Metabolic stimulation-elicited transcripotional responses and biosynthesis of acylated triterpenoids precursors in the medicinal plant Helicteres angustifolia

Yuying Huang†, Wenli An†, Zerui Yang, Chunzhu Xie, Shanshan Liu, Ting Zhan, Huageng Pan and Xiasheng Zheng*

Abstract

Background: 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 saponins. It is worth noting that some acylated 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 is not enough to meet the market requirement. Therefore, the engineering of its metabolic pathway is of high research value for enhancing the production of secondary metabolites. Unfortunately, there are few studies on the biosynthetic pathways of triterpenoids in H. angustifolia, hindering its further investigation.

Results: 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 oxidosqualene 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.

Conclusion: In general, the transcriptomic data of H. angustifolia was first reported and analyzed to study functional genes. Three OSCs, four CYP450s and three acyltransferases were screened out as candidate genes to perform further functional verification, which demonstrated that HaOSC1 and HaCYPi3 encode for lupeol synthase and β-amyrin oxidase, which produce corresponding products of lupeol and oleanolic acid, respectively. Their successful identification revealed pivotal steps in the biosynthesis of acylated triterpenoids precursors, which laid a foundation for further study on acylated triterpenoids. Overall, these results shed light on the regulation of acylated triterpenoids biosynthesis.

© The Author(s) 2022. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.
Introduction

*Helicteres angustifolia*, a conversant plant in *Helicteres* within the family Malvaceae, is a traditional medicinal herb used in Southern China [1]. Ethanolic 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 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 lupane type (Fig. 1A) and the oleanane type (Fig. 1B). Intriguingly, some of these pentacyclic triterpenoids, like helicteric acid and helicterilic acid, are decorated with acyl groups at the C-3 and/or C-27 positions, which are the characteristic components of the genus *Helicteres* [8, 9].

Modern pharmacological studies indicated 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. A significant increase in the content of target products in native plants through biosynthesis and gene regulation could be a more effective strategy for addressing this shortage than the traditional extraction approach [15, 16]. However, the complete lack of research on partial key biosynthetic pathways of triterpenoids in *H. angustifolia*, which makes it difficult to improve the content of acylated triterpenoids by means of transcriptional regulation and genetic engineering. Due to the rapid development of high-throughput sequencing, de novo transcriptome analysis provides a powerful tool for screening candidate genes involved in the biosynthesis pathway of target component in plants without genomic reference, which will support further genetic engineering research [17, 18].

Upstream biosynthesis pathways for triterpenoid have been largely clarified [19, 20], mainly through the mevalonate pathway [21, 22] (Fig. 1C). In this pathway, acetyl-CoA is catalyzed by a series of enzymes to 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 prenyl transferases (GPPS, FPPS, and GGPPS). Among them, FPP is an important intermediate for triterpenoid biosynthesis. Two units of FPP can be catalyzed by the squalene synthase (SS) in a tail-to-tail manner to yield the hydrocarbon squalene. Subsequently, another important precursor 2,3-dihydroxy-3-methyl-2-buta-noic acid (DMBA) is converted into cholesterol by the mevalonate pathway. The subsequent elongation and isomerization of squalene yield a series of minor triterpenoids in *H. angustifolia*. The biosynthetic pathway for triterpenoids in *H. angustifolia* is summarized in Fig. 1C.
3-oxidosqualene, is generated under the catalysis of the squalene monooxygenase (SM) with the presence of O2 and the coenzyme NADPH [23, 24]. 2, 3-oxidosqualene can then be derived into various triterpenoid skeletons under the catalysis of oxidosqualene cyclases (OSCs) [25]. Based on those known compounds and intermediates, downstream synthetic routes of triterpenoids in H. angustifolia were speculated. As showed in Fig. 2, 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, cytochrome P450 enzymes (CYP450s) are responsible for modifying lupeol and β-amyrin by C-28 oxidation to form betulinic acid and oleanolic acid. The acylation reaction is presumably as the final step in the modification of those 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 precursors like betulinic acid and oleanolic acid (Fig. 1C).

