

RESEARCH ARTICLE

Open Access

Transcriptomic analysis of Siberian ginseng (*Eleutherococcus senticosus*) to discover genes involved in saponin biosynthesis

Hwan-Su Hwang¹, Hyoshin Lee² and Yong Eui Choi^{1*}

Abstract

Background: *Eleutherococcus senticosus*, Siberian ginseng, is a highly valued woody medicinal plant belonging to the family Araliaceae. *E. senticosus* produces a rich variety of saponins such as oleanane-type, noroleanane-type, 29-hydroxyoleanan-type, and lupane-type saponins. Genomic or transcriptomic approaches have not been used to investigate the saponin biosynthetic pathway in this plant.

Result: In this study, *de novo* sequencing was performed to select candidate genes involved in the saponin biosynthetic pathway. A half-plate 454 pyrosequencing run produced 627,923 high-quality reads with an average sequence length of 422 bases. *De novo* assembly generated 72,811 unique sequences, including 15,217 contigs and 57,594 singletons. Approximately 48,300 (66.3%) unique sequences were annotated using BLAST similarity searches. All of the mevalonate pathway genes for saponin biosynthesis starting from acetyl-CoA were isolated. Moreover, 206 reads of cytochrome P450 (CYP) and 145 reads of uridine diphosphate glycosyltransferase (UGT) sequences were isolated. Based on methyl jasmonate (MeJA) treatment and real-time PCR (qPCR) analysis, 3 CYPs and 3 UGTs were finally selected as candidate genes involved in the saponin biosynthetic pathway.

Conclusions: The identified sequences associated with saponin biosynthesis will facilitate the study of the functional genomics of saponin biosynthesis and genetic engineering of *E. senticosus*.

Keywords: Cytochrome P450, UDP-glycosyltransferase, Saponin, Transcriptome analysis, De novo sequencing, Next-generation sequencing, *Eleutherococcus senticosus*

Background

Eleutherococcus senticosus Maxim (= *Acanthopanax senticosus*) is a thorny shrub belonging to Araliaceae that grows in the Russian Far East, Northeast China, Korea and Japan. There are approximately 38 species of *Eleutherococcus*. *E. senticosus* is popularly known as Siberian ginseng because of its remarkable pharmacological effects. The cortical root and stem tissues of the plant are used as a tonic and sedative and to treat rheumatism and diabetes [1,2]. Its main ingredients are triterpenoid saponins, lignans, and phenolic compounds [3].

E. senticosus produces various types of triterpene saponins. Huang et al. [3] reviewed 43 types of triterpene saponins isolated from *E. senticosus*. The representative

saponins of *E. senticosus* are oleanane-type triterpene saponins (referred to as eleutherosides I, K, L, and M and ciwujianoside A1, C3, C4, and D1). Moreover, noroleanane-type (ciwujianoside A2, B, C1, C2, D2, and E), 29-hydroxyoleanan type (ciwujianoside A3, A4, and D3) and lupane-type triterpene saponins (chiisanoside) have been isolated from *E. senticosus* [4].

Saponin synthesis starts from the acetylated coenzyme A (acetyl CoA) molecule, from which all triterpene carbon atoms are derived. The first diversifying step in triterpenoid biosynthesis is the cyclisation of 2,3-oxidosqualene catalysed by oxidosqualene cyclase (OSC) [5]. The molecular diversity of OSCs enables more than 100 skeletal variations of triterpenoids in plants [6]. Saponins are thought to be synthesised from subsequent hydroxylation or oxidation of triterpene skeletons by CYP and glycosylation by UGT. These enzymes exist as supergene families in the plant genome. However, the key genes involved in

* Correspondence: yechoi@kangwon.ac.kr

¹Department of Forest Resources, College of Forest and Environmental Sciences, Kangwon National University, Chuncheon 200-701, South Korea
Full list of author information is available at the end of the article

saponin biosynthesis in *E. senticosus* have not been identified.

Expressed sequence tag (EST) analysis is a powerful method to discover novel genes [7]. Next-generation sequencing (NGS) technologies have enabled a genomics and genetics revolution in which the discovery of useful genes has been greatly accelerated [8,9]. NGS sequencing has been used in saponin-rich plant species such as the *Panax* species [10,11], *Siraitia grosvenorii* [12], and *Buplerum chinense* [13], and *Ilex asperlla* [14] to identify triterpene biosynthetic genes.

Despite the economic and pharmacological value of *E. senticosus*, it has not been characterised using genomic and transcriptomic approaches. In this research, 627,923 reads were generated using the Roche GS FLX titanium platform from a leaf cDNA library from *E. senticosus*. The reads were assembled to 15,217 contigs and 57,594 singletons. We focused on discovering genes encoding enzymes involved in the saponin biosynthesis pathway. Genes involved in saponin skeleton biosynthesis as well as a number of candidate genes that might be involved in modification of the triterpene saponin biosynthetic pathway skeleton, including CYPs and UGTs, were screened by elicitor treatment. Candidate CYP and UGT genes were selected based on their putative involvement in saponin biosynthesis in *E. senticosus*.

Results

Sequencing using the 454 genome sequencer FLX system and *de novo* assembly

A cDNA library constructed from total RNA extracted from *E. senticosus* leaves was sequenced on a one-half plate using the GS FLX Titanium platform. After trimming adapter sequences and removing repeat sequences or short sequences of less than 50 bp, a total of 627,923 reads were generated as 371,784 reads with an average length of 422 bp. The 371,784 reads were then used for assembly by Roche Newbler Software as 15,217 contigs and 57,594 singletons. The longest contig was 6,537 bp, with an average total contig length of 785 bp. The singletons ranged in size from 50 to 948 bp, with an average length of 368 bp. Information on bases, contigs and singletons is presented in Table 1. The size distribution of the contigs is shown in Figure 1.

Functional annotation and classification based on gene ontology (GO)

The unique sequences were compared with the NCBI non-redundant nucleotide database (Nt) and three major protein databases (KEGG, Nr, and UniProt) using the BLASTN and BLASTX algorithms with an E-value cut-off of $< 10^{-5}$. A total of 48,300 (66.3%) unique sequences with a significant match were annotated (Table 2).

Table 1 Summary of the total 454 sequencing and the assembly results for *E. senticosus* leaf tissues

Items	No. of sequences	No. of bases
Total number of reads	627,923	264,936,636
Average read length (bp)	422	
Reads used in assembly	375,482 (59.80%)	145,595,016 (54.95%)
Number of contigs	15,217	11,958,995
Average length of contigs (bp)	785	
Range of contig length (bp)	100-6,537	
Number of singletons (bp)	57,594	21,194,393
Average length of singletons (bp)	368	
Range of singleton length (bp)	50-948	

The nineteen sequences listed in Table 3 are the most abundant transcripts in the 454 cDNA library, with greater than 2,000 reads. These include the genes encoding ATP synthase, chlorophyll a/b binding protein, cell wall-associated hydrolase, and cytochrome P450. The most abundant transcript, with 7,284 reads, was annotated as a chloroplast-unknown-protein. Gene ontology (GO) analysis revealed three major categories: biological process, molecular function and cellular component. A total of 41,746 (53.4%) of unique sequences were annotated based on The Arabidopsis Information Resource (TAIR) proteins and assigned using gene ontology terms (Figure 2). The major groups of the molecular function category were transferase activity, nucleotide binding, hydrolase activity, nucleic acid binding, and kinase activity. In the cellular component group, many sequences were annotated as plasma membrane, nuclear structure, and Golgi apparatus. The best represented groups were response to the stimulus, protein metabolism, and transport in biological process categories.

Mevalonate pathway genes as candidates for involvement in saponin backbone biosynthesis

Triterpenes are assembled from a five-carbon isoprene unit through the cytosolic mevalonate pathway. Mevalonate is a product of the sequential condensation of three acetyl-CoA units to generate 3-hydroxy-3-methylglutaryl CoA (HMG-CoA), which is converted to mevalonate by HMG-CoA reductase (HMGR). The mevalonate is sequentially phosphorylated and decarboxylated to generate isopentenyl pyrophosphate (IPP). Condensation of dimethylallyl diphosphate (DMAPP) with one IPP generates geranyl diphosphate (GPP), and the addition of a second IPP unit generates farnesyl pyrophosphate (FPP). Squalene synthase (SS) catalyses the first enzymatic step from the central isoprenoid pathway toward sterol and triterpenoid biosynthesis [5]. Squalene epoxidase (SQE)

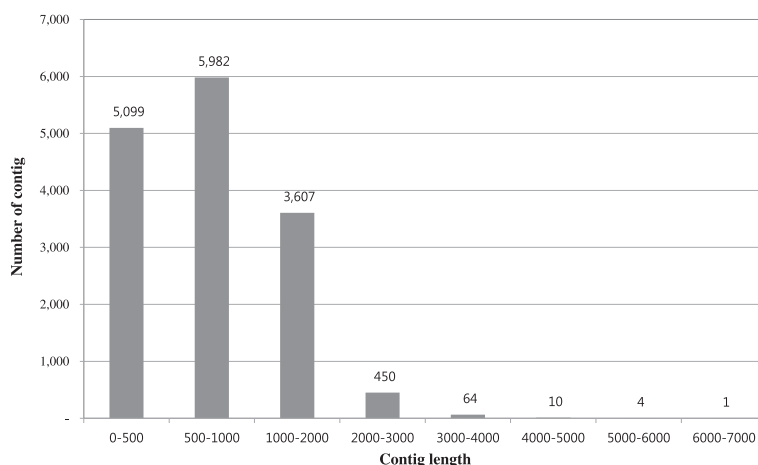


Figure 1 Length distribution of the assembled contigs of *E. senticosus*.

catalyses the first oxygenation step in phytosterol and triterpenoid saponin biosynthesis. Both phytosterols and triterpenes in plants are synthesised from the product of OSC-catalysed cyclisation of 2,3-oxidosqualene.

It has been suggested that the HMGR, SS, and SQE enzymes of the mevalonate pathway represent the rate-limiting or regulatory enzymes for saponin biosynthesis [15]. The diverse triterpene skeletons are determined by OSC. All the genes encoding enzymes involved in the upstream regions of saponin biosynthesis were successfully identified in the leaf transcriptome of *E. senticosus* (Table 4). All transcripts were annotated by more than one unique sequence as the same enzyme. A putative sequence with high similarity to SQE was found to comprise the most abundant 17 unique sequences (Table 4). The OSC sequences with high similarity to β -amyrin synthase gave the greatest number of reads (Table 4).

Oxidosqualene cyclase

Triterpenes are one of the largest classes of plant metabolites and have important functions. A diverse array of triterpenoid skeletons are synthesised via the isoprenoid pathway by OSC. The major saponins in *E. senticosus*

are eleutherosides I, K, L, and M and ciwujianosides A1, C3, C4, and D1, all these are called oleanane-type saponin derived from β -amyrin. We suggest that the aglycone of ciwujianoside E may be formed from 30-noroleanolic acid, which is frequently observed in natural compounds [16] and may be derived from 30-nor β -amyrin. The aglycone of chiisanoside is reportedly 3,4-seco-betulinic acid (Chiisanogenin) [17]. Because betulinic acid is derived from lupeol, the triterpene precursor of 3,4-seco-betulinic acid may be 3,4-seco-lupeol. Thus, *E. senticosus* may have special types of OSC genes for the production of 30-nor β -amyrin and 3,4-seco-lupeol. In Figure 3, we propose that putative 4 OSC genes are involved in triterpene biosynthesis in *E. senticosus*.

The 454 pyrosequencing of *E. senticosus* revealed a total of 15 OSC sequences, among which 4 transcripts with 323 reads were putative β -amyrin synthases, 10 transcripts with 36 reads were cycloartenol synthases, and one transcript with 31 reads was a putative lupeol synthase. An OSC full sequence (*EsBAS*) with high similarity to β -amyrin synthase was obtained (Additional file 1). The *EsBAS* cDNA was 2,738 bp long and included a 2,295 bp full open reading frame (ORF) fragment. The deduced amino acid sequence of *EsBAS* (769 amino acids with a predicted molecular mass of 88.4 kDa) is 92% and 84% identical to β -amyrin synthase (PgPNY1) in *P. ginseng* and OSCBPY in *Betula platyphylla* (Figure 4). The relatively high identities of the *EsBAS* protein with other β -amyrin proteins suggest that this gene encodes a β -amyrin synthase in *E. senticosus*.

Cytochrome P450s

CYP is a superfamily of monooxygenases, a large and diverse group of enzymes that catalyse the oxidation of organic substances. CYP is involved in a wide range of

Table 2 Summary of the annotation of the 454 assembled unique sequences

Annotation database	Annotation number	Annotation percentage (%)
KEGG	43,041	59.1
Nt	40,712	55.9
Nr	44,712	61.4
UniProt	43,300	59.5
Total	48,300	66.3

The annotations were obtained by comparing the assembled sequences with sequences from KEGG, Nr, and UniProt of public databases.

Table 3 Most abundant transcripts in *E. senticosus* leaf transcriptome

Contig ID	Length (bp)	Read	Target accession no.	Target description
EPT001TT0600C012714	253	7284	KC844054	Chloroplast, complete genome [<i>Aconitum barbatum</i>]
EPT001TT0600C012563	263	7080	KF856619	Chloroplast, complete genome [<i>Cercis canadensis</i>]
EPT001TT0600C015018	111	5674	CAA42617	Ribulose biphosphate carboxylase [<i>Phaseolus vulgaris</i>]
EPT001TT0600C012209	292	4515	CAN59721	Hypothetical protein VITISV_032350 [<i>Vitis vinifera</i>]
EPT001TT0600C000003	5621	3273	XP_003621695	ATP synthase subunit beta [<i>Medicago truncatula</i>]
EPT001TT0600C013281	215	3222	ZP_06388631	Hypothetical protein Ssol98_08391 [<i>Sulfolobus solfataricus</i>]
EPT001TT0600C012224	291	3026	AFO67221	Putative chlorophyll a/b binding protein, partial [<i>Aralia elata</i>]
EPT001TT0600C012058	305	2893	BAE46384	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit [<i>Panax ginseng</i>]
EPT001TT0600C014082	163	2852	XP_003637074	Cell wall-associated hydrolase, partial [<i>Medicago truncatula</i>]
EPT001TT0600C014525	137	2717	XP_003544026	PREDICTED: uncharacterised protein LOC100801029 [<i>Glycine max</i>]
EPT001TT0600C008793	616	2650	CAA48410	Light harvesting chlorophyll a /b binding protein [<i>Hedera helix</i>]
EPT001TT0600C013420	206	2624	BAE46384	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit [<i>Panax ginseng</i>]
EPT001TT0600C013790	181	2602	BAD26579	Cytochrome P450-like TBP [<i>Citrullus lanatus</i>]
EPT001TT0600C011532	359	2376	AFO67218	Putative ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit [<i>Aralia elata</i>]
EPT001TT0600C014447	141	2347	XP_003064992	Senescence-associated protein [<i>Micromonas pusilla</i>]
EPT001TT0600C010797	426	2319		No hit
EPT001TT0600C001408	1540	2219	XP_003588355	Mitochondrial protein, putative [<i>Medicago truncatula</i>]
EPT001TT0600C011162	393	2219	XP_003637074	Cell wall-associated hydrolase, partial [<i>Medicago truncatula</i>]
EPT001TT0600C015121	106	2045	ZP_04821298	Conserved hypothetical protein [<i>Clostridium botulinum</i>]

biosynthetic pathways, including those for lignin, terpenoids, sterol, fatty acids, and saponins [18,19]. In our sequencing results for *E. senticosus*, 84 contigs and 122 singletons were annotated as CYPs. These sequences were grouped into 32 CYP families with single and multiple copies (Additional file 2). The most abundant CYP transcripts (more than 500 454 sequencing reads) in the *E. senticosus* leaf belonged to the CYP72, CYP76, and CYP716 families. Of the 32 CYP families, we selected 22 CYP families with more than 40 copies of transcript reads as shown in Additional file 3. Among these 22 family sequences, 9 sequences had a full ORF region.

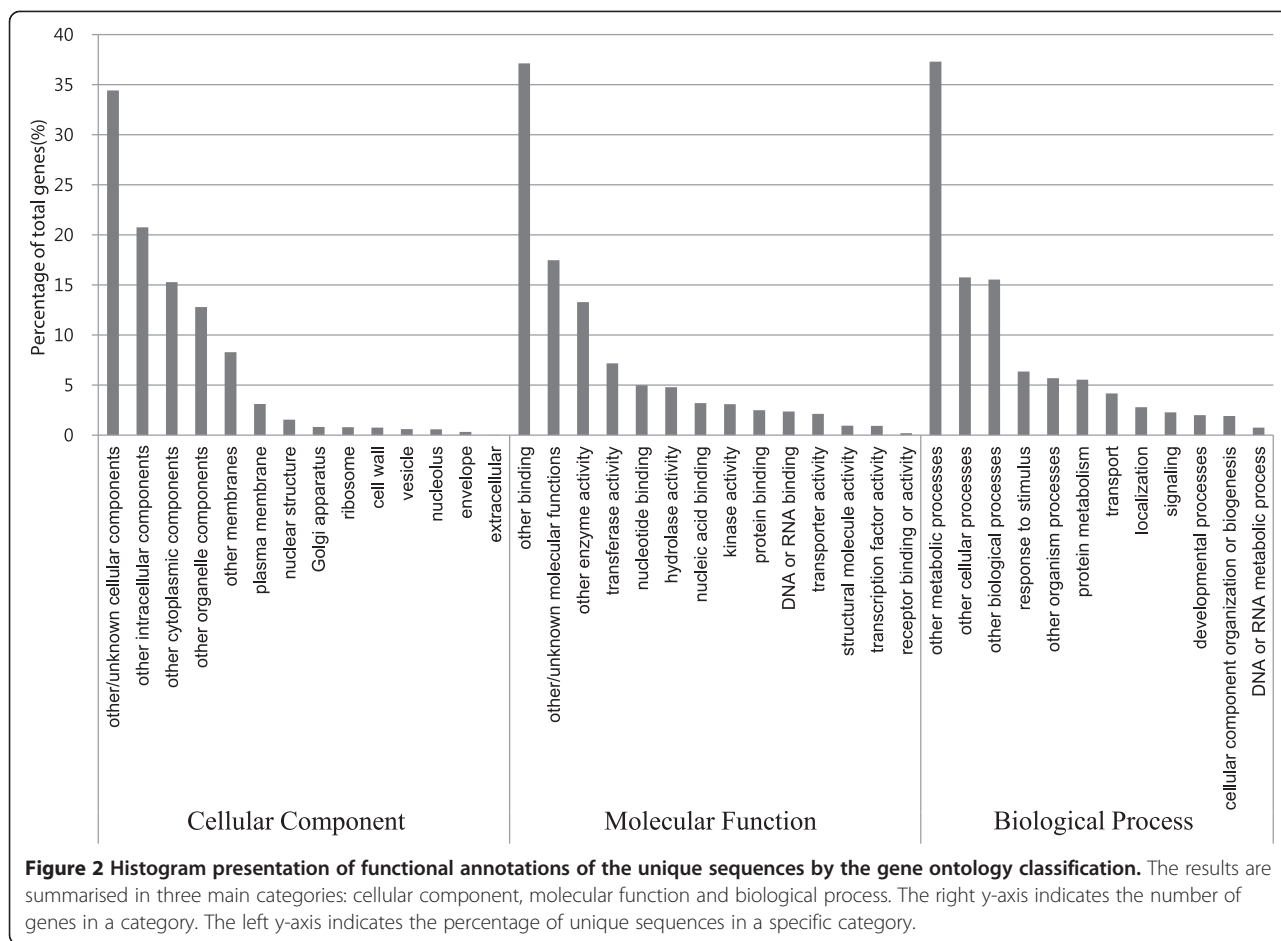
Based on the structure of the saponin aglycone, the non-saccharide portion of saponins, saponins from *E. senticosus* can be classified as several types of triterpenoid aglycones (oleanolic acid, 29-hydroxyoleanolic acid, 30-noroleanolic acid, and 3,4-seco-betulinic acid) and steroid aglycone (β -sitosterol), as shown in Figure 3. Thus, we propose that several CYP enzymes are involved in saponin biosynthesis in *E. senticosus*.

Methyl jasmonate (MeJA), a type of elicitor, has been used to increase saponin production in plant cell culture [20]. MeJA treatment also induces the strong up-regulation of enzymes related to saponin metabolism [21]. To discover genes involved in saponin biosynthesis in *E. senticosus*, the transcription of 22 putative CYP

genes in MeJA-treated leaves was monitored by qPCR for 1 day. Because the genes involved in the saponin biosynthetic pathway are simultaneously enhanced after MeJA treatment in many species, the putative β -amyrin gene (*EsBAS*) was used as a control to screen the putative CYP genes involved in saponin biosynthesis in *E. senticosus*. The transcription of the *EsBAS* gene was increased 3-fold after MeJA treatment compared to non-treatment. Three sequences of putative CYPs (CYP-3, CYP-17, and CYP-18) were clearly up-regulated by MeJA at least more than 2-fold (Figure 5).

Phylogenetic analysis revealed that CYP-3 belongs to the CYP72A subfamily (Figure 6). In *Glycyrrhiza* (licorice), CYP72A154 catalyses C-30 oxidation of β -amyrin [22], and CYP72A61v2 and CYP72A68v2 in *Medicago truncatula* modify 24-OH- β -amyrin and oleanolic acid, respectively [23]. CYP-17 is similar to *P. ginseng* CYP716A47, which is dammarenediol 12-hydroxylase [24] (Han et al. 2011). The deduced amino acid sequence of CYP-17 is 49% homologous to CYP716A47.

The saponin structure of eleutherosides I, K, L, and M and ciwujianosides A1, C3, C4, and D1 in *E. senticosus* is oleanolic acid which is derived from β -amyrin (Figure 3). β -amyrin is converted to oleanolic acid after hydroxylation by CYP716A subfamily enzymes [23,25,26]. *CYP716A12* from *Medicago truncatula* and *CYP716A52v2*

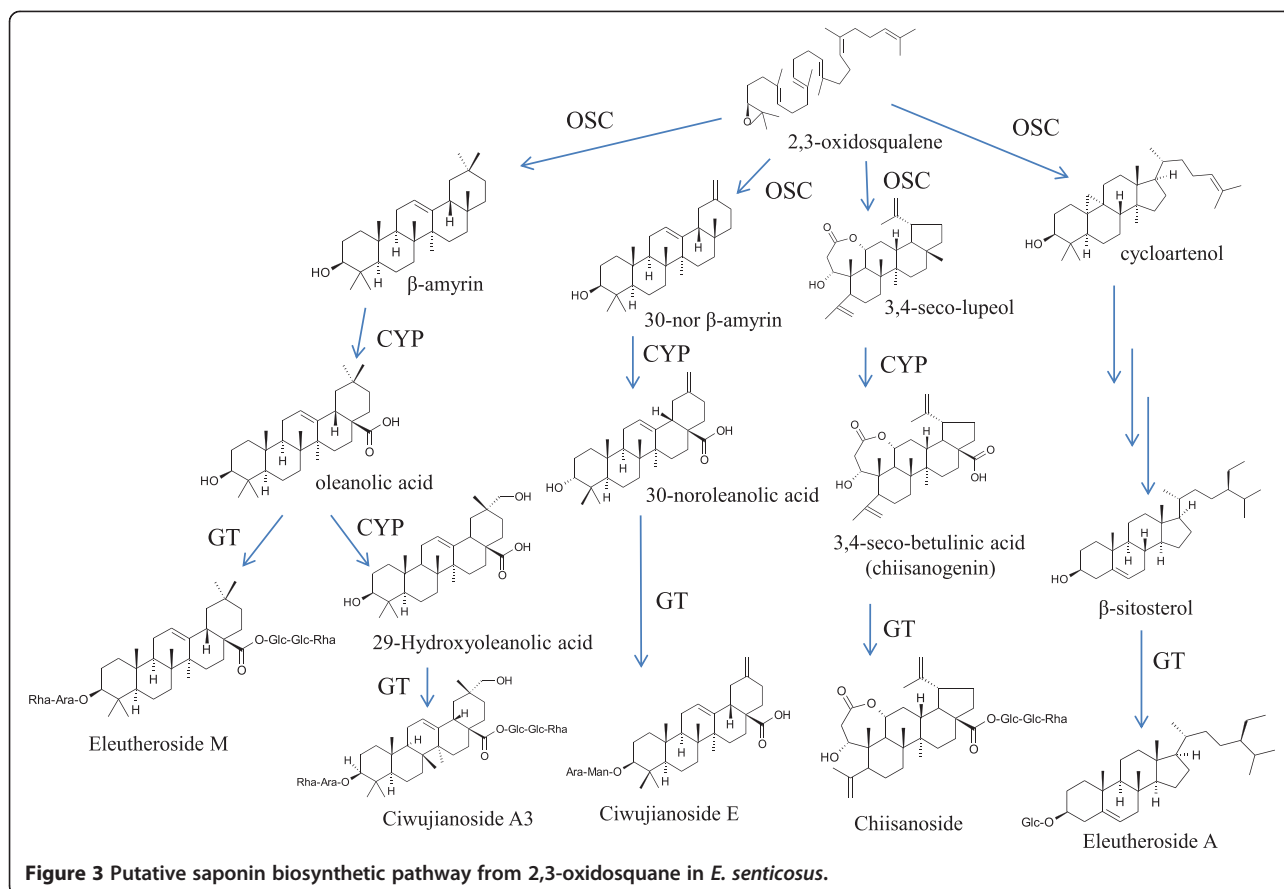


from *P. ginseng* are β -amyrin 28-oxidases (oleanolic acid synthases) belonging to the CYP85 clan [23,26]. The full sequences of the CYP-18 gene are 92% and 95% similar to *CYP716A52v2* from *P. ginseng* and *CYP716A12* from *M. truncatula*, respectively. Thus, we propose that the

CYP-18 sequence is the best candidate CYP gene determining saponin formation in the biosynthesis of eleutherosides I, K, L, and M and ciwujianosides A1, C3, C4, and D1. The enzymes responsible for 29-hydroxyoleanolic acid formation from oleanolic acid, 30-noroleanolic acid formation

Table 4 Number of putative unique sequences involved in saponin skeleton biosynthesis

Enzyme code	Enzyme name	Number of unique sequences	Number of 454 reads
2.3.1.9	Acetyl-CoA acetyltransferase	8	13
2.3.3.10	HMG-CoA synthase	8	84
1.1.1.34	HMG-CoA reductase	11	125
2.7.1.36	Mevalonate kinase	7	7
2.7.4.2	Phosphomevalonate kinase	1	8
4.1.1.33	Mevalonate-5-diphosphate decarboxylase	4	20
5.3.3.2	Isopentenyl-PP isomerase	1	22
2.5.1.10	Farnesyl diphosphate synthase	2	121
2.5.1.21	Squalene synthase	6	71
1.14.99.7	Squalene epoxidase	17	213
5.4.99.39	β -Amyrin synthase	4	323
5.4.99.41	Lupeol synthase	1	31
5.4.99.8	Cycloartenol synthase	10	36



from 30-nor β -amyrin, and 3,4-seco-betulinic acid formation from 3,4-seco-lupeol have not been identified. Thus, *E. senticosus* CYP enzymes and their effect on saponin aglycone formation merit further study.

UDP-glycosyltransferases

Saponins are high molecular weight glycosides consisting of a sugar moiety linked to a triterpenoid or steroid aglycone. All saponins feature one or more sugar chains attached to the aglycone. Glycosylation contributes to the highly diverse nature of plant secondary metabolites. UGT is a superfamily of enzymes that catalyses the addition of the glycosyl group from a UTP-sugar to a saponin molecule. Thus, UGTs are important for the regulation of saponin biosynthesis. Normally, UGTs act at the last stage of natural plant secondary metabolites and have a significant role in the stability of products and modification of biological activity [27]. In this study, 144 unique UGT sequences were identified in the *E. senticosus* transcriptome. They were classified into 18 UGT families as shown in Additional file 4. The UGT85 family gene had the most reads, with 39 unique sequences and 309 reads. The UGT73 family had the second highest number of reads, including 4 subfamilies and 14 unique sequences.

Fifteen unique sequences from each UGT family were screened to discover genes involved in saponin biosynthesis (Additional files 5 and 6). MeJA treatment also resulted in strong up-regulation of UGT enzymes related to saponin metabolism [21]. The transcription profiles of 15 UGT sequences were examined in MeJA-treated leaves of *E. senticosus* to screen the UGT genes involved in saponin biosynthesis. As shown in Figure 7, the expression of three UGT (UGT-3, UGT-10, and UGT-11) sequences was increased at least 1.5-fold after MeJA treatment. Transcription of *EsBAS* was increased approximately three-fold in MeJA-treated leaves compared to the control. The transcription of UGT-10 and UGT-11, which belong to the UGT85A subfamily, was enhanced approximately 2.5-fold in MeJA-treated leaves compared to the untreated control (Figure 7). The UGT-3 sequence belongs to the UGT73C subfamily (Figure 8). The involvement of UGT73 family genes in saponin glycosylation has been reported previously for other plants [28].

Discussion

In the present study, transcriptomic analysis of *E. senticosus* leaves was performed using the GS FLX Titanium platform. A total of 15,217 contigs and 57,594 singletons

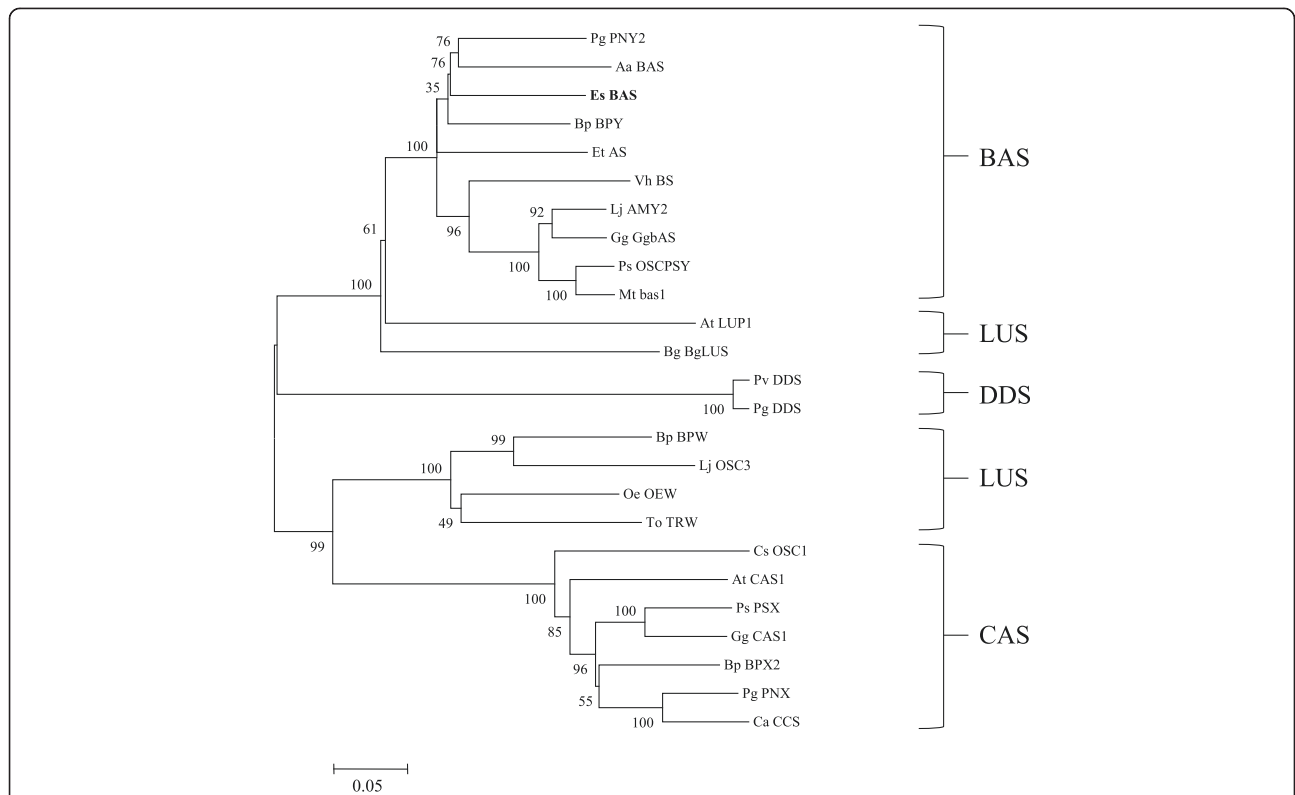


Figure 4 Phylogenetic tree of the deduced amino acid sequences of EsBAS and other plant OSCs. Phylogenetic trees of plant OSC distances between each clone and group were calculated using the program CLUSTAL W. The distance between each clone was calculated using CLUSTAL W. Bootstrap analysis values are shown at the nodal branches. The indicated scale represents 0.1 amino acid substitutions per site. Pg, *Panax ginseng*; Aa, *Artemisia annua*; Es, *Eleutherococcus senticosus*; Bp, *Betula platyphylla*; Et, *Euphorbia tirucalli*; Vh, *Vaccaria hispanica*; Lj, *Lotus japonicas*; Gg, *Glycyrrhiza glabra*; Ps, *Pisum sativum*; Mt, *Medicago truncatula*; At, *Arabidopsis thaliana*; Bg, *Bruguiera gymnorrhiza*; Pv, *Panax vietnamensis*; Oe, *Olea europaea*; To, *Taraxacum officinale*; Cs, *Crocus speciosus*; Ca, *Centella asiatica*.

were generated by assembling 627,923 reads. The most abundant cDNA sequences of the leaf transcriptome of *E. senticosus* were chloroplast-specific genes. However, we identified all sequences involved in the upstream region of the mevalonate pathway for saponin biosynthesis,

from acetyl-CoA to SS, in *E. senticosus* by searching these transcripts against sequence databases using the blastX algorithm. *E. senticosus* leaves are rich in saponins. Of a total of 43 triterpenoid saponins in *E. senticosus*, 26 were isolated from leaves [3].

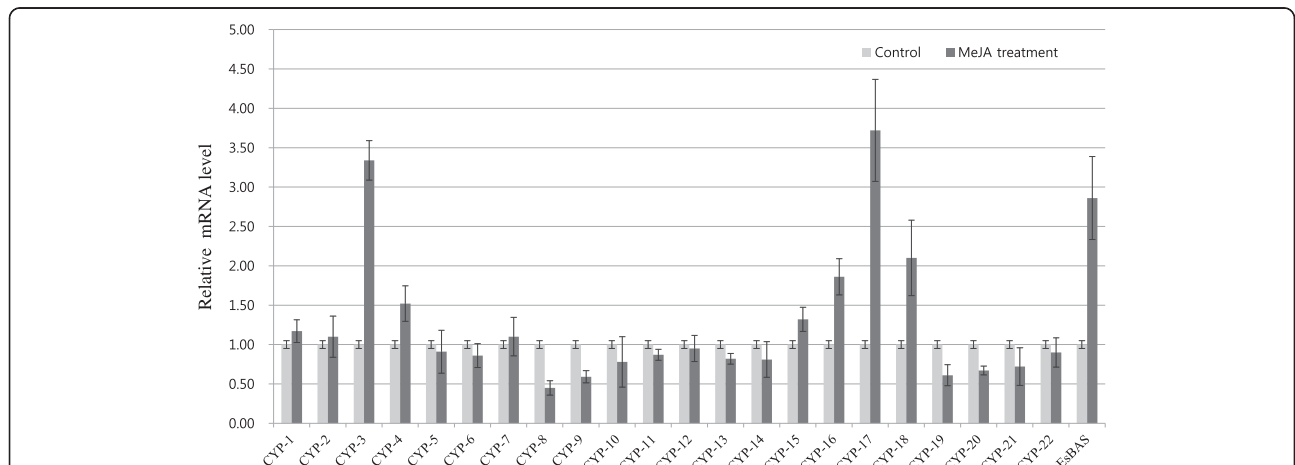


Figure 5 qPCR analysis of 22 CYPs and EsBAS in MeJA-treated leaves of *E. senticosus*. The relative fold expression of genes in MeJA-treated leaves and untreated controls is shown. EsBAS, putative β -amyrin synthase in *E. senticosus*.

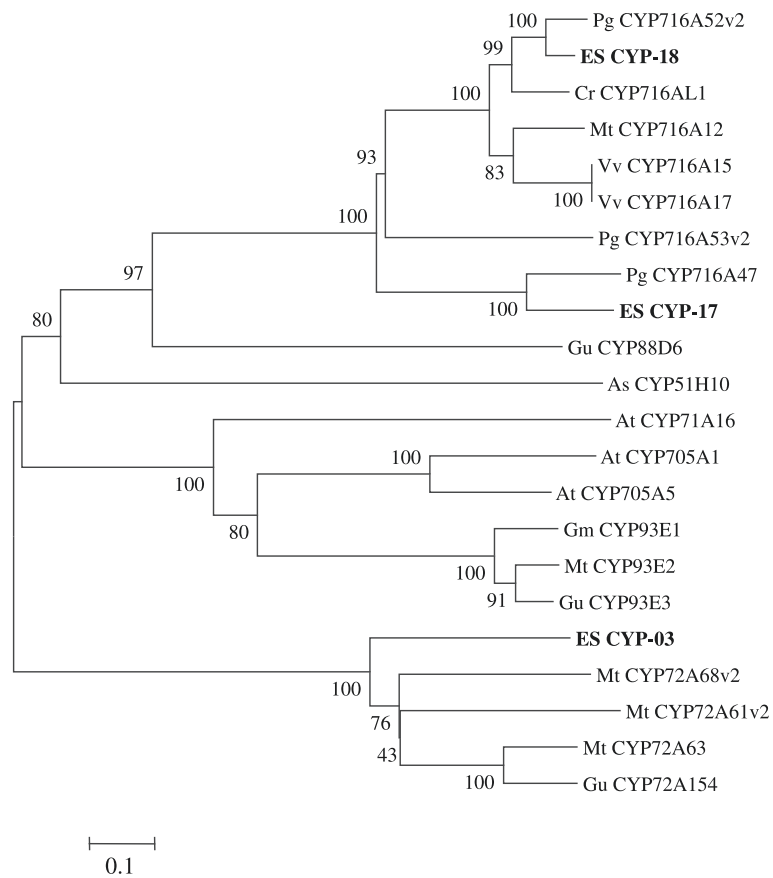


Figure 6 Phylogenetic tree of the deduced amino acid sequences of EsCYP-03, 17, 18 and other plant CYPs. Phylogenetic trees of plant OSC distances between each clone and group were calculated using the program CLUSTAL W. The distance between each clone was calculated using CLUSTAL W. Bootstrap analysis values are shown at the nodal branches. The indicated scale represents 0.1 amino acid substitutions per site. Pg, *Panax ginseng*; Es, *Eleutherococcus senticosus*; Cr, *Catharanthus roseus*; Mt, *Medicago truncatula*; Vv, *Vitis vinifera*; Gu, *Glycyrrhiza uralensis*; As, *Avena strigosa*; At, *Arabidopsis thaliana*; Gm, *Glycine max*.

SQE enzymes catalyse the conversion of squalene to 2,3-oxidosqualene. In the transcriptomic analysis of *E. senticosus*, sequences encoding SQE represented the highest number (17) of unique sequences among transcripts associated with the mevalonate metabolic pathway, with

213 sequence reads. SQE is likely an important regulatory enzyme in this pathway [15]. Single copies of the SQE gene are found in yeast and mouse, and thus disruption of SQE in these organisms is lethal [29]. By contrast, plants examined thus far have two or more copies of SQE genes.

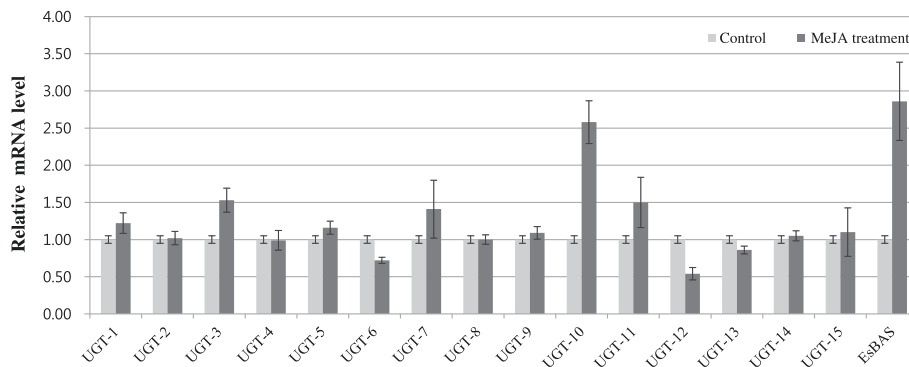


Figure 7 qPCR analysis of 15 selected UGTs of *E. senticosus* in MeJA-treated materials.

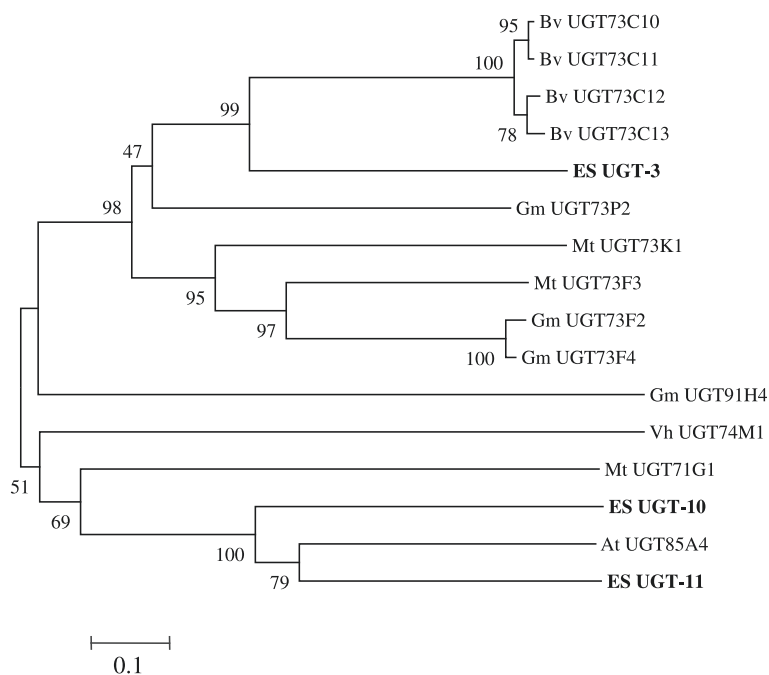


Figure 8 Phylogenetic tree of the deduced amino acid sequences of EsUGT-3, 10, 11 and other plant UGTs. Phylogenetic trees of plant OSC distances between each clone and group were calculated using the program CLUSTAL W. The distance between each clone was calculated using CLUSTAL W. Bootstrap analysis values are shown at the nodal branches. The indicated scale represents 0.1 amino acid substitutions per site. Bv, *Barbarea vulgaris*; Es, *Eleutherococcus senticosus*; Gm, *Glycine max*; Mt, *Medicago truncatula*; Vh, *Vaccaria hispanica*; At, *Arabidopsis thaliana*.

In *Arabidopsis thaliana*, 6 SQE isoforms have been identified [30], of which *SQE1*, *SQE2*, and *SQE3* encode functional SQEs, while *SQE4*, *SQE5*, and *SQE6* fail to complement the yeast *erg1* mutation. Rasbery et al. [30] suggested that *SQE* genes have different isoform-dependent functions in *Arabidopsis*. In *Medicago truncatula* cell cultures, the *SQE* gene *MtSQE2* was up-regulated by treatment with MeJA, while *MtSQE1* was not [31]. Han et al. [32] reported that the expression of *PgSQE1* and *PgSQE2* regulated in different manners and that *PgSQE1* regulates the biosynthesis of ginsenoside but not phytosterols in *P. ginseng*. The *SQE* gene responsible for saponin biosynthesis among the 17 unique *SQE* sequences in *E. senticosus* remains to be identified.

The cyclisation of 2, 3-oxidosqualene is a branch point of phytosterol and saponin synthesis, which play an important role in carbon flux regulation in other metabolic branches. In the transcriptomic analysis of *E. senticosus*, sequences encoding lupeol and cycloartenol synthase were represented in 31 and 35 reads, respectively. β -Amyrin synthase represented four unique sequences with 323 reads and thus had the most reads among the upstream genes of saponin biosynthesis (Table 4). This result suggests that transcriptional activity for oleanane-type saponin biosynthesis starting from β -amyrin may be very high in leaves of *E. senticosus*. Among the *Arabidopsis*

OSC enzymes, *ATBAS* (AT1G78950) encodes a multifunctional OSC yielding more than nine products, including β -amyrin, α -amyrin and lupeol [33]. Tomato *SITTS1* (*SIBAS*) forms β -amyrin as its sole product, while *SITTS2* catalyses the formation of seven different triterpenoids, with δ -amyrin as the major product [34]. *E. senticosus* may produce 30-nor β -amyrin and 3,4-seco-lupeol, and the characterisation of the OSC genes involved in the biosynthesis of these triterpenes will be of interest in future work.

The cyclised triterpenes undergo two additional transformations (hydroxylation and glycosidation). In oleanane-type saponin biosynthesis, the oleanolic acid sapogenin is synthesised from β -amyrin after oxidation by CYP [23,26], and this sapogenin is further glycosylated by UGT to produce various type of saponins.

The major saponins of *E. senticosus* are the eleutherosides I, K, L, and M and the ciwujianosides A1, C3, C4, and D1, which are oleanane-type triterpenoids derived from β -amyrin triterpene. Among CYP ESTs, the sequence encoding the β -amyrin gene has the most abundant reads in *E. senticosus*. This result suggests that some CYPs and UGT genes involved in oleanane-type saponin biosynthetic pathway may be abundant in EST sequences of *E. senticosus*. *CYP716A12* in *M. truncatula* [25], *CYP716A16* and *CYP716A17* in *Vitis* [23], and *CYP716A52v2* in *P. ginseng* [26] have been identified as

genes encoding β -amyrin 28-oxidase (oleanolic acid synthase). In the 454 dataset, we observed that CYP-18 of *E. senticosus* is highly homologous (90%) with *CYP716A52v2*. We suggest that this gene may convert β -amyrin to oleanolic acid in *E. senticosus*. Two other genes (CYP-3, CYP-17) that were increased by MeJA treatment compared to the control are also likely involved in saponin biosynthesis in *E. senticosus*. *CYP-17* has high similarity (86%) with *CYP716A47* from *P. ginseng*, which catalyses protopanaxadiol sapogenin formation from dammarenediol-II [24]. Fukushima et al. [23] reported that CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. Thus, the *CYP-17* gene may be involved in saponin biosynthesis in *E. senticosus*. The *CYP-3* gene is 67% identical to *CYP72A63* from *M. truncatula* and 69% identical to *CYP72A154* from *Glycyrrhiza uralensis*. All three genes encode enzymes that catalyse β -amyrin oxidation to produce different types of aglycones in saponin biosynthesis [22].

UGTs involved in saponin biosynthesis belonging to the UGT 71, 73, 74, and 91 clans have been identified previously [28]. UGT73C10 to UGT73C13 in *Barbarea vulgaris* have been reported to be involved in C-3 glycosylation of hederagenin and oleanolic acid [35]. UGT73F2 and UGT73P2 from *Glycine max* catalyse the addition of Xyl and Glc, respectively, to the Ara residue at the C-22 position of soyasapogenol A [36]. Thus, the UGT 73 clan is the best candidate group for oleanane-type saponin biosynthesis. In this study, we identified a gene (UGT-3) belonging to the UGT 73 clan whose transcription was enhanced approximately 2-fold by MeJA treatment compared to the control. However, the UGT85 family of sequences (UGT-10 and UGT-11) exhibited the highest enhancement of transcription after MeJA treatment. Based on the most abundant transcripts in the *E. senticosus* transcriptome analysis, the UGT-10 and UGT-11 sequences belonging to the UGT85A subfamily are the best candidate genes for saponin biosynthesis in *E. senticosus*. As shown in Figure 3, the huge biodiversity of saponins in *E. senticosus* suggests that various UGT genes are involved in each specific step of saponin biosynthesis. Further characterisation of UGT family enzymes is needed to validate the pathway of saponin biosynthesis in *E. senticosus*.

Conclusions

In this research, a large-scale EST sequencing was performed in leaf tissues from *E. senticosus*. The obtained EST dataset provides a useful information for gene discovery and genetic analysis in this plants. The genes involved in saponin biosynthesis pathway as well as candidate genes that might be involved in the triterpene formation, hydroxylation or oxidation of triterpene skeletons by CYP and glycosylation by UGT will help the

further research for conducting the functional genomics and transcriptomics of *E. senticosus*.

Methods

Plant materials

Fresh leaves of *E. senticosus* were collected from Mt. Odae, Pyeongchang, Kangwon-do, Korea. To determine the effect of elicitor treatment on the transcriptional activities of specific genes, leaves were exposed to 200 μ M MeJA for 8 h, and control leaves were treated with 0.25% ethanol. All tissues were immediately frozen in liquid nitrogen and stored at -80°C until use.

RNA extraction

Total RNA was extracted from leaves using Trizol reagent (MRC, USA) and RNeasy[®] Plant Mini Kit (QIAGEN, Germany) according to the manufacturer's instructions. Genomic DNA was removed from the total RNA using DNase following the manufacturer's protocol (TAKARA, Japan). mRNA was isolated from 100 μ g of total DNase-treated RNA using an mRNA purification kit (Stratagene, USA) according to the manufacturer's instructions. Agarose gel electrophoresis and the OD260/280 ratio were used to assess the quality of RNA before cDNA synthesis.

cDNA preparation and sequencing

mRNA was purified using poly-T oligo-attached magnetic beads and then fragmented with the RNA fragmentation solution supplied with the GS Titanium Library Preparation kit (454 Life Sciences, Branford, CT) following the manufacturer's recommendations. The first- and second-strand cDNAs were synthesised and end-repaired. Adaptors were ligated at the 5' and 3' ends. cDNA libraries were validated using a High Sensitivity Chip on the Agilent2100 Bioanalyzer[™] (Agilent Technologies, CA). emPCR reactions were performed on enriched cDNA templates.

The emulsions were broken, and the DNA capture beads were collected. The enriched bead samples were counted according to the manufacturer's instructions (Roche). Tagged libraries were combined in a picotitre plate for sequencing. A one-plate reaction of 454 pyrosequencing was conducted using the Roche 454 Genome Sequencer FLX System (Branford, CT, USA).

De novo assembly

The 454 Genome Sequencer FLX system collects the data and generates a standard flow gram file (.sff) that contains raw data for all the reads. The raw data were quality-filtered using a quality cut-off value of 40. The primer and adapter sequences that were incorporated during cDNA synthesis and normalisation were removed. Sequences of less than 50 bp were removed before contig

assembly. De novo contig assembly of the reads was performed using GS De Novo Assembler software provided by 454 Life Sciences Corp, CT, USA. The assembly parameters were a minimum overlap length of 40 bp and a minimum overlap identity of 95%.

A total of 627,923 reads were assembled as 15,217 contigs and 57,594 singletons, which were functionally annotated using the BLASTN program. Putative protein-encoding sequences were compared with the databases KEGG (<http://www.genome.jp/kegg/>) and UniProt (<http://www.uniprot.org/>) and searched against the Nr (www.ncbi.nlm.nih.gov/) database using the BLASTX algorithm with a cut-off E value of 10^{-5} . The functional categories of these sequences were matched to the gene ontology (GO) algorithm.

qPCR analysis

RNA was isolated from control and MeJA-treated leaves and reverse transcribed using the ImProm-II Reverse Transcription System (Promega, Madison, WI, USA). qPCR was performed using a Qiagen Rotor Gene Q Real-time PCR detector system with SYBR Green PCR Kit (Qiagen, Germany). Two-step amplification conditions for all real-time PCRs were 95°C for 5 min, followed by 40 cycles of 95°C for 5 sec and 60°C for 10 sec. The qPCR data are presented as the average relative quantities \pm SE from at least three replicates. For the MeJA inducibility experiment, the expression of each gene was used as the calibrator. The relative expression value of each gene was calculated using the $^{-\Delta\Delta C_T}$ method [37]. The *E. senticosus* β -actin gene was used for normalisation. All primers used in the present study are listed in Additional files 4 and 7.

Phylogenetic analysis

The deduced amino acid sequences of the *EsBAS*, CYP and UGT genes of *E. senticosus* and those of other plants were obtained from DDBJ/GenBank/EMBL for phylogenetic analysis. Multiple sequence alignments were generated using the CLUSTAL W program [38]. Phylogenetic analysis was performed using the neighbour-joining method with the MEGA 5.0 software program [39]. A bootstrap of 1,000 replications was used to estimate the strength of nodes in the tree [40].

Availability of supporting data

The transcriptome sequence data have been deposited into the NCBI Short Read Archive (SRA, <http://www.ncbi.nlm.nih.gov/sra/>) under the accession numbers SRR1611617. The phylogenetic alignments have been deposited in TreeBase; submission ID 17087, (<http://treebase.org/treebase-web/search/study/summary.html?id=17087&x-access-code=aef0b055f66288e54b73754f03fe0316>).

Additional files

Additional file 1: Sequence of beta-amyrin (*EsBAS*) gene in *E. senticosus*.

Additional file 2: Summary of family classification of the annotated CYPs from the 454 assembled unique sequences.

Additional file 3: Lists of sequences of selected 22 CYP genes in *E. senticosus*.

Additional file 4: List of CYPs gene-specific primer sequences used for qPCR analysis.

Additional file 5: Summary of family classification of the annotated UGTs.

Additional file 6: Lists of sequences of selected 15 UGT genes in *E. senticosus*.

Additional file 7: List of UGTs gene-specific primer sequences used for qPCR analysis.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YEC designed this study, participated in the data analysis. HSH prepared the figures and wrote the article and performed qPCR. HL performed the sample preparation. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by a grant from the Next-Generation BioGreen 21 Program (PJ011285) of the Rural Development Administration, and by Post-Genome Multi-Ministry Genome Project, and Basic Science Research Program through the NRF of Korea funded by the Ministry of Education (NRF-2013R1A1A4A01009460).

Author details

¹Department of Forest Resources, College of Forest and Environmental Sciences, Kangwon National University, Chuncheon 200-701, South Korea. ²Biotechnology Division, Korea Forest Research Institute, Suwon 441-350, South Korea.

Received: 16 October 2014 Accepted: 19 February 2015

Published online: 14 March 2015

References

- Umeyama A, Shoji N, Takei M, Endo K, Arihara S. Ciwujianosides D1 and C1: Powerful inhibitors of histamine release induced by anti-immunoglobulin E from rat peritoneal mast cells. *J Pharm Sci*. 1992;81:661–2.
- Davydov M, Krikorian AD. *Eleutherococcus senticosus* (Rupr. & Maxim.) Maxim. (Araliaceae) as an adaptogen: a closer look. *J Ethnopharmacol*. 2000;72:345–93.
- Huang LZ, Zhao HF, Huang BK, Zheng CJ, Peng W, Qin LP. *Acanthopanax senticosus*: review of botany, chemistry and pharmacology. *Pharmazie*. 2011;66:83–97.
- Shao CJ, Kasai R, Xu JD, Tanaka O. Saponins from leaves of *Acanthopanax senticosus* harms, ciwujia. Structures of ciwujianosides B, C1, C2, C3, C4, D1, D2 and E. *Chem Pharm Bull*. 1988;36:601–8.
- Abe I, Rohmer M, Prestwich GD. Enzymatic cyclization of squalene and oxidosqualene to sterols and triterpenes. *Chem Rev*. 1993;93:2189–206.
- Xu R, Fazio GC, Matsuda SPT. On the origins of triterpenoid skeletal diversity. *Phytochemistry*. 2004;65:261–91.
- Parkinson J, Blaxter M. Expressed sequence tags: an overview. *Methods Mol Biol*. 2009;533:1–12.
- Emrich SJ, Barbazuk WB, Li L, Schnable PS. Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Res*. 2007;17:69–73.
- Hahn DA, Ragland GJ, Shoemaker DD, Denlinger DL. Gene discovery using massively parallel pyrosequencing to develop ESTs for the flesh fly *Sarcophaga crassipalpis*. *BMC Genomics*. 2009;10:234.
- Sun C, Li Y, Wu Q, Luo HM, Sun YZ, Song JY, et al. De novo sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. *BMC Genomics*. 2010;11:262.

11. Chen S, Luo H, Sun Y, Wu Q, Niu Y, Song J, et al. 454 EST analysis detects genes putatively involved in ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Rep.* 2011;30:1593–601.
12. Tang Q, Ma X, Mo C, Wilson IW, Song C, Zhao H, et al. An efficient approach to finding *Siraitia grosvenorii* triterpene biosynthetic genes by RNA-seq and digital gene expression analysis. *BMC Genomics.* 2011;12:343.
13. Sui C, Zhang J, Wei JH, Chen SL, Li Y, Xu JS, et al. Transcriptome analysis of *Bupleurum chinense* focusing on genes involved in the biosynthesis of saikosaponins. *BMC Genomics.* 2011;12:539.
14. Zheng X, Xu H, Ma X, Zhan R, Chen W. Triterpenoid Saponin Biosynthetic Pathway Profiling and Candidate Gene Mining of the *Ilex asprella* Root Using RNA-Seq. *Int J Mol Sci.* 2014;15:5970–87.
15. Ryder NS. Squalene epoxidase as a target for the allylamines. *Biochem Soc Trans.* 1991;19:774–7.
16. Qu Y, Liang J, Feng X. Research in nor-oleanane triterpenoids. *Nat Prod Res Dev.* 2011;23:577–81.
17. Bae EA, Yook CS, Oh OJ, Chnag SY, Nohara T, Kim DH. Metabolism of chiisanoside from *Acanthopanax divaricatus* var. *albeofructus* by human intestinal bacteria and its relation to some biological activities. *Biol Pharm Bull.* 2001;24:582–5.
18. Meijer AH, Souer E, Verpoorte R, Hoge JH. Isolation of cytochrome P-450 cDNA clones from the higher plant *Catharanthus roseus* by a PCR strategy. *Plant Mol Biol.* 1993;22:379–83.
19. Morant M, Bak S, Moller BL, Werck-Reichhart D. Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. *Curr Opin Biotech.* 2003;14:151–62.
20. Gundlach H, Müller MJ, Kutchan TM, Zenk MH. Jasmonic acid is a signal transducer in elicitor-induced plant cell cultures. *Proc Natl Acad Sci U S A.* 1992;89:2389–93.
21. Zhao CL, Cui XM, Chen YP, Liang QA. Key enzymes of triterpenoid saponin biosynthesis and the induction of their activities and gene expressions in plants. *Nat Prod Commun.* 2010;5:1147–58.
22. Seki H, Sawai S, Ohyama K, Mizutani M, Ohnishi T, Sudo H, et al. Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. *Plant Cell.* 2011;23:4112–23.
23. Fukushima EO, Seki H, Ohyama K, Ono E, Umemoto N, Mizutani M, et al. CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. *Plant Cell Physiol.* 2011;52:2050–61.
24. Han JY, Kim HJ, Kwon YS, Choi YE. The cyt P450 enzyme CYP716A47 catalyzes the formation of protopanaxadiol from dammarenediol-II during ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Physiol.* 2011;52:2062–73.
25. Carelli M, Biazzi E, Panara F, Tava A, Scaramelli L, Porceddu A, et al. *Medicago truncatula* CYP716A12 is a multifunctional oxidase involved in the biosynthesis of hemolytic saponins. *Plant Cell.* 2011;23:3070–81.
26. Han JY, Kim MJ, Ban YW, Hwang HS, Choi YE. The involvement of β -amyrin 28-oxidase (CYP716A52v2) in oleanane-type ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Physiol.* 2013;54:2034–46.
27. Hefner T, Arend J, Warzecha H, Siems K, Stockigt J. Arbutin synthase, a novel member of the NRD1 glycosyltransferase family, is a unique multifunctional enzyme converting various natural products and xenobiotics. *Bioorg Med Chem.* 2002;10:1731–41.
28. Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A. Triterpene biosynthesis in plants. *Annu Rev Plant Biol.* 2014;65:225–57.
29. Landl KM, Klösch B, Turnowsky F. ERG1, encoding squalene epoxidase, is located on the right arm of chromosome VII of *Saccharomyces cerevisiae*. *Yeast.* 1996;12:609–13.
30. Rasbery JM, Shan H, LeClair RJ, Norman M, Matsuda SP, Bartel B. *Arabidopsis thaliana* squalene epoxidase 1 is essential for root and seed development. *J Biol Chem.* 2007;282:17002–13.
31. Suzuki H, Achnine L, Xu R, Matsuda SPT, Dixon RA. A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula*. *Plant J.* 2002;32:1033–48.
32. Han JY, In JG, Kwon YS, Choi YE. Regulation of ginsenoside and phytosterol biosynthesis by RNA interferences of squalene epoxidase gene in *Panax ginseng*. *Phytochemistry.* 2010;71:36–46.
33. Shibuya M, Katsube Y, Otsuka M, Zhang H, Tansakul P, Xiang T, et al. Identification of a product specific β -amyrin synthase from *Arabidopsis thaliana*. *Plant Physiol Biochem.* 2009;47:26–30.
34. Wang Z, Guhling O, Yao R, Li F, Yeats TH, Rose JKC, et al. Two oxidosqualene cyclases responsible for biosynthesis of tomato fruit cuticular triterpenoids. *Plant Physiol.* 2011;155:540–52.
35. Augustin JM, Drok S, Shinoda T, Sanmiya K, Nielsen JK, Khakimov B, et al. UDP-glycosyltransferases from the UGT73C subfamily in *Barbarea vulgaris* catalyze saponin 3-O-glucosylation in saponin-mediated insect resistance. *Plant Physiol.* 2012;160:1881–95.
36. Sayama T, Ono E, Takagi K, Takada Y, Horikawa M, Nakamoto Y, et al. The *Sg-1* glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. *Plant Cell.* 2012;24:2123–38.
37. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using realtime quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods.* 2001;25:402–8.
38. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22:4673–80.
39. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol.* 2011;28:2731–9.
40. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* 1985;39:783–91.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

