosphate glycosyltransferases
WBC: white-brown complex homologs

Author Contributions

LH and GY designed the experiments and GM performed the experiments. All authors analyzed data and co-wrote the manuscript.

References

[1] Hao DC, Gu XJ, Xiao PG. Anemone medicinal plants: ethnopharmacology, phytochemistry and biology [J]. *Acta Pharm Sin B*, 2017, **7** (2): 146-158.

[2] Editorial Committee of Flora of China. *Flora of China* [M]. Beijing: Science Press, 1980, **28**:16-18.

[3] Liu Q, Zhu XZ, Feng RB, *et al*. Crude triterpenoid saponins from *Anemone flaccida* (Di Wu) exert anti-arthritic effects on type II collagen-induced arthritis in rats [J]. *Chin Med*, 2015, **10**: 20.

[4] Cao P, Wu FE, Ding LS. Advances in the studies on the chemical constituents and biologic activities for *Anemone* species [J]. *Nat Prod Res Dev*, 2004, **16**: 581-584.

[5] Sun YX, Liu JC, Liu DY. Phytochemicals and bioactivities of Anemone raddeana Regel: a review [J]. *Pharmazie*, 2011, **66** (11): 813-821.

[6] Zou ZJ, LH, Yang JS. Phytochemicals components and pharmacological activities of the genus Anemone [J]. *Chin Pharm J*, 2004, **39**: 493-495.

[7] Hao DC, Xiao PG, Ma HY, *et al*. Mining chemodiversity from biodiversity: pharmacophylogeny of medicinal plants of Ranunculaceae [J]. *Chin J Nat Med*, 2015, **13** (7): 507-520.

[8] Hao DC, Gu XJ, Xiao PG. *Medicinal plants: chemistry, biology and omics* [M]. Oxford，Elsevier-Woodhead, 2015.

[9] Hao DC, Gu XJ, Xiao PG, *et al*. Chemical and biological research of Clematis medicinal resources [J]. *Chin Sci Bull*, 2013, **58**(10): 1120-1129.

[10] Hao DC, Gu XJ, Xiao PG, *et al*. Recent advance in chemical and biological studies on Cimicifugeae pharmaceutical resources [J]. *Chin Herb Med*, 2013, **5**(2): 81-95.

[11] Han LT, Li J, Huang F, *et al*. Triterpenoid saponins from Anemone flaccida induce apoptosis activity in HeLa cells [J]. *J Asian Nat Prod Res*, 2009, **11**(2): 122-127.

[12] Liu J. Pharmacology of oleanolic acid and ursolic acid [J]. *J Ethnopharmacol*, 1995, **49**(2): 57-68.

[13] Han LT, Fang Y, Li MM, *et al*. The antitumor effects of triterpenoid saponins from the *Anemone flaccida* and the underlying mechanism [J]. *Evid Based Complement Alternat Med*, 2013, **2013**: 517931.

[14] Han LT, Fang Y, Cao Y, *et al*. Triterpenoid saponin flaccidoside II from Anemone flaccida triggers apoptosis of NF1-associated malignant peripheral nerve sheath tumors via the MAPK-HO-1 pathway [J]. *Onco Targets Ther*, 2016, **9**: 1969-1979.

[15] Huang XJ, Tang JQ, Li MM, *et al*. Triterpenoid saponins from the rhizomes of *Anemone flaccida* and their inhibitory activities on LPS-induced NO production in macrophage RAW264.7 cells [J]. *J Asian Nat Prod Res*, 2014,**16**(9): 910-921.

[16] Kong X, Wu W, Yang Y, *et al*. Total saponin from *Anemone flaccida* Fr. Schmidt abrogates osteoclast differentiation and bone resorption *via* the inhibition of RANKL-induced NF-kappaB, JNK and p38 MAPKs activation [J]. *J Transl Med*, 2015, **13**: 91.

[17] Liu C, Yang Y, Sun D, *et al*. Total saponin from *Anemone flaccida* Fr. schmidt prevents bone destruction in experimental rheumatoid arthritis *via* inhibiting osteoclastogenesis [J]. *Rejuvenation Res*, 2015, **18**(6): 528-542.

[18] Kong X, Yang Y, Wu W, *et al*. Triterpenoid saponin W3 from *Anemone flaccida* suppresses osteoclast differentiation through inhibiting activation of MAPKs and NF-kappaB pathways [J]. *Int J Biol Sci*, 2015, **11**(10): 1204-1214.

[19] Liu Q, Xiao XH, Hu LB, *et al*. Anhuienoside C ameliorates collagen-induced arthritis through inhibition of MAPK and NF-kappaB signaling pathways [J]. *Front Pharmacol*, 2017, **8**: 299.