Triterpenoids are important metabolic substances [26, 27], and their biosynthesis can be affected by the metabolic stimulation of their source plants, including exogenous phytohormone treatment and mechanical damage [28, 29]. In this study, transcriptomes of *H. angustifolia* samples under different treatments of metabolic stimulation were sequenced and analyzed to obtain a genetic basis 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, 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.

**Materials and methods**

**Plant material and metabolic stimulation treatment**

*H. angustifolia* was collected in Guanzhu Town, Maoming City, Guangdong Province in 2015, and then planted in the medicinal botanical garden of Guangzhou University of Chinese Medicine (GZUCM). *H. angustifolia* used in this experiment were all collected from the medicinal botanical garden of GZUCM. Permission was not necessary for collecting these species, which have not been included in the list of national key protected plants. The formal identification of the plant material was undertaken by Prof. Chaomei Pan (GZUCM). A voucher specimen of *H. angustifolia* was deposited at the Chinese medicine herbarium of GZUCM (voucher number 440923LY0127). Healthy leaves of cultivated *H. angustifolia* with a length of 5 cm and a width of 1.5 cm were selected as experimental samples, which were rinsed with water, removing sundries on the surface of leaves, and excess water was sucked up by absorbent papers. 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 h of treatment, the content of total triterpenoids in *H. angustifolia* was analyzed, and the optimal treatment concentration was determined. Fifteen individual plants with the same growth stage were separated into 5 groups. The mechanical damage (MD) group was scratched with a scissor for one-third of the blade area of its leaves. The MeJA and SA groups were sprayed with the optimal concentration of MeJA and SA on their leaves, respectively. The solvent control group was treated the same as MeJA group, only with a different solvent of EtOH. The negative control group (NC) contains three plants without any treatment. Leaves from these five groups were sampled 24 h after treatment and placed in liquid nitrogen for total triterpenoids content determination and RNA extraction.

**Chemical analyses of total triterpenoids**

Procedures of total triterpenoids extraction and measurement were adapted from Cai et al. [30] and Oludemi et al. [31]. Leaves were sampled and extracted with 80% ethanol at a material-to-liquid ratio of 1:30 and an ultrasonic extraction time of 2 h at 70 °C. Then the dry extract was redissolved in an 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 utilized 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 triterpenoids extraction and measurement were performed in independent triplicates.

**RNA extraction and transcriptome sequencing**

Total RNA of 15 samples (Fig. S1) was 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 sent to Majorbio (Shanghai, China) to construct the cDNA library and then be sequenced on an Illumina HiSeq4000 (Illumina Inc., San Diego, CA, USA). Paired-end reads were generated and examined with the fastx-toolkit_0.0.14 software (http://hannonlab.cshl.edu/fastx-toolkit/) to assess the quality of those reads.

**Differential expressed genes (DEGs) analysis**

Fragments per kilo bases per million reads (FPKM) was measured to gauge the relative expression levels, and the FPKM value of the transcripts was determined using the software RSEM v1.2.15 (http://deweylab.github.io/RSEM/) with default parameters. The DESeq2 version 1.10.1 software was used to analyze the raw counts based on the negative binomial distribution. Expressions of transcripts in different groups were calculated using the FPKM method. The absolute value of the log2 multiple of the FPKM expression ratio between pairwise groups was greater than 1 and the false discovery ratio was less than 0.05 as the standard. DEGs screening was carried out. KEGG enrichment analysis of the DEGs was performed with the 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 presented in Table S1.

**Mining of candidate genes involved in triterpenoid biosynthesis**

Three approaches were employed to screen 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 performed 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 considered as the putative genes.

**Sequence analysis of candidate enzyme proteins and weighted gene co-expression network analysis (WGCNA)**

Those candidate gene sequences verified by sequencing were analyzed for the open reading frame (ORF) through the CLC software and were translated into amino acid sequences. The online ExPASy-ProtParam tool (https://web.expasy.org/protparam/) was used to predict physical and chemical properties of those proteins. SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) and SWISS-MODEL (https://swissmodel.expasy.org) online tools were employed to predict secondary and tertiary structure of target proteins. Conserved motif and domain of candidate enzyme proteins were analyzed using the online analysis software MEME (http://meme-suite.org/tools/meme) and Pfamscan (https://www.ebi.ac.uk/Tools/pfa/pfamscan/). Phylogenetic analysis of protein sequences was carried out by the MEGA 7.0 software. All selected protein sequences were firstly aligned and exported in the MEGA format. Then the phylogenetic tree was constructed by the Neighbor-joining (NJ) method with the Poisson model using the MEGA alignment output. The bootstrap value was set as 1000 with other parameters as default. The online tool Evolview (https://www.evolgenius.info/evolview/#login) was used to edit the evolutionary tree. Expression data was normalized by square root transformation, then be used to infer co-expression gene network modules employing the WGCNA R package according to step-by-step network construction and the module detection method, soft-thresholding method was used to select a proper power-law coefficient β, and a dynamic hierarchical tree cut algorithm was used to detect the co-expression modules [32, 33].

**Recombinant expressions**

The open reading frame (ORF) of candidate genes was amplified from cDNA, using primers summarized in Table S2. The resulting amplicon was conducted to the construction of recombinant vector, which was subsequently transferred into _Escherichia coli_ for preservation after sequencing verification. ORFs of candidate OSCs and CYP450s were amplified with gene-specific primers that contained _SalI_ and 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 those cloned genes through Sanger sequencing, recombinant plasmids were transformed into _S. cerevisiae_ WAT11 by lithium acetate (LiAc) transfer method, adapting specific operation referred to the previous literature [34]. After being induced by 2% D-galactose, the recombinant yeast activates the GAL10 and GAL1 promoter on the pESC-TRP vector, thus inducing the expression of the target (6×His)-tagged proteins. The western-blot method was used to detect the expression level of the fusion protein at different time points to determine the optimal induction time, through specific methods referred to in the literature reported previously [35]. And ORFs of TATs and TBT were subcloned into an N-terminal 6×His tag fusion expression vector pCOLD-TF. Upon sequence confirmation, recombinant plasmids were then transferred into _E. coli_ Rosetta (DE3) for heterologous expression. All the gene-specific forward and reverse primers were displayed in Table S3.

**Subcellular location of HaCYPi3 proteins**

Full-length putative _HaCYPi3_ gene sequences without stop codons were fused with an enhanced green fluorescent protein (EGFP) and ligated into the pAN580 vector. The recombinant vectors were then transferred into Arabidopsis protoplasts using the polyethylene glycol (PEG). Protocols used for protoplast isolation and recombinant vector transfer protocols were as described in the previous study [36]. EGFP fluorescent signals were observed using a Zeiss laser scanning microscope (LSM) 800 (Zeiss, Germany).

**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 dropped, the pH value was adjusted to 2.0 with concentrated hydrochloric acid. Yeast lysate was extracted three times with an 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°C for 60 min before GC-MS analysis. The GC system was 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 a temperature of 250°C. The GC oven temperature was programmed from 80°C (held for 1 min) to 300°C at 40°C/min and kept for 10 min. For the identification...
of metabolites, complete mass spectra was generated by scanning within the mass-to-charge ratio range of 50 to 600. Triterpenoid products in yeast were monitored according to the base peak. Empty pESC-TRP vector was used as the negative control, and standard substance was used as the positive control. Triterpenoid was determined by comparing both the retention time and mass spectra with the authentic standards.

For the characterization of the acyltransferase, an enzymatic assay 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) was carried out in a final volume of 200 μL. The reaction system was placed at 30°C for 2 h. After the reaction, 500 μL ethyl acetate was added to extract twice and the extracted solution was combined. The residue was redissolved with 200 μL acetonitrile and centrifuged at 16,000 rpm for 5 min, with the supernatant isolated for subsequent HPLC analysis. The HPLC equipment was equipped with a Hypersil Gold AQ-C18 column (250 × 4.5 mm × 5 μm). 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). The elution procedure was set as followed: 0–60 min, 80% B, 60–75 min, 80–100% B, with a flow rate of 1.2 mL/min. The detection wavelength was set at 210 nm.

**Statistical analysis**

Three biological replicates were carried out for each experiment, and statistical analysis was performed using the IBM SPSS Statistics 23 statistical software. A Shapiro-Wilk test was used to test whether the data was normally distributed. If yes, the t-test of independent samples was used for pairwise comparison. One-way Analysis of Variance was used for comparison between groups, in which homogeneity of Variance was tested for the data, assuming that the Variance was not homogenous. A significance P > 0.05 was used to indicate homogeneity of variance. The Bonferroni comparison test was used. The Dunnett’s T3 multiple comparison test was used when in heterogeneity variance, and the data obtained were stated as (mean ± standard) deviation, with a P < 0.05 indicated statistically significant differences. If the single-factor data did not follow normal distribution, Kruskal Wallis single-factor ANOVA (K sample) and Mann-Whitney U (two samples) rank sum test were used. A P < 0.05 indicated that the difference was statistically significant. Data statistics were performed using the Graph Pad Prism 8.0 software.

**Results**

**Transcriptome analysis induced by metabolic stimulation**

The optimal concentration of 500 μM methyl jasmonate (MeJA) and 400 μM salicylic acid (SA) was determined by the single factor experiment (Fig. S2, Table S4). Consequently, the total triterpenoids content of samples from five groups was measured. The results showed that elevated contents of total triterpenoids in the MD-, MeJA-, and SA-induced groups were observed, compared to the NC group (Fig. S3, Table S5). This result was consistent with other research, suggesting that the synthesis of triterpenoids could be affected by metabolic stimulation [37]. 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 those 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 were characterized in Table S6. 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 length N50 of 1300 bp (Table 2).

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 (Fig. S4, Table S7). 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 (Fig. S5).

**Analysis of DEGs under different treatments**

The statistics of expressed genes showed that the control group had the least number of genes, while the MD group had the most (Fig. 2A). Then transcriptome statistics of those five groups were compared pairwise to establish potential DEGs among different groups. Compared to the NC group, there was the greatest amount of DEGs in the MD group (Fig. 2A), and the number of up-regulated DEGs in each metabolic stimulation treatment group

| Items                  | Number   |
|------------------------|----------|
| Total sequence base    | 381,166,620|
| Total transcripts      | 581,934  |
| Total unigenes         | 424,824  |
| Largest (bp)           | 16,851   |
| Smallest (bp)          | 201      |
| Length N50 (bp)        | 1294     |
| Mean Length(bp)        | 655      |
and EtOH group were higher than that of the down-regulated genes (Fig. 2B) However, there is a very interesting phenomenon that EtOH group appears to be more influential on gene expression than either of the hormone treatments. Ethanol was selected as the solvent because ethanol treatment had no significant effect on the content of total triterpenoids. This phenomenon revealed that although ethanol has no significant effect on the expression of genes related to triterpenoids pathway, it may affect the expression of other genes. The Venn diagram presented in Fig. 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 (Fig. 2D). KEGG enrichment analyses of DEGs in four pairwise groups (Fig. S6, Tables S8, S9, S10, S11) suggested that pathways with the most enrichment were mostly associated with triterpenoid synthesis, revealing that the expression of genes related to triterpenoid synthesis was up-regulated after metabolic stimulation, leading to the enhancement of triterpenoid accumulation for resistance, which is consistent with the results of previous research [29, 38].

In addition, weighted gene co-expression network analysis (WGCNA) was conducted to further investigate potential genes involved in the triterpenoid 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 with those DEGs, resulting in 10 co-expression modules. These modules were named according to their colors, namely the 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) module (Fig. 3). Genes within each module were selected for enrichment in

![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](image-url)
the KEGG pathway. Statistically significantly enriched genes ($P < 0.05$) were filtrated for further analyses. Results showed that five modules (Blue, turquoise, black, green and magenta), consisting of 4936 genes in total, were related to triterpenoid biosynthesis.

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

Based on transcriptome annotation results, we screened 41 genes, whose annotations were related to triterpenoid skeleton synthesis according to KEGG pathway, and their expression patterns (Fig. S7, Table S12) were analyzed thereafter. Among them, 18 genes presented significantly different expression levels among five treatment groups, nine of which were chosen to validation on their expression level using qRT-PCR, aiming to further verify the reliability of our RNA-Seq data. As shown, almost all of these genes were significantly increased after treatment with hormones and mechanical damage, which is consistent with the RNA-Seq data in expression (Fig. 4, Table S13), confirming the accuracy of our transcriptome data.

**Cloning and sequence analysis of candidate genes**

In order to elucidate the biosynthetic pathway of acylated triterpenoid precursors and shed light on acylated triterpenoids biosynthesis in *H. angustifolia*, putative key enzymes encoding genes, including three OSCs (named *HaOSC1*, *HaOSC2* and *HaOSC3*) and four CYP450s (named *HaCYPi1*, *HaCYPi2*, *HaCYPi3* and *HaCYPi4*), were targeted for functional analysis. Their corresponding protein sequences were summarized and exhibited in Table S14 and physicochemical properties of these proteins were documented in the Table S15, two-dimensional structure information of these proteins was classified in the Table S16, while three-dimensional structures were displayed in Fig. 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, while *HaOSC2* was classified as a β-amyrin synthase (Fig. 5A, Table S17). In addition, the conserved domain and motif analysis of candidate HaOSCs showed that they had high similarity to the plant oxidosqualene...
cyclase (OSC) family. They all contained conserved SQHOP-cyclase-N and SQHOP-cyclase-C domains with three conserved motifs, including DCTAE, MWCYCR and QW repeat (Fig. 5B). Among them, the DCTAE motif might be crucial for the interaction between the protein and its substrate [38]. The QW repeat is a negatively charged aromatic region that equalizes carbocation in the cyclization reaction, which may play a role in the structural stabilization of proteins [39].

The results of phylogenetic analyses also implied that four CYP450 candidate genes were well documented into the CYP716A subfamily (Fig. 5C, Table S18), which primarily includes CYP450 enzymes with a C-28 oxidation function [40, 41]. Conserved cytochrome P450 domain (heme active sites) shared by candidate CYP450 genes were found except for HaCYPi2, including the oxygen-binding domain (A/G)GX(D/E)T, the K helix (EXXXR), and the C-terminal heme binding domain (FXXGXRXCXGX) (Fig. 5D). Among them, the heme binding domain is the main characteristic structure for the identification of P450 protein [42].

**Heterologous expression and functional characterization of three OSCs**

Three candidate OSCs were functionally validated by expressing their ORFs in the expression vector pESC-TRP. Recombinant plasmids were transferred into yeast (S. cerevisiae strain WAT11) by the lithium acetate conversion method. Protein expression of the recombinant yeast within 24 h was investigated, which implied that the recombinant yeast pESC-TRP-HaOSC1 and
pESC-TRP-HaOSC2 successfully expressed the protein with molecular mass of about 86 kDa (Fig. 6A) and 87 kDa (Fig. S9) except for the recombinant yeast pESC-TRP-HaOSC3, and the protein expression reached the highest level 8 h after induction. 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 enzymatic products with authentic standards (α-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 (Fig. 6B-C). The yeast expression experiments were repeated three times, always with very similar results. These results clearly manifested that HaOSC1 encodes a lupeol synthase, which could use 2, 3-oxidosqualene in yeast as the substrate to synthesize the target product lupeol (Fig. 6D).

**Functional validation of four 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 was observed. The induction time was also investigated for HaCYPi3, with a result that the protein expression reached the highest level at about 8 h after induction (Fig. 7A).

Based on the previous sequence analysis, HaCYPi3 was predicted to be a potential C-28 oxidase, which might use α-amyрин, β-amyрин and lupeol as substrates. Therefore, we firstly transferred pESC-HaCYPi3 and pESC-HaOSC1 recombinant plasmids 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-amyрин synthase IaAS from *Ilex asprell* that is identified in our previous work. 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 the new product were consistent with that of the oleanolic acid authentic standards.
Consistene results were gained by three repeated experiments, which clearly demonstrated that HaCYPi3 encoded a β-amyrin oxidase, which could use the β-amyrin catalyzed by IaAS in yeast as a substrate to synthesize the target product oleanolic acid (Fig. 7D). Furthermore, the subcellular localization of HaCYPi3 was observed, and results displayed that it was localized in the cytoplasm (Fig. 7E), which is consistent with reports that triterpenoids were synthesized in the cytoplasm [43].

Discussions

Triterpenoids, as a class of natural products with extensive biological activities, have attracted the attention of many researchers. Currently, the biosynthetic pathways of many triterpenoids have been successfully elucidated, such as ginsenosides [44], ganoderic acid [45] and glycyrrhizic acid [46]. However, acylated triterpenoids, especially benzoylated triterpenoids, are uncommonly seen and only present in plants belonging to the Helicteres genus. Hence, H. angustifolia can be used as an ideal research material to elucidate the biosynthetic 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 [47–49]. 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 [50, 51]. As a means of abiotic stress, mechanical damage can effectively regulate the growth, development, stress response and metabolites of plants [27]. In this study, the content of total triterpenoids in H. angustifolia was analyzed after different concentration hormone stimulation treatments, with the optimal treatment conditions investigated. Furthermore, transcriptome sequencing of H. angustifolia under the optimized treatment conditions was carried out to screen genes and assisted in revealing the biosynthetic pathway at molecular level, which provided theoretical basis for the effective development and utilization of triterpenoids in H. angustifolia.

One lupeol cyclase and one β-amyrin oxidase were successfully screened and identified by means of
metabolic stimulation combined with transcriptomics, and corresponding products lupeol and oleanolic acid were successfully synthesized in *S. cerevisiae*, which clearly reveal the biosynthesis pathways of some important precursors of acylated triterpenoids. Furthermore, acylation is the final and most critical step in the biosynthesis of acylated triterpenoids. Functional identification of the corresponding acyltransferases including acetyl transferases and benzoyl transferases is the target of our research group in the future.

So far, we have successfully spotted three candidate acyltransferases that belong to the BAHD acyltransferase family through phylogenetic analysis (Fig. S10A, Table S16). This family is named after the initials of four representative enzymes: benzylalcohol acetyltransferase, anthocyanin ohydroxycinnamoyltransferase, anthranilate N-hydroxycinnamoyl/benzoyltransferase, and deacetylvindoline acetyltransferase [52, 53]. Domain analysis showed that the transferase family domain existed in these three candidates. And they all possess three amino acid motifs, HXXXD, YFGNC, and DFGWG (Fig. S10B), which are highly conserved in BAHD enzymes. The HXXXD motif is located in the center of the reaction channel, which is considered to facilitate catalytic reactions to form amide or acylated esters [54]. The DFGWG motif is proposed to play a structural role and has a significant impact on the integrity of the CoA binding pocket [55].

Then, prokaryotic expression vectors carrying three candidate acyltransferases coding genes were constructed, and the resulting 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 (Fig. S11). Unfortunately, functions of the three candidate enzymes failed to be characterized via in vitro enzyme activity test. We hypothesized that prokaryotic expression could not provide appropriate post-translational modifications for these proteins, leading to failure of enzyme activity responses in vitro. Therefore, further investigation on their function shall be carried out through the combination of eukaryotic expression and substrate feeding in the future.

**Conclusion**

*Helicteres angustifolia*, a shrub distributed in Southern China and be used in traditional Chinese medicine, is rich in triterpenoids such as betulinic acid, oleanolic acid, hederic acid, hederilic acid, and similar derivatives. The biosynthetic pathway of these compounds is of great value due to their valuable medicinal activity. Here, we first reported the transcriptomic study of *H. angustifolia* for the exploration of functional genes. Three OSCs, four CYP450s and three acyltransferases were screened out as candidate genes. Their functional verification demonstrated that HaOSC1 and HaCYP13 encode for a lupeol synthase and β-amyrin oxidase, respectively. Corresponding products of lupeol and oleanolic acid, the important precursors of acylated triterpenoids, were synthesized successfully. These results in this study built a basis for future analyses of the biosynthetic pathway of acylated triterpenoids, and laid a foundation for future genetic engineering to increase the yield of medicinal triterpenoids in *H. angustifolia*.

**Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s12870-022-03429-8.

---

**Additional file 1:** Figure S1. The electrophoresis diagrams of RNA extracted from different treatment groups of *H. angustifolia* tree leaves.

**Additional file 2:** Figure S2. Investigation of hormone concentration.

**Additional file 3:** Figure S3. Comparisons of the total triterpenoid content among different treatment groups.

**Additional file 4:** Figure S4. Functional annotation of the transcripts and unigenