Metabolic Stimulation-Elicited Transcriptional Responses and Biosynthesis ofAcylated Triterpenoids in the Medicinal Plant *Helicteres Angustifolia*

Yuying Huang
Guangzhou University of Chinese Medicine
Wenli An
Guangzhou University of Chinese Medicine
Zerui Yang
Guangzhou University of Chinese Medicine
Chunzhu Xie
Guangzhou University of Chinese Medicine
Shanshan Liu
Guangzhou University of Chinese Medicine
Ting Zhan
Guangzhou University of Chinese Medicine
Xiasheng Zheng (xszheng@gzucm.edu.cn)
Guangzhou University of Chinese Medicine
Huaigeng Pan
Guangzhou University of Chinese Medicine

Research Article

Keywords: Helicteres Angustifolia, Acylated triterpenoids, Transcriptome, , Functional expressions Triterpenoids acetyl transferases, Triterpenoids benzoyl transferases.

Posted Date: April 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-395080/v1

Abstract

Helicteres angustifolia has long been used in Chinese traditional medicine. It has multiple pharmacological benefits, including anti-inflammatory, anti-viral and anti-tumor effects. Its main active chemicals include betulinic acid, oleanolic acid, helicteric acid, helicterilic acid, and other triterpenoid sapions. It is worth noting that the acylation of triterpenoids, such as helicteric acid and helicterilic acid, are characteristic components of Helicteres and are relatively rare among other plants. However, reliance on natural plants as the only sources of these restricts the potential use of H. angustifolia. Therefore, engineering of its metabolic pathway is of high research value for enhancing the production of secondary metabolites. Unfortunately, little is known of the biosynthesis of acylated triterpenoids, hindering its further investigation. Here, the RNAs of different groups treated by metabolic stimulation were sequenced with an Illumina high-throughput sequencing platform, resulting in 121 gigabases of data. A total of 424,824 unigenes were obtained after the trimming and assembly of the raw data, and 22,430 unigenes were determined to be differentially expressed. In addition, three oxidized squalene cyclases (OSCs) and four Cytochrome P450 (CYP450s) were screened, of which one OSC (HaOSC1) and one CYP450 (HaCYPi3) achieved functional verification, suggesting that they could catalyze the production of lupeol and oleanolic acid, respectively. At the same time, we also screened two triterpenoid acetyl transferases (TATs) and one triterpenoid benzoyl transferase (TBT), as subsequent structural modificatory genes, their preliminary study laid a foundation for the acylation of triterpenoids. In a nutshell, these results shed light on the regulation of acylated triterpenoids biosynthesis.

1. Introduction

Helicteres angustifolia, a conversant plant in *Helicteres* within the family Sterculiaceae, is a traditional medicinal herb used in Southern China [1] . Alcoholic and aqueous extracts of *H. angustifolia* are often used clinically to treat influenza, headache, carbuncles, hemorrhoids, tonsillitis, pharyngitis, parotitis, inflammatory diseases, and cancer, according to modern pharmacological studies [2-4]. These pharmacological benefits are caused by its chemical components, particularly the triterpenoids [5]. So far, 14 triterpenoids have been isolated and identified from *H. angustifolia* (summarized in Table 1), including betulinic acid [6], oleanolic acid [7], helicteric acid and helicterilic acid [6]. These compounds can be grouped into two parent scaffold types, namely, the lupine type (Figure 1A) and the oleanane type (Figure 1B). Intriguingly, most of these pentacyclic triterpenoids are decorated with acyl groups at the C-3 and/or C-27 position. In addition, triterpenoids are found to be acylated only in the genus *Helicteres* and rarely in other plants, to the best of our knowledge [8, 9].

Modern pharmacological studies indicate that acylated triterpenoids exhibit multiple pharmacological effects, such as easing transaminase [9], promoting apoptosis in cancer cells, and providing anti-tumor [4], anti-hepatitis B virus [10], anti-liver fibrosis[11, 12], and liver-protecting effects [13, 14]. However, in *H. angustifolia*, the content of these compounds is limited. Hence, exclusive reliance on isolating ingredients from herbal materials is impractical. Enhancement of the biosynthetic pathway of target products in homologous or heterogenous organs could be a more effective strategy for addressing this shortage than

the traditional extraction approach [15, 16]. For example, researchers have successfully improved the yield of triterpenoids in S. cerevisiae strains by introducing and overexpressing the genes that encode the key rate-limiting enzymes involved triterpenoid-synthesis pathways in Crataegus pinnatifida [17]. Regrettably, the complete lack of basic knowledge on the metabolic network of *H. angustifolia* hinders further investigation of biosynthesis and metabolic engineering. Due to the rapid development of high-throughput sequencing, de novo transcriptome analyses provide a powerful tool for screening candidate genes involved in the metabolic biosynthesis pathway of a target herb without genomic reference, which will support further investigation [18, 19].

The metabolic biosynthesis pathways for triterpenoid have been largely clarified [20, 21], mainly through the mevalonate pathways [22, 23] (Figure 1C), which produce the C5 unit IPP, which can be transformed into the isomer DMAPP by the IDI (isopentenyl diphosphate isomerase). IPP and DMAPP can be catalyzed into GPP, FPP, and GGPP through a series of pentyl transferases (GPPS, FPPS, and GGPPS). Among them, FPP is an important intermediate for triterpenoid biosynthesis. Two units of FPP can be catalyzed by squalene synthase (SS) in a tail-to-tail manner to yield the hydrocarbon squalene. Subsequently, another important precursor 2, 3-oxidesqualene, is generated under the catalysis of the squalene monooxygenase (SM) with the presence of O2 and coenzyme NADPH [24, 25]. 2, 3-oxidesqualene can then be derived into various triterpenoid skeletons under the catalysis of oxidized squalene cyclases (OSCs) [26]. Based on the known compounds and intermediates, synthetic routes were speculated for triterpenoids in *H. angustifolia*. As shown in Figure 2, the OSCs involved in the triterpenoid synthesis pathway in *H. angustifolia* are mainly lupeol synthase and β-amyrin synthase, which act on 2, 3- oxidosqualene to produce lupeol and β-amyrin, respectively. Then, C-28 oxidase catalyzed by cytochrome P450 enzyme (CYP450) is essential for modifying lupeol and β-amyrin to form betulinic acid and oleanolic acid. The acylation reaction is presumably the final step in the modification of the triterpenoids of *H. angustifolia*. Those unique triterpenoids, such as helicteric acid and helicterilic acid, are formed by acetylation at C-3 and benzoylation at C-27 on betulinic acid and oleanolic acid (Figure 2).

Triterpenoids are important metabolic substances [27, 28], and their biosynthesis can be affected by the metabolic stimulation of their source plants, including exogenous phytohormone treatment and mechanical damage [29, 30]. In this study, transcriptomes of *H. angustifolia* samples under different treatments of metabolic stimulation were sequenced and analyzed to obtain a genetic basic for the biosynthesis of the metabolic regulation of triterpenoids in vivo. Moreover, cloning and functional characterization of two key enzymes encode genes revealed their involvement in the biosynthesis of the direct precursors of medicinally acylated triterpenoids. In general, the results presented here support our understanding of the regulation and synthesis of triterpenoids in *H. angustifolia*, laying the groundwork for further studies of the potential manipulation of this pharmaceutical resource.

Table.1 Main triterpenoids in H. angustifolia

A and B correspond to parent nucleus A and parent nucleus B in Figure 1 (A-B) respectively.

Parent nucleus	R_1	R_2	R_3	Compound name
A	OH	CH_3	COOH	betulonicacid
	$OCOCH_3$	CH_2OCOBz	$COOCH_3$	helicteric acid methyl ester
	$OCOCH_3$	$CH_2OCO(4\text{-}OH)\text{-}Bz$	$COOCH_3$	27-(p-hydroxyl) benzoylhelicteric acid methyl ester
	$OCOCH_3$	CH_2OCOBz	COOH	helicteric acid
	$OCOCH_3$	CH_3	OH	betulin-3-acetate
	$OCOCH_3$	CH_3	COOH	betulonicacid-3-acetate
	OH	CH_2OCOBz	COOH	27-benzoybetulinic acid
	OH	CH_2OCOBz	$COOCH_3$	27-benzoybetulinic acid methyl ester
B	OH	CH_3	COOH	oleanic acid
	$OCOCH_3$	CH_2OCOBz	COOH	helicterilic acid
	$OCOCH_3$	CH_2OCOBz	$COOCH_3$	helicterilic acid methyl ester
	OH	CH_2OCOBz	$COOCH_3$	3β-hydroxyphenyl-27-benzoyoleanic acid
	$OCOCHCH(4\text{-}OH)Bz$	CH_3	COOH	3-o-p-coumaroyloleanolic acid
	$OCOCH_3$	$CH_2OCO(4\text{-}OH)Bz$	$COOCH_3$	27-(p-hydroxyl) benzoylhelicteric acid methyl ester

2. Materials And Methods

2.1 Plant material and metabolic stimulation treatment

Helicteres angustifolia were collected according to institutional guidelines and grown in medicinal botanical garden of Guangzhou University of Chinese Medicine (GUCM, GuangZhou, GuangDong). All plant materials were kept in permissible botanical garden at GUCM. Experimental research on plants comply with institutional, national, or international guidelines. The authenticity of plant materials has been verified by professor Huang Song in the field of traditional Chinese medicine at GUCM. Young and healthy leaves of cultivated *H. angustifolia* were rinsed with water, removing sundries on the surface of leaves, and excess water was sucked up by absorbent paper. Methyl jasmonate (MeJA) and salicylic acid (SA) with concentrations of 100-600 μM were prepared and sprayed on the surface of plant leaves. Taking the surface of the leaves as the degree of basic wetting, after 24 hours of treatment, the content of total triterpenoids in *H. angustifolia* was analyzed, and the optimal treatment concentration was determined. Then, one-third of the blade area was scratched with scissors, and the samples was taken 24 h later as the mechanical damage (MD) group. The leaves were sprayed with the optimal concentration of MeJA and SA, and sampled was taken as the MeJA treatment group and SA treatment group, respectively, after 24 h treatment. The leaves were sprayed with the configuration solvent (0.04% ethanol solution) of MeJA and SA and treated for 24 h, and then sampled as the solvent control group (EtOH). The samples without any treatment were used as negative control group (NC). 15 individuals from five different treatment groups were placed in liquid nitrogen for the determination of total triterpenoid content and the extraction of RNA.

2.2 Chemical analyses of total triterpenoid

The procedures of total triterpenoid extraction and measurement were adapted from Cai et al [31] and Oludemi et al [32]. In a nutshell, leaves were sampled and extracted with 80% ethanol in a material-to-liquid ratio of 1:30 and an ultrasonic extraction time of 2 h at 70 ℃. Then the dry extract was redissolved in n-butanol solution. With 5% vanillin-glacial acetic acid solution and perchloric acid as the chromogenic agents, the mixture's absorbance was measured with ultraviolet spectrophotometry at 548 nm. Betulinic acid served as a reference chemical marker and was used to draw the standard curve. The curve showed a good linear relationship, with an equation of y = 4.2893x + 0.0136 (r2 = 0.9995). Total triterpenoid extraction and measurement were performed in independent triplicates.

2.3 RNA extraction and transcriptome sequencing

Total RNA of the samples (Figure S1) were extracted using a plant total RNA purification kit, following the manufacturer's instructions (TIANGEN, China). The integrity and quantity of the RNA samples were analyzed by 1% agarose gel electrophoresis and with a NanoDrop 2000C Spectrophotometer (Thermo Scientific, USA). To analyze the transcriptome, qualified RNA samples were send to Majorbio (Shanghai, China) to construct the cDNA library and for sequencing on an Illumina HiSeq4000 (Illumina Inc., San Diego, CA, USA). Paired-end reads were generated and examined with fastx_toolkit_0.0.14 software (http://hannonlab.cshl.edu/fastx_toolkit/) to assess the quality of the sequence.

DEG analyses

Fragments per kilo bases per million reads (FPKM) was measured to gauge the relative expression levels, and the FPKM value of the transcripts were determined using the software RSEM v1.2.15 (http://deweylab.github.io/RSEM/) with the default parameters. DESeq2 version 1.10.1 software was used to analyze the raw counts based on the negative binomial distribution. Transcripts with expression differences between the groups were obtained using the parameters of p-adjust < 0.05 and |log2FC| >= 1. KEGG enrichment analyses of the DEGs were performed with KOBAS version 2.1.1 (http://www.genome.jp/kegg/). Nine differentially expressed unigenes related to triterpenoid biosynthesis were selected for qPCR verification. The gene and primer information was shown in Table S1.

2.4 Mining of the candidate genes involved in triterpenoid biosynthesis

Three approaches were employed to screen the genes related to the biosynthesis of triterpenoids. Firstly, BLASTx was performed using the resulting unigenes mentioned above against public protein databases, including NR, Swiss-Prot, KEGG, GO, PFAM, and COG (E value< 0.00001). Secondly, BLAST-p was adapted using the protein sequences of a set of characterized genes against the Transdecoder-predicted peptide sequences of the *H. angustifolia* transcriptome assembly (e-value < 1e-5, identify > 40%, score > 200). Lastly, using the Pfam profiles (PF13243.6, PF13249.6, PF00067.22, and PF02458.15) as queries, pfamscan based on the HMMER suite (http://hmmer.janelia.org/) was conducted (e-value < 10^−5). Positive hits were manually inspected by confirming the presence of the corresponding domains, which were classified as the putative genes. Then those candidate genes were selected to conduct phylogenetic analyses using the neighbor-joining (N-J) method in MEGA7.0, with 1000 bootstrap replicates. Expression

data were normalized by square root transformation, then used to infer co-expression gene network modules employing WGCNA R software package according to step-by-step network construction and the module detection method; soft-thresholding method were used to select a proper power-law coefficient β, and a dynamic hierarchical tree cut algorithm was used to detect the co-expression modules [33, 34].

2.5 Recombinant expression

The open reading frame (ORF) of candidate genes were amplified from cDNA of samples, primers were summarized in Table S2. And the clone vector was further constructed, which was transformed into Escherichia coli for preservation after sequencing verification. Then ORFs of the candidate OSCs and CYP450s were amplified with gene-specific primers that contained SalI and NheI restriction sites, respectively. The PCR products were subcloned into an expression vector pESC-TRP to create pESC-TRP-OSCs and pESC-TRP-CYP450s. After verifying the integrity of the gene through sequencing, recombinant plasmids were transformed into Saccharomyces cerevisiae WAT11. After being induced by 2% D-galactose, the recombinant yeast activate the GAL10 and GAL1 promoter on the pESC-TRP vector, thus inducing the expression of the target (6×His)-tagged proteins. Western-blot method was used to detect the expression levels of the fusion protein at different time points to determine the optimal induction time, the specific methods were referred to the literature reported preciously [35]. And the ORFs of the TATs and TBT were subcloned into an N-terminal 6×His tag fusion expression vector pCOLD-TF. Upon sequence confirmation, plasmids were then be transformed into E. coli Rossetta (DE3) for heterologous expression. All the gene-specific forward and reverse primers were displayed in Table S3.

2.6 GC-MS analysis of yeast extracts and enzymatic assays

The recombinant yeast cells were harvested by centrifugation after 7 days of Gal induction, then added by 20 mL lysis solution (20% KOH, 50% EtOH) for condensation reflux maintained 30 min at a slight boil. After the temperature drops, the pH value is adjusted to 2.0 with concentrated hydrochloric acid. Yeast lysate was extracted three times with equal volume of n-hexane. After the hexane phase was dried, the residue was derivatized using bis-N, O-(trimethylsilyl) trifluoroacetamide (Sigma-Aldrich) at 80 ℃ for 60 min before GC-MS analysis. The GC system equipped with an HP-5MS (30 m×0.25 mm×0.25 μm) column. The sample was injected with a flow rate of 1.2 mL/min and temperature of 250 ℃. The GC oven temperature was programmed from 80 ℃ (held for 1 min) to 300 ℃ at 40 ℃/min and kept for 10 min. For the identification of metabolites, complete mass spectra were generated by scanning within the mass-to-charge ratio range of 50 to 600. Triterpenoid products in yeast were monitored according to base peak. Empty pESC-TRP vector was used as negative control, and standard substance was used as positive control. Triterpenoid was determined by comparing both the retention time and mass spectra with the authentic standards.

For the characterization of the acyltransferase, a enzymatic assays which containing 50 mM Tris-HCl buffer (pH 7.0), 300 mM NaCl, 1 mM dithiothreitol, 30-50 μg protein, 200 μM acyl donor (250 μM acetyl CoA/benzoyl CoA) and substrates (500 μM oleanolic acid/betulinic acid) were carried out in a final volume of 200 μL. The reaction system was placed at 30 ℃ for 2 h. After the reaction, 500 μL ethyl

acetate was added to extract for twice and the extracted solution was combined. The residue was redissolved with 200 μL acetonitrile and centrifuged at 16,000 rpm for 5 min, and the supernatant was absorbed for HPLC analysis. HPLC was equipped with a Hypersil Gold AQ-C18 column (250×4.5 mm). Samples (20 μL) were injected at a column temperature of 25 °C. The liquid phase consisted of solvent A, 0.2% phosphoric acid aqueous solution, and solvent B, acetonitrile. Chemicals were separated in 75 min using the following gradient: 0-60 min, 80% B, 60-75 min, 80%-100% B, with a flow rate of 1.2 mL/min. The detection wavelength was 210 nm.

3. Results

3.1 Transcriptome analysis induced by metabolic stimulation

The optimal concentration of 500 μM methyl jasmonate (MeJA) and 400 μM salicylic acid (SA) were determined by single factor experiment (Figure S2). Consequently, the total triterpenoid content of samples from five groups was measured. The results showed elevated contents for total triterpenoids in the MD-, MeJA-, and SA-induced groups relative to the NC group (Figure S3), which was consistent with other research showing that synthesis of triterpenoids was affected by metabolic stimulation [36]. Then the RNAs of different groups treated by metabolic stimulation were sequenced with an Illumina high-throughput sequencing platform to obtain an overview of the transcriptome profile of *H. angustifolia* among the five groups. In all, 121.62 Gb high-quality data (at least 6.91 Gb for each sample and 23.46 Gb for each group) with 43.67%, GC contents were generated from 15 samples. The raw data and quality control data are characterized in Table S4. Then 581,934 transcripts and 424,824 unigenes were obtained after trimming and assembly of the clean data. The mean size of the unigenes was 655 bp, with a N50 length of 1,300 bp (Table 2).

Table 2 Analysis of assembly transcriptome

Item	Number
Total sequence base	381,166,620
Total transcripts num.	581,934
Total unigenes num.	424,824
Largest (bp)	16,851
Smallest (bp)	201
Length N_{50} (bp)	1,294
Mean Length(bp)	655

Sequence annotation was performed using the BLAST algorithm against six public databases. In all, 196,252, 179,036, 167,974, 25,390, 110,706, and 138,066 unigenes, respectively, were annotated in these

databases (Figure S4, Table S5). In total, among the 424,824 unigenes generated by transcriptional sequencing, 245,709 (57.84%) received at least one hit in those databases, with 13,320 unigenes sharing common annotation (Figure S5).

3.2 Analysis of DEGs under different treatments

The transcriptome statistics of those five groups were compared pairwise to establish potential DEGs among the different groups. Compared to the NC group, there was the greatest amount of DEGs in the MD group (Figure 2A), and the number of up-regulated DEGs in each metabolic stimulation treatment group and EtOH group were higher than that of the downregulated genes (Figure 2B). The Venn diagram presented in Figure 2C indicates that 774 DEGs were shared among the four aforementioned pairwise comparisons, and a total of 22,430 DEGs were obtained among the four comparison groups. Heat maps were produced based on the expression patterns of all DEGs, and these could be divided into eight clusters (Figure 2D). KEGG enrichment analyses of DEGs in four pairwise groups（Figure S6, Table S6-9） suggested that pathways with the most enrichment were mostly associated with triterpenoid synthesis, revealing the expression of genes related to triterpenoid synthesis was upregulated after metabolic stimulation, leading to the enhancement of triterpenoid accumulation for resistance, which is consistent with the results of previous research [30, 37].

In addition, WGCNA was conducted to further investigate the potential genes involved in the triterpenoids biosynthesis of *H. angustifolia*. All DEGs were submitted to construct a scale-free co-expression network. The dynamic hierarchical tree algorithm was employed to divide cluster trees constructed by DEGs, resulting in 10 co-expression modules. These modules are named according to their colors, namely blue (1619 genes), brown (891 genes), turquoise (2505 genes), grey (17 genes), yellow (759 genes), pink (55 genes), green (673 genes), black (87 genes), magenta (52 genes), and red (115 genes) (Figure 3). The genes within each module were selected for enrichment in the KEGG pathway. Statistically significantly enriched genes ($P < 0.05$) were filtrated for deep analyses. Results showed that five modules (Blue, turquoise, black, green and magenta), consisting of 4,936 genes in total, were related to triterpenoid biosynthesis.

3.3 Quantitative real-time PCR (qRT-PCR) to verify the DEGs.

The expression patterns of 41 key candidate genes (Figure S7, Table S10) involved in triterpenoids biosynthesis were analyzed. Among them, 18 genes presented significantly different expression levels among five treatment groups, nine of them were chosen to analyze the expression level using qRT-PCR to further verify the reliability of our RNA-Seq data. As shown, the expression patterns of these nine DEGs were consistent with the RNA-Seq data (Figure 4), which confirmed the accuracy of our transcriptome data.

3.4 Cloning and sequence analysis of the candidate genes

To shed light on acylated triterpenoids biosynthesis in *H. angustifolia*, putative key enzymes encoding genes including three OSCs (named HaOSC1, HaOSC2 and HaOSC3), four CYP450s (named HaCYPi1, HaCYPi2, HaCYPi3 and HaCYPi4), two TATs (named HaTAT1 and HaTAT2), and one TBT (named HaTBT) were targeted for the functional analysis. Their nucleotide sequences were summarized and exhibited in Table S11 and physicochemical properties of corresponded proteins were documented in the Table S12, two-dimensional structure information of proteins was classified in the Table S13 and three-dimensional structures were displayed in Figure S8.

Phylogenetic analyses of three OSCs with a set of well characterized OSCs revealed that HaOSC1 and HaOSC3 were grouped into the lupeol synthase clade, and HaOSC2 was classified as a β-amyrin synthase (Figure 5A, Table S14). In addition, we found that all three candidate OSCs contained complete SQHop-cyclase-N and SQHop-cyclase-C domains. Furthermore, several conserve motifs were present in all three candidate OSCs, including DCTAE, MWCYCR, and QW repeat (Figure 5B).

The results of phylogenetic analyses also implied that four CYP450 candidate genes were well documented in the CYP716A subfamily (Figure 5C, Table S15), which primarily includes CYP450 enzymes with a C-28 oxidation function [38, 39](Fukushima et al. 2011; Suzuki et al. 2018). Analysis of conserved domains demonstrated that four candidate genes all had a complete Cytochrome P450 domain. Motif elucidation revealed that HaCYPi1, HaCYPi3, and HaCYPi4 possessed conserved motifs, including EXXXR, (A/G)GX(D/E)T and FXXGXRXCXGX, while HaCYPi2 lacks the motif FXXGXRXCXGX (Figure 5D).

Meanwhile, the three candidate acyltransferase coding genes were selected to perform phylogenetic analyses with other functionally characterized BAHDs. HaTBT falls into branch that principally includes acyltransferases related to benzenoid ester production. HaTAT1 and HaTAT2 were aggregated into the same branch as other anthocyanin-modifying acetyl transferases (Figure S9A, Table S16). Domain analysis showed that the transferase family domain existed in these three candidate genes. And they all possess three amino acid motifs, HXXXD, YFGNC, and DFGWG (Figure S9B), which are highly conserved in BAHD enzymes [40].

Heterologous expression and functional characterization of the OSCs

Three candidate OSCs were functionally validated by expressing their ORFs in expression vector pESC-TRP. Recombinant plasmids were transferred into yeast (Saccharomyces cerevisiae strain WAT11) by lithium acetate conversion method. Protein expression of the recombinant yeast within 24 hours were investigated, which implied that the recombinant yeast pESC-TRP-HaOSC1 successfully expressed the protein with the molecular mass of about 86 kDa and the protein expression reached the highest level 8 hours after induction (Figure 6A).

Gal-induced transgenic yeast cells were also analyzed for the accumulation of triterpenoids metabolites. Yeast cells were extracted and metabolites were identified through gas chromatography-mass spectrometry (GC-MS) analysis. By comparing the retention time and mass spectra of those new

products with authentic standards (a-amyrin, β-amyrin and lupeol), we found that HaOSC1-expressing transgenic yeast accumulated lupeol, which was not present in the control yeast cells transformed with empty vector (Figure 6B-C). The yeast expression experiments were repeated three times, always with very similar results. These results clearly manifested that HaOSC1 encoded a lupeol synthase, which could use 2, 3-oxidized squalene in yeast as substrate to synthesize the target product lupeol (Figure 6D).

3.5 Functional validation of the CYP450s

To determine the product specificities of HaCYP450s, ORF of four candidates was expressed in yeast under the control of the gal-inducible GAL10 promoter. Induced cells were further used for Western blot analysis, the fusion protein bands of HaCYPi3 of about 54 kDa were observed. The induction time was also investigated, same with HaOSC1, the protein expression reached the highest level at about 8 hours after induction (Figure 7A).

Based on the previous sequence analysis, HaCYPi3 was predicted to be a potential C-28 oxidase, catalyzing α-amyrin, β-amyrin and Lupeol substrates. Therefore, we firstly transferred pESC-HaCYPi3 and the above pESC-HaOSC1 recombinant plasmid into the same yeast for co-expression. The metabolites after induction were analyzed by GC-MS. Unfortunately, the target product betulinic acid was not detected, which suggested that HaCYPi3 failed to encode for a lupeol oxidase. Then, we co-expressed HaCYPi3 with α-/β-amyrin synthase IaAS previously identified by our research group from Ilex asprell. It is surprising that IaAS and HaCYPi3 co-expressing transgenic yeast accumulated new products compared with empty vector. Except for the corresponding substrate peaks, the retention time and mass spectra of one new product was consistent with oleanolic acid authentic standards (Figure 7 B-C). The results of three repeated experiments were consistent, which clearly demonstrated that HaCYPi3 encoded a β-amyrin oxidase, which could use β-amyrin catalyzed by IaAS in yeast as substrate to synthesize the target product oleanolic acid (Figure 7D). Furthermore, the subcellular localization of HaCYPi3 was observed, results displayed that it was localized in the cytoplasm (Figure 7E)，which is consistent with reports that triterpenoids were synthesized in the cytoplasm.

4. Discussion

Triterpenoids, as a natural plant product with extensive biological activities, have attracted the attention of many researchers. Currently, the biosynthetic pathways of many triterpenoids have been successfully identified, such as ginsenosides [41], ganoderic acid [42] and glycyrrhizic acid [43]. However, acylated triterpenoids are uncommon, which has only been reported in the plant belonging to *Helicteres* so far. Hence, *H. angustifolia* can be used as a good research material to clarify the biosynthesis pathway of acylated triterpenoids.

The biosynthesis of triterpenoids has been reported to be affected by plant metabolic stimulation, including exogenous plant hormone treatment and mechanical damage [44-46]. Exogenous methyl jasmonate (MeJA) and salicylic acid (SA) can act as signal transduction molecules in plant cells and induce the production of terpenoids, phenols and other secondary metabolites [47, 48]. As a means of

abiotic stress, mechanical damage can effectively regulate the growth, development, stress response and metabolites of plants [28]. In this study, the content of total triterpenoids in *H. angustifolia* was analyzed after different concentration hormone stimulation treatment, and the optimal treatment conditions were optimized. Furthermore, transcriptome sequencing of *H. angustifolia* under the optimized treatment conditions was carried out to reveal the biosynthetic pathway at molecular level, which provided theoretical basis for the effective development and utilization of triterpenoids in *H. angustifolia*.

Bioinformatics analysis including phylogenetic and conserved motifs suggested that HaOSC2 might be a potential β-amyrin synthase, but in this experiment, new product was failed to be detected in induced yeast metabolites, the reason might be that the protein expressed in yeast cannot be properly folded, resulting in its inactivity. In the future, we will try to construct the plant expression vector of HaOSC2, and further study its function in model plants such as Nicotiana benthamiana or Arabidopsis thaliana.

Acylation is the final and most critical step in the biosynthesis of acylated triterpenoids, which plays an important role in many modifications of plant metabolism, such as changing the polarity, volatility, chemical stability, and biological activity of the metabolites [49]. However, there have been only a few studies of plant acyltransferases, predominately focusing on Arabidopsis thaliana [50], Oryza sativa [51], and Populus simonii [52], which limits the further understanding of this family of enzymes. Therefore, our study of triterpenoid acyl transferase in *H. angustifolia* is helpful to fill in these gaps. In this study, prokaryotic expression vectors carrying three candidate acyltransferases were successfully constructed, and the fusion proteins were induced to express in Escherichia coli Rosetta (DE3) strain. Three purified target proteins were successfully obtained by purifying poly (His) (6×His)-tagged proteins using a nickel-nitrilotriacetic acid agarose column (Figure S10). Unfortunately, by comparing the enzyme activity system of the inactivated protein with that of the non-inactivated protein, the results showed that no new products were produced in the enzyme activity reactions of the three candidate proteins. Reason can be speculated that the catalytic efficiency of the enzyme is too low, or the substrates catalyzed by these three proteins are not oleanolic acid and betulinic acid and so on. Hence, their functions still need to be further confirmed.

5. Conclusion

Helicteres angustifolia, a shrub distributed in Southern China and used in traditional Chinese medicine, is rich in triterpenoids such as betulinic acid, oleanolic acid, helicteric acid, helicterilic acid, and similar derivatives. The biosynthetic pathway of these compounds is of great value due to their valuable medicinal activity. Here, we firstly report the transcriptomic study of *H. angustifolia* for the exploration of functional genes. Three OSCs, four CYP450s, two TATs, and one TBT were screened out as candidate genes. Their further functional verification demonstrated that HaOSC1 and HaCYPi3 encodes for a lupeol synthase and β-amyrin oxidase, respectively. The results of this study can form a basis for future analyses of the biosynthetic pathway of acylated triterpenoids, and lay a foundation for future genetic engineering to increase the yield of medicinal triterpenoids in *H. angustifolia*.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The sequencing datasets generated during the current study are available at CNGBdb: CNP0001265 (https://db.cngb.org/search/project/CNP0001265/).

Competing interests

The authors declare that they have no conflict of interest.

Funding

This research was supported by the National Natural Science Foundation of China (Grant number: 81903741). The funding agencies had no role in research design, data collection and analysis, or manuscript writing.

Authors' contributions

All authors have read and approved the manuscript. XSZ and HGP conceived and designed the experiments. YYH and WLA, conduct the experiments, analyzed the data and wrote the manuscript. ZRY participated in bioinformatics analysis, SSL, CZX and TZ participated in experiments and wrote the manuscript.

Acknowledgements

We apologize to the many colleagues whose work could not be cited in this manuscript due to space limitation.

References

1. Li K, Lei Z, Hu X, Sun S, Li S, Zhang Z: **In vitro and in vivo bioactivities of aqueous and ethanol extracts from Helicteres angustifolia L. root**. *J Ethnopharmacol* 2015, **172**:61-69.
2. Li K, Yu Y, Sun S, Liu Y, Garg S, Kaul SC, Lei Z, Gao R, Wadhwa R, Zhang Z: **Functional Characterisation of Anticancer Activity in the Aqueous Extract of Helicteres angustifolia L. Roots**. *PLoS One* 2016, **11**(3):e0152017.

3. Yin X, Lu Y, Cheng ZH, Chen DF: **Anti-Complementary Components of Helicteres angustifolia**. *Molecules (Basel, Switzerland)* 2016, **21**(11).
4. Sun S, Li K, Xiao L, Lei Z, Zhang Z: **Characterization of polysaccharide from Helicteres angustifolia L. and its immunomodulatory activities on macrophages RAW264.7**. *Biomed Pharmacother* 2019, **109**:262-270.
5. Pan MH, Chen CM, Lee SW, Chen ZT: **Cytotoxic triterpenoids from the root bark of Helicteres angustifolia**. *Chem Biodivers* 2008, **5**(4):565-574.
6. Su D, Gao YQ: **Helicteric Acid, Oleanic Acid, and Betulinic Acid, Three Triterpenes from Helicteres angustifolia L., Inhibit Proliferation and Induce Apoptosis in HT-29 Colorectal Cancer Cells via Suppressing NF-κB and STAT3 Signaling**. 2017, **2017**:5180707.
7. Wang GC, Li T, Wei YR, Zhang YB, Li YL, Sze SC, Ye WC: **Two pregnane derivatives and a quinolone alkaloid from Helicteres angustifolia**. *Fitoterapia* 2012, **83**(8):1643-1647.
8. Chang YS, Ku YR, Lin JH, Lu KL, Ho LK: **Analysis of three lupane type triterpenoids in Helicteres angustifolia by high-performance liquid chromatography**. *J Pharm Biomed Anal* 2001, **26**(5-6):849-855.
9. Chen L, Liu J, Ge X, Xu W, Chen Y, Li F, Cheng D, Shao R: **Simulated digestion and fermentation in vitro by human gut microbiota of polysaccharides from Helicteres angustifolia L**. *Int J Biol Macromol* 2019, **141**:1065-1071.
10. Huang Q, Huang R, Wei L, Chen Y, Lv S, Liang C, Zhang X, Yin F, Li H, Zhuo L *et al*: **Antiviral activity of methyl helicterate isolated from Helicteres angustifolia (Sterculiaceae) against hepatitis B virus**. *Antiviral Res* 2013, **100**(2):373-381.
11. Biggs BW, Lim CG, Sagliani K, Shankar S, Stephanopoulos G, De Mey M, Ajikumar PK: **Overcoming heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in Escherichia coli**. *Proceedings of the National Academy of Sciences of the United States of America* 2016, **113**(12):3209-3214.
12. Zhang XL, Chen ZN, Huang QF, Bai FC, Nie JL, Lu SJ, Wei JB, Lin X: **Methyl Helicterate Inhibits Hepatic Stellate Cell Activation Through Modulation of Apoptosis and Autophagy**. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology* 2018, **51**(2):897-908.
13. Lin X, Huang R, Zhang S, Zheng L, Wei L, He M, Zhou Y, Zhuo L, Huang Q: **Methyl helicterate protects against CCl4-induced liver injury in rats by inhibiting oxidative stress, NF-κB activation, Fas/FasL pathway and cytochrome P4502E1 level**. *Food Chem Toxicol* 2012, **50**(10):3413-3420.
14. Wen S, Wei Y, Zhang X, Bai F, Tan S, Nie J, Wei J, Lin X: **Methyl helicterilate ameliorates alcohol-induced hepatic fibrosis by modulating TGF-β1/Smads pathway and mitochondria-dependent pathway**. *Int Immunopharmacol* 2019, **75**:105759.
15. Erb TJ, Jones PR, Bar-Even A: **Synthetic metabolism: metabolic engineering meets enzyme design**. *Curr Opin Chem Biol* 2017, **37**:56-62.

16. Choi KR, Jang WD, Yang D, Cho JS, Park D, Lee SY: **Systems Metabolic Engineering Strategies: Integrating Systems and Synthetic Biology with Metabolic Engineering**. *Trends Biotechnol* 2019, **37**(8):817-837.
17. Dai Z, Liu Y, Sun Z, Wang D, Qu G, Ma X, Fan F, Zhang L, Li S, Zhang X: **Identification of a novel cytochrome P450 enzyme that catalyzes the C-2α hydroxylation of pentacyclic triterpenoids and its application in yeast cell factories**. *Metab Eng* 2019, **51**:70-78.
18. Heather JM, Chain B: **The sequence of sequencers: The history of sequencing DNA**. *Genomics* 2016, **107**(1):1-8.
19. Zhang X, Li T, Liu F, Chen Y, Yao J, Li Z, Huang Y, Wang J: **Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems**. *Mol Cell* 2019, **73**(1):130-142.e135.
20. Haralampidis K, Trojanowska M, Osbourn AE: **Biosynthesis of triterpenoid saponins in plants**. *Adv Biochem Eng Biotechnol* 2002, **75**:31-49.
21. Biswas T, Dwivedi UN: **Plant triterpenoid saponins: biosynthesis, in vitro production, and pharmacological relevance**. *Protoplasma* 2019, **256**(6):1463-1486.
22. Yang C, Gao X, Jiang Y, Sun B, Gao F, Yang S: **Synergy between methylerythritol phosphate pathway and mevalonate pathway for isoprene production in Escherichia coli**. *Metab Eng* 2016, **37**:79-91.
23. Kobayashi K, Suzuki M, Muranaka T, Nagata N: **The mevalonate pathway but not the methylerythritol phosphate pathway is critical for elaioplast and pollen coat development in Arabidopsis**. *Plant biotechnology (Tokyo, Japan)* 2018, **35**(4):381-385.
24. Wölwer-Rieck U, May B, Lankes C, Wüst M: **Methylerythritol and mevalonate pathway contributions to biosynthesis of mono-, sesqui-, and diterpenes in glandular trichomes and leaves of Stevia rebaudiana Bertoni**. *Journal of agricultural and food chemistry* 2014, **62**(11):2428-2435.
25. Gastaldo C, Lipko A, Motsch E, Adam P, Schaeffer P, Rohmer M: **Biosynthesis of Isoprene Units in Euphorbia lathyris Laticifers vs. Other Tissues: MVA and MEP Pathways, Compartmentation and Putative Endophytic Fungi Contribution**. *Molecules (Basel, Switzerland)* 2019, **24**(23).
26. Xue Z, Duan L, Liu D, Guo J, Ge S, Dicks J, ÓMáille P, Osbourn A, Qi X: **Divergent evolution of oxidosqualene cyclases in plants**. *The New phytologist* 2012, **193**(4):1022-1038.
27. Cao J, Zhang X, Qu F, Guo Z, Zhao Y: **Dammarane triterpenoids for pharmaceutical use: a patent review (2005 - 2014)**. *Expert Opin Ther Pat* 2015, **25**(7):805-817.
28. Yin J, Yang J, Ma H, Liang T, Li Y, Xiao J, Tian H, Xu Z, Zhan Y: **Expression characteristics and function of CAS and a new beta-amyrin synthase in triterpenoid synthesis in birch (Betula platyphylla Suk.)**. *Plant science : an international journal of experimental plant biology* 2020, **294**:110433.
29. Jin ML, Lee WM, Kim OT: **Two Cycloartenol Synthases for Phytosterol Biosynthesis in Polygala tenuifolia Willd**. *International journal of molecular sciences* 2017, **18**(11).
30. Sharma A, Rather GA, Misra P, Dhar MK, Lattoo SK: **Jasmonate responsive transcription factor WsMYC2 regulates the biosynthesis of triterpenoid withanolides and phytosterol via key pathway genes in Withania somnifera (L.) Dunal**. 2019, **100**(4-5):543-560.

31. Cai C, Ma J, Han C, Jin Y, Zhao G, He X: **Extraction and antioxidant activity of total triterpenoids in the mycelium of a medicinal fungus, Sanghuangporus sanghuang**. *Scientific reports* 2019, **9**(1):7418.
32. Oludemi T, Barros L, Prieto MA, Heleno SA, Barreiro MF, Ferreira I: **Extraction of triterpenoids and phenolic compounds from Ganoderma lucidum: optimization study using the response surface methodology**. *Food & function* 2018, **9**(1):209-226.
33. Wan Q, Tang J, Han Y, Wang D: **Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma**. *Experimental eye research* 2018, **166**:13-20.
34. Yin L, Cai Z, Zhu B, Xu C: **Identification of Key Pathways and Genes in the Dynamic Progression of HCC Based on WGCNA**. *Genes* 2018, **9**(2).
35. Hnasko TS, Hnasko RM: **The Western Blot**. *Methods in molecular biology (Clifton, NJ)* 2015, **1318**:87-96.
36. de Costa F, Yendo AC, Fleck JD, Gosmann G, Fett-Neto AG: **Accumulation of a bioactive triterpene saponin fraction of Quillaja brasiliensis leaves is associated with abiotic and biotic stresses**. *Plant physiology and biochemistry : PPB* 2013, **66**:56-62.
37. Zeng G, Li Z, Zhao Z: **Comparative Analysis of the Characteristics of Triterpenoid Transcriptome from Different Strains of Wolfiporia cocos**. *International journal of molecular sciences* 2019, **20**(15).
38. Fukushima EO, Seki H, Ohyama K, Ono E, Umemoto N, Mizutani M, Saito K, Muranaka T: **CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis**. *Plant & cell physiology* 2011, **52**(12):2050-2061.
39. Suzuki H, Fukushima EO, Umemoto N, Ohyama K, Seki H, Muranaka T: **Comparative analysis of CYP716A subfamily enzymes for the heterologous production of C-28 oxidized triterpenoids in transgenic yeast**. *Plant biotechnology (Tokyo, Japan)* 2018, **35**(2):131-139.
40. Zhang T, Huo T, Ding A, Hao R, Wang J, Cheng T, Bao F, Zhang Q: **Genome-wide identification, characterization, expression and enzyme activity analysis of coniferyl alcohol acetyltransferase genes involved in eugenol biosynthesis in Prunus mume**. *PLoS One* 2019, **14**(10):e0223974.
41. Kim YJ, Zhang D, Yang DC: **Biosynthesis and biotechnological production of ginsenosides**. *Biotechnology advances* 2015, **33**(6 Pt 1):717-735.
42. Wang WF, Xiao H, Zhong JJ: **Biosynthesis of a ganoderic acid in Saccharomyces cerevisiae by expressing a cytochrome P450 gene from Ganoderma lucidum**. *Biotechnology and bioengineering* 2018, **115**(7):1842-1854.
43. Wang J, Li J, Li J, Li J, Liu S, Gao W: **LSP1, a responsive protein from Meyerozyma guilliermondii, elicits defence response and improves glycyrrhizic acid biosynthesis in Glycyrrhiza uralensis Fisch adventitious roots**. *Journal of cellular physiology* 2017, **232**(12):3510-3519.
44. Nguyen KV, Pongkitwitoon B, Pathomwichaiwat T, Viboonjun U, Prathanturarug S: **Effects of methyl jasmonate on the growth and triterpenoid production of diploid and tetraploid Centella asiatica (L.) Urb. hairy root cultures**. *Scientific reports* 2019, **9**(1):18665.
45. Fu J, Liu G, Yang M, Wang X, Chen X, Chen F, Yang Y: **Isolation and functional analysis of squalene synthase gene in tea plant Camellia sinensis**. *Plant physiology and biochemistry : PPB* 2019, **142**:53-

58.

46. Vergara Martínez VM, Estrada-Soto SE, Arellano-García JJ, Rivera-Leyva JC, Castillo-España P, Flores AF, Cardoso-Taketa AT, Perea-Arango I: **Methyl Jasmonate and Salicylic Acid Enhanced the Production of Ursolic and Oleanolic Acid in Callus Cultures of Lepechinia Caulescens**. *Pharmacognosy magazine* 2018, **13**(Suppl 4):S886-s889.
47. Yi GE, Robin AH, Yang K, Park JI, Hwang BH, Nou IS: **Exogenous Methyl Jasmonate and Salicylic Acid Induce Subspecies-Specific Patterns of Glucosinolate Accumulation and Gene Expression in Brassica oleracea L**. *Molecules (Basel, Switzerland)* 2016, **21**(10).
48. Cappellari LDR, Santoro MV, Schmidt A, Gershenzon J, Banchio E: **Improving Phenolic Total Content and Monoterpene in Mentha x piperita by Using Salicylic Acid or Methyl Jasmonate Combined with Rhizobacteria Inoculation**. *International journal of molecular sciences* 2019, **21**(1).
49. Thinon E, Hang HC: **Chemical reporters for exploring protein acylation**. *Biochem Soc Trans* 2015, **43**(2):253-261.
50. Lee S, Kaminaga Y, Cooper B, Pichersky E, Dudareva N, Chapple C: **Benzoylation and sinapoylation of glucosinolate R-groups in Arabidopsis**. *The Plant journal : for cell and molecular biology* 2012, **72**(3):411-422.
51. Peng M, Gao Y: **Evolutionarily Distinct BAHD N-Acyltransferases Are Responsible for Natural Variation of Aromatic Amine Conjugates in Rice**. 2016, **28**(7):1533-1550.
52. Xu Z, Ma J, Qu C, Hu Y, Hao B, Sun Y, Liu Z, Yang H, Yang C, Wang H *et al*: **Identification and expression analyses of the alanine aminotransferase (AlaAT) gene family in poplar seedlings**. *Scientific reports* 2017, **7**:45933.

Figures

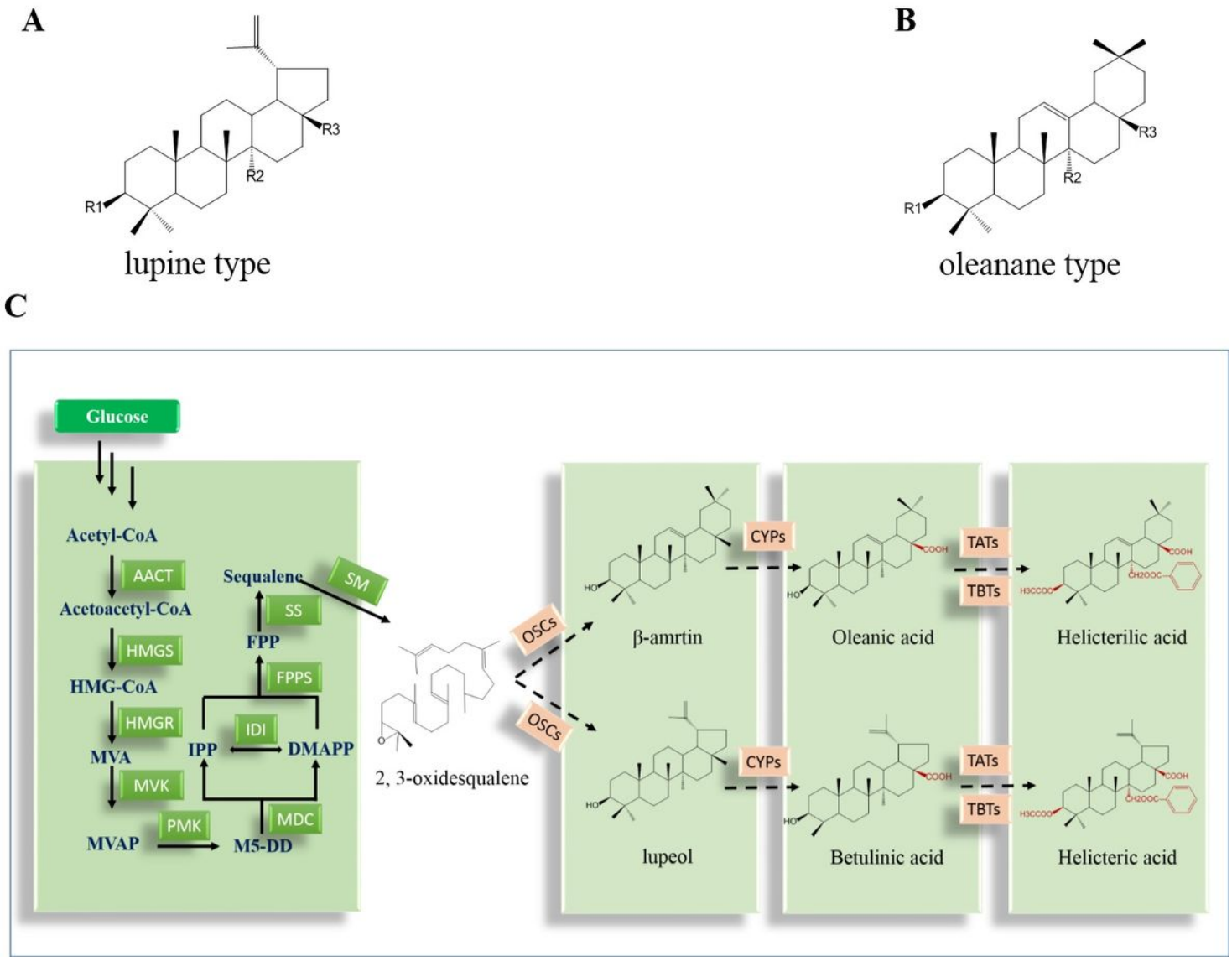

Figure 1

Two parent scaffold types of main triterpenoids and diagram of speculated triterpenoids biosynthetic pathway in H. angustifolia. A Lupine type parent scaffold. B Oleanane type parent scaffold. C Diagram of speculated triterpenoids biosynthetic pathway according to the KEGG database. Enzymes related to the triterpenoids biosynthesis were shown in red frames. The solid arrows represent validated steps and the dotted arrows represent hypothetical steps. AACT: acetyl-CoA C-acetyltransferase; HMGS: hydroxymethylglutaryl-CoA synthase; HMGR: hydroxymethylglutaryl-CoA reductase; MVK: mevalonate kinase; PMK: phosphomevalonate kinase; MDC: diphosphomevalonate decarboxylase; IDI: isopentenyl-diphosphate Delta-isomerase. FPPS: farnesyl pyrophosphate synthase. SS: squalene synthase; SM: squalene monooxygenase; OSCs: Oxidosqualene cyclases; CYPs: cytochrome P450 enzymes; TATs: triterpenoids acetyl transferases; TBTs: triterpenoids benzoyl transferases.

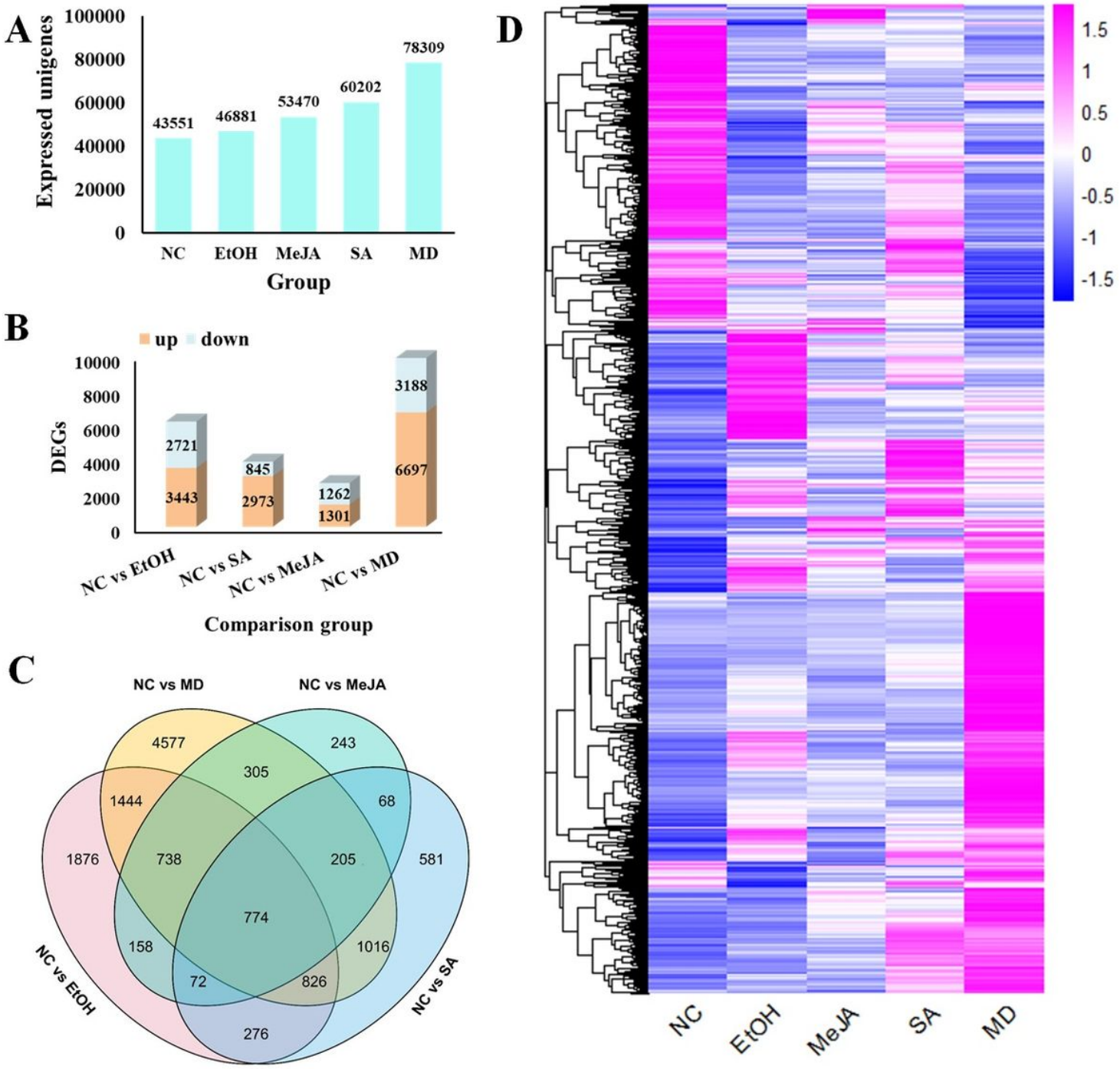

Figure 2

Identification of DEGs among five treatment groups. A Numbers of expressed unigenes between five different groups; B Numbers of DEGs among five treatment groups; C Venndiagram showing the DEGs among five treatment groups; D Heatmap showing the hierarchical cluster of DEGs among five different treatment groups.

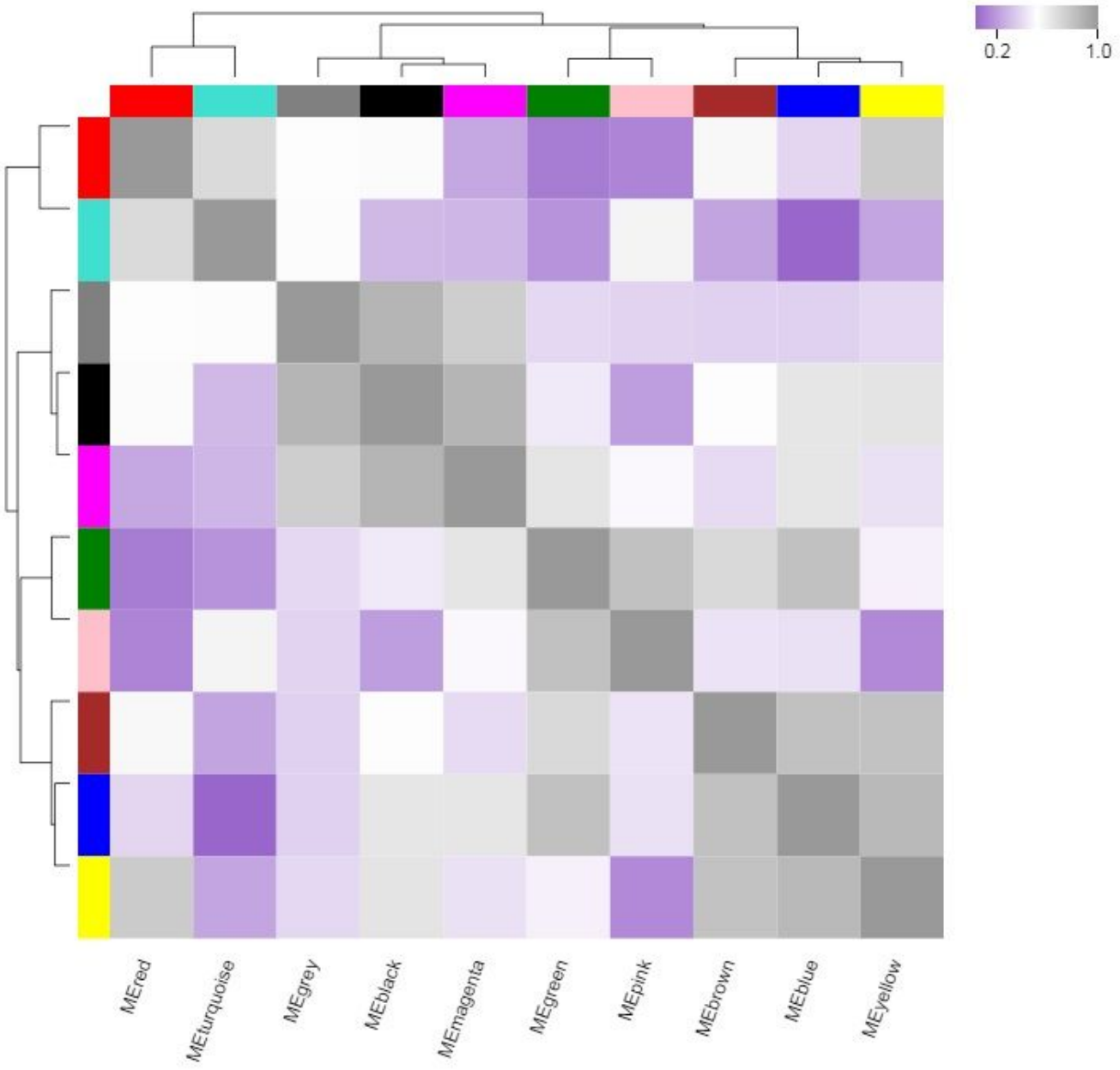

Figure 3

Gene co-expression modules in hormone and mechanical damage-treated transcriptome indicated the cluster hierarchical tree constructed by the eigengenes of the modules and the correlation coefficient between modules with a heatmap.

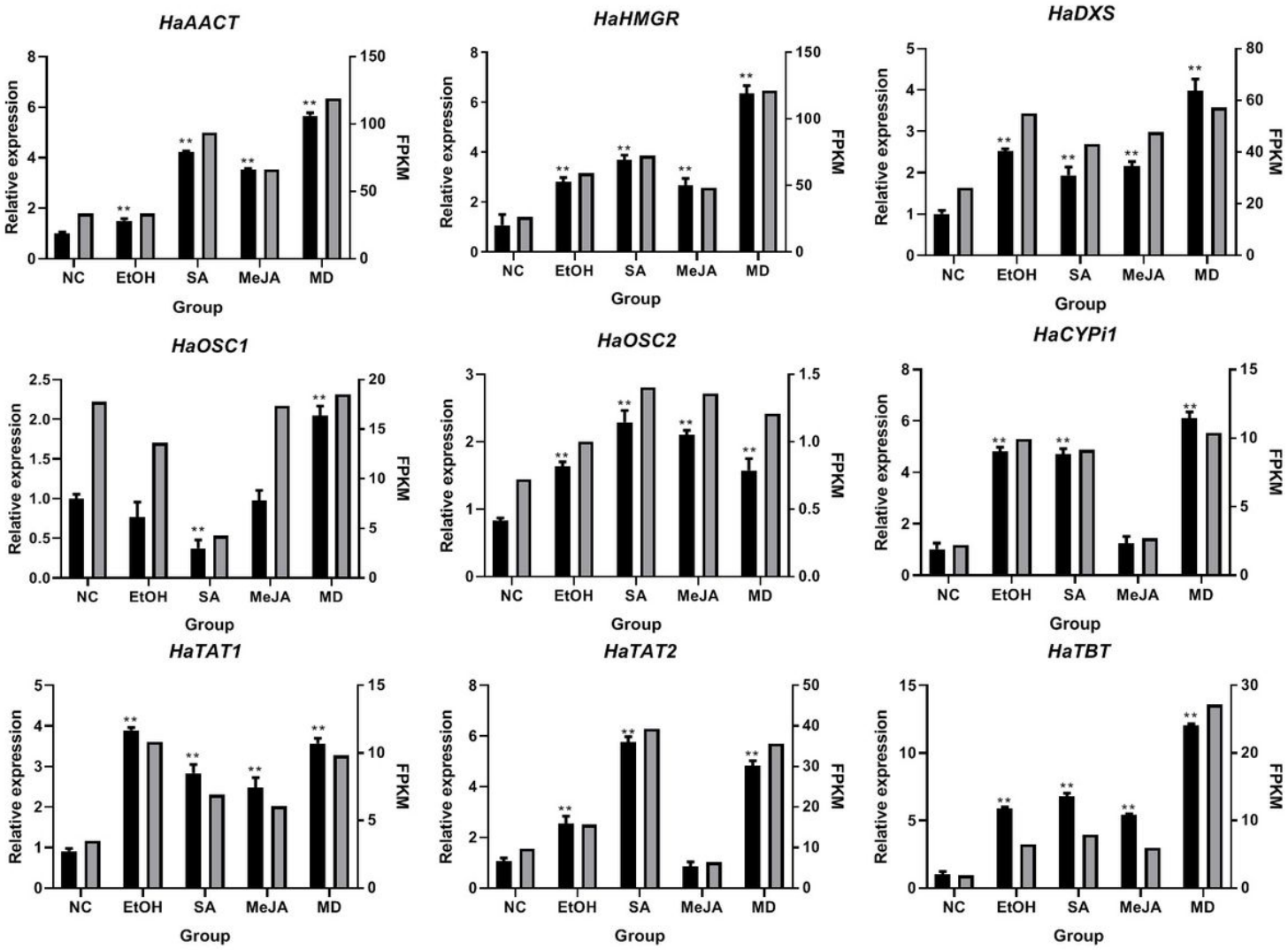

Figure 4

Validation of genes involved in triterpenoid biosynthesis using qRT-PCR. The left Y-axis and black bars represents the relative expression of qPCR, while the right Y-axis and grey bars exhibits the FPKM value of the RNA-Seq data.

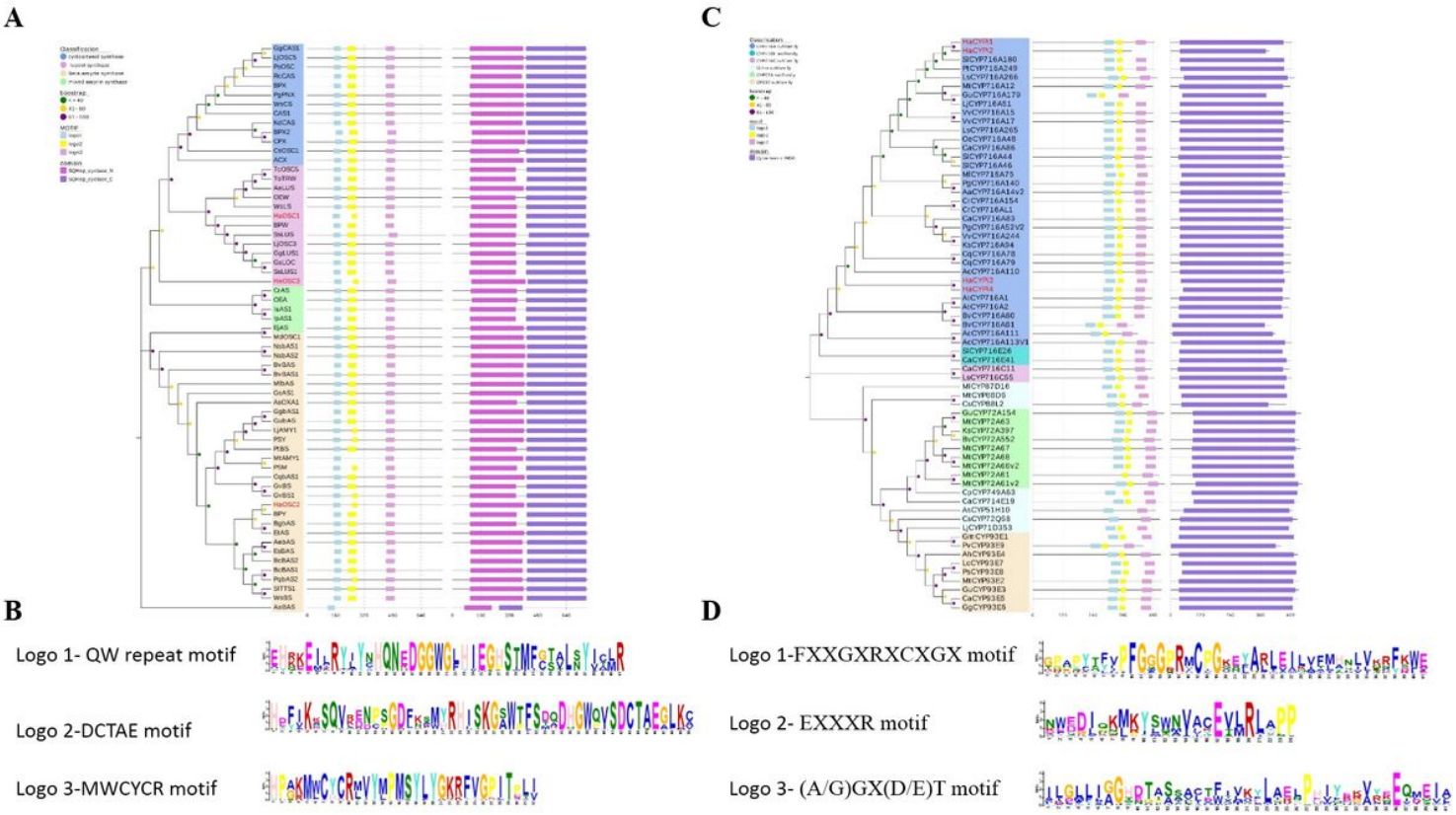

Figure 5

Analysis of amino acid sequences of OSCs and CYP450s. A Phylogenetic tree，conserved domain and conserved motif analysis of OSCs; B Conservative motifs of OSCs corresponding sequences; C Phylogenetic tree，conserved domain and conserved motif analysis of CYP450s; D Conservative motifs of CYP450s corresponding sequences.

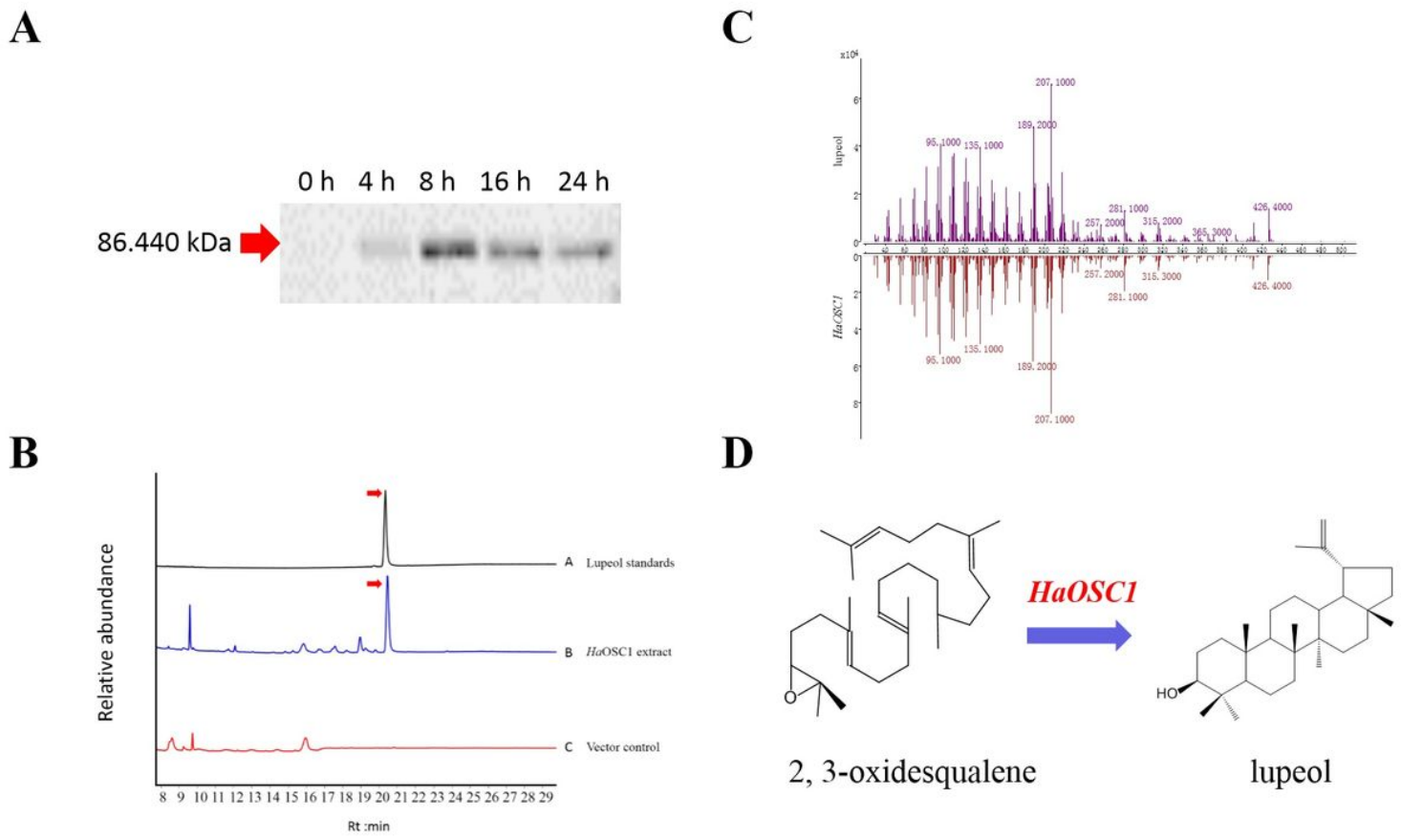

Figure 6

Functional Analysis of HaOSC1. A Westenrn blot analysis of pESC-TRP-HaOSC1 transformants□B extracted ion chromatograms of lupeol standards□A□, yeast cell transformed with expression constructs pESC-TRP- HaOSC1 (B) and empty vector pESC-TRP (C). C the MS data of lupeol and pESC-TRP-HaOSC1. D Schematic diagram of catalytic function of HaOSC1.

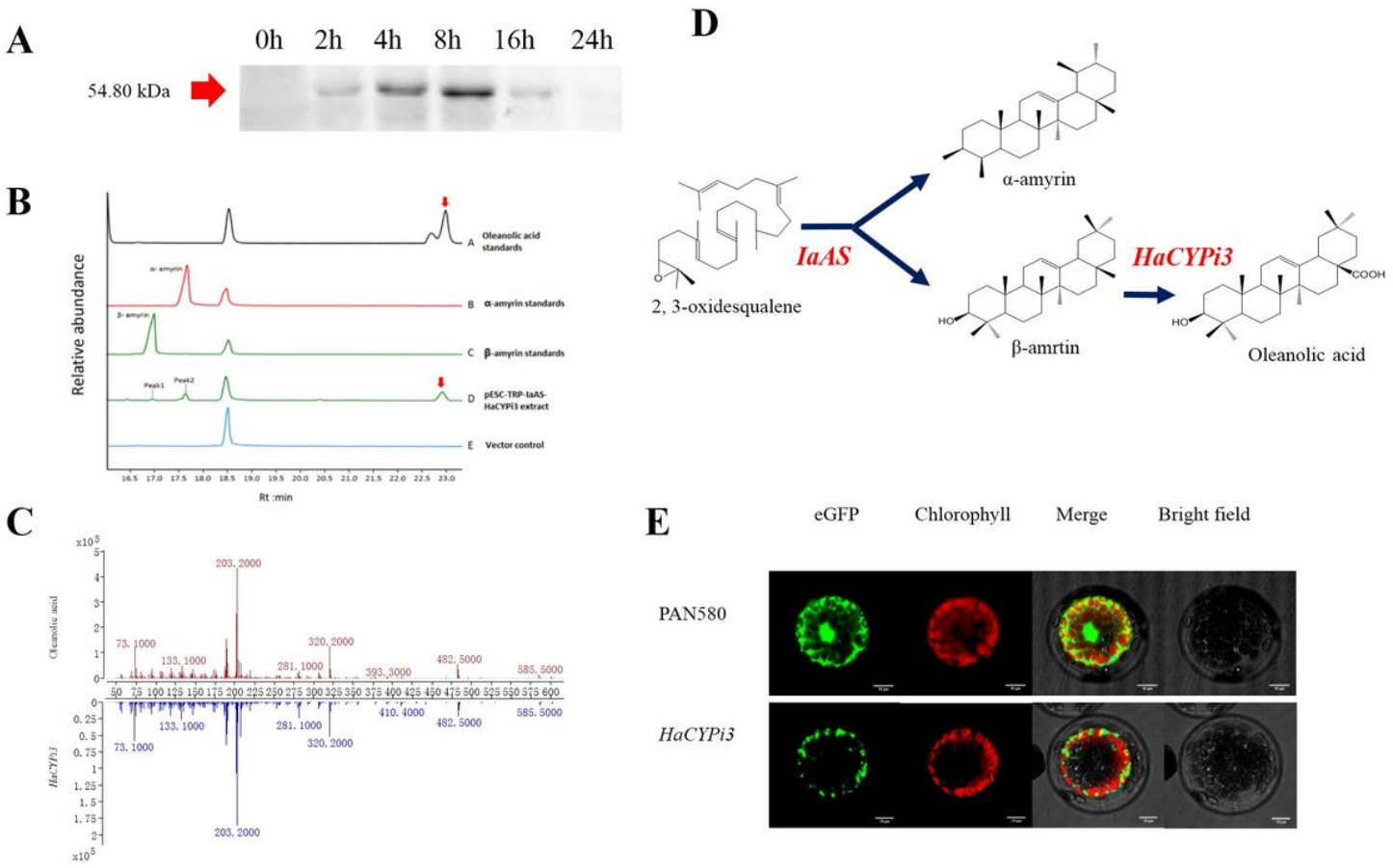

Figure 7

Functional Analysis of HaCYPi3. A Protein expression analysis of HaCYPi3□B extracted ion chromatograms of oleanolic acid standards (A), α-amyrin standards (B), β-amyrin standards (C), yeast cell transformed with expression constructs pESC-TRP-IaAS and pESC-TRP-HaCYPi3 (D) and empty vector pESC-TRP (E). C the MS data of oleanolic acid standards and IaAS-HaCYPi3 co-expression yeast cell extracts. D Schematic diagram of catalytic function of HaCYPi3.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.png
- FigureS2.png
- FigureS3.png
- FigureS4.png
- FigureS5.png
- FigureS6.png
- FigureS7.jpg

- FigureS8.png
- FigureS9.png
- FigureS10.png
- TableS1.doc
- TableS2.doc
- TableS3.doc
- TableS4.doc
- TableS5.doc
- TableS7.doc
- TableS8.doc
- TableS12.doc
- TableS13.doc
- TableS14.doc
- TableS16.doc