[20] Pollier J, Goossens A. Oleanolic acid [J]. *Phytochemistry*, 2012, **77**: 10-15.

[21] Altschul SF, Gish W, Miller W, *et al*. Basic local alignment

search tool [J]. *J Mol Biol*, 1990, **215**: 403-410.

[22] Kanehisa M, Araki M, Goto S, *et al*. KEGG for linking genomes to life and the environment [J]. *Nucleic Acids Res*, 2008, **36**: D480-484.

[23] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes [J]. *Nucleic Acids Res*, 2000, **28**(1): 27-30.

[24] Yin G, Sun Z, Liu N, *et al*. Production of double-stranded RNA for interference with TMV infection utilizing a bacterial prokaryotic expression system [J]. *Appl Microbiol Biotechnol*, 2009, **84**(2): 323-333.

[25] Huang L, Yan H, Jiang X, *et al*. Identification of candidate reference genes in perennial ryegrass for quantitative RT-PCR under various abiotic stress conditions [J]. *PLos One*, 2014, **9**(4): e93724.

[26] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method [J]. *Methods*, 2001, **25**: 402-408.

[27] Pop M, Salzberg SL. Bioinformatics challenges of new sequencing technology [J]. *Trends Genet*, 2008, **24**(3): 142-149.

[28] Tholl D. Biosynthesis and biological functions of terpenoids in plants [J]. *Adv Biochem Eng Biotechnol*, 2015, **148**: 63-106.

[29] Haralampidis K, Trojanowska M, Osbourn AE. Biosynthesis of triterpenoid saponins in plants [J]. *Adv Biochem Eng/Biotechnol*, 2002, **75**: 31-49.

[30] Rohdich F, Kis K, Bacher A, *et al*. The non-mevalonate pathway of isoprenoids: genes, enzymes and intermediates [J]. *Curr Opin Chem Biol*, 2001, **5**(5): 535-540.

[31] Zheng X, Xu H, Ma X, *et al*. Triterpenoid saponin biosynthetic pathway profiling and candidate gene mining of the *Ilex asprella* root using RNA-Seq [J]. *Int J Mol Sci*, 2014, **15**(4): 5970-5987.

[32] Pollier J, Moses T, Gonzalez-Guzman M, *et al*. The protein quality control system manages plant defence compound synthesis [J]. *Nature*, 2013, **504**(7478): 148-152.

[33] Chang WC, Song H, Liu HW, *et al*. Current development in isoprenoid precursor biosynthesis and regulation [J]. *Curr Opin Chem Biol*, 2013, **17**(4): 571-579.

[34] Phillips DR, Rasbery JM, Bartel B, *et al*. Biosynthetic diversity in plant triterpene cyclization [J]. *Curr Opin Plant Biol*, 2006, **9**(3): 305-314.

[35] Kalinowska M, Zimowski J, Pączkowski C, *et al*. The formation of sugar chains in triterpenoid saponins and glycoalkaloids [J]. *Phytochem Rev*, 2005, **4**: 237-257.

[36] Vincken JP, Heng L, de Groot A, *et al*. Saponins, classification and occurrence in the plant kingdom [J]. *Phytochemistry*, 2007, **68**(3): 275-297.

[37] Abe I. Enzymatic synthesis of cyclic triterpenes [J]. *Nat Prod Rep*, 2007, **24**(6): 1311-1331.

[38] Xu R, Fazio GC, Matsuda SP. On the origins of triterpenoid skeletal diversity [J]. *Phytochemistry*, 2004, **65**(3): 261-291.

[39] Thoma R, Schulz-Gasch T, D'Arcy B, *et al*. Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase [J]. *Nature*, 2004, **432**: 118-122.

[40] Corey EJ, Matsuda SP, Bartel B. Isolation of an *Arabidopsis thaliana* gene encoding cycloartenol synthase by functional expression in a yeast mutant lacking lanosterol synthase by the use of a chromatographic screen [J]. *Proc Natl Acad Sci USA*, 1993, **90**(24): 11628-11632.

[41] Kushiro T, Shibuya M, Ebizuka Y. *β*-amyrin synthase: cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants [J]. *Eur J Biochem*, 1998, **256**(1): 238-244.

[42] Kolesnikova M, Wilson W, Lynch D, *et al*. *Arabidopsis* camelliol C synthase evolved from enzymes that make pentacycles [J]. *Org Lett*, 2007 , **9**(25): 5223-5226.

[43] Zhao L, Chen W, Fang Q. Triterpenoid saponins from *Anemone flaccida* [J]. *Planta Med*, 1990, **56**: 92-93.

[44] Huhman DV, Berhow MA, Sumner LW. Quantification of saponins in aerial and subterranean tissues of Medicago truncatula [J]. *J Agric Food Chem*, 2005, **53**(6): 1914-1920.

[45] Hayashi H, Sakai T, Fukui H, *et al*. Formation of soyasaponins in liquorice cell suspension cultures [J]. *Phytochemistry*, 1990, **29**: 3.

[46] Croteau R, Kutchan TM, Lewis NG. *Biochemistry & Molecular Biology of Plants* [M]. American Society of Plant Biologists, 2000: 1250-1318.

[47] Ross J, Li Y, Lim EK, *et al*. Higher plant glycosyltransferases. genome biology 2: reviews [J]. *Genome Biol*, 2001, **2(2)**: reviews3004.1

[48] Augustin JM, Kuzina V, Andersen SB, *et al*. Molecular activities, biosynthesis and evolution of triterpenoid saponins [J]. *Phytochemistry*, 2011, **72**(6): 435-457.

[49] Luo M, Brown RL, Chen ZY, *et al*. Transcriptional profiles uncover *Aspergillus flavus*-induced resistance in maize kernels [J]. *Toxins*, 2011, **3**(7): 766-786.

[50] Campbell JA, Davies GJ, Bulone V, *et al*. A classification of nucleotide-diphospho-sugar glycosyltranszferases based on amino acid sequence similarities [J]. *Biochem J*, 1998, **329**(Pt 3): 719.

[51] Vogt T, Jones P. Glycosyltransferases in plant natural product synthesis: characterization of a supergene family [J]. *Trends Plant Sci*, 2000, **5**(9): 380-386.

[52] Bowles D, Lim EK, Poppenberger B, *et al*. Glycosyltransferases of lipophilic small molecules [J]. *Ann Rev Plant Biol*, 2006, **57**: 567-597.

[53] Yonekura-Sakakibara K, Hanada K. An evolutionary view of functional diversity in family 1 glycosyltransferases [J]. *Plant J*, 2011, **66**(1): 182-193.

[54] Achnine L, Huhman DV, Farag MA, *et al*. Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula* [J]. *Plant J*, 2005, **41**(6): 875-887.

[55] Meesapyodsuk D, Balsevich J, Reed DW, *et al*. Saponin biosynthesis in *Saponaria vaccaria*. cDNAs encoding *β*-amyrin synthase and a triterpene carboxylic acid glucosyltransferase [J].

Plant Physiol, 2007, **143**(2): 959-969.

[56] Shibuya M, Nishimura K, Yasuyama N, *et al*. Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in *Glycine max* [J]. *FEBS Lett*, 2010, **584**(11): 2258-2264.

[57] Naoumkina MA, Modolo LV, Huhman DV, *et al*. Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in *Medicago truncatula* [J]. *The Plant Cell*, 2010, **22**(3): 850-866.

[58] Seo HS, Song JT, Cheong JJ, *et al*. Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses [J]. *Proc Natl Acad Sci USA*, 2001, **98**(8): 4788-4793.

[59] Wani SH KV, Shriram V, Sah SK. Phytohormones and their metabolic engineering for abiotic stress tolerance in crop plants [J]. *Crop J*, 2016, **4**(3): 162-176.

[60] Kazan K. Auxin and the integration of environmental signals into plant root development [J]. *Ann Bot*, 2013, **112**(9): 1655-1665.

[61] Duca D, Lorv J, Patten CL, *et al*. Indole-3-acetic acid in plant-microbe interactions [J]. *Antonie Van Leeuw*, 2014, **106**(1): 85-125.

[62] Rivas-San Vicente M, Plasencia J. Salicylic acid beyond defence: its role in plant growth and development [J]. *J Exp Bot*, 2011, **62**(10): 3321-3338.

[63] Strickler SR, Bombarely A, Mueller LA. Designing a transcriptome next-generation sequencing project for a nonmodel plant species [J]. *Am J Bot*, 2012, **99**(2): 257-266.

[64] Sui C, Zhang J, Wei J, *et al*. Transcriptome analysis of Bupleurum chinense focusing on genes involved in the biosynthesis of saikosaponins [J]. *BMC Genomics*, 2011, **12**: 539.

Cite this article as: MO Guo-Yan, HUANG Fang, FANG Yin, HAN Lin-Tao, Kayla K. Pennerman, BU Li-Jing, DU Xiao-Wei, Joan W. Bennett, YIN Guo-Hua. Transcriptomic analysis in *Anemone flaccida* rhizomes reveals ancillary pathway for triterpene saponins biosynthesis and differential responsiveness to phytohormones [J]. *Chin J Nat Med*, 2019, **17**(2): 131-144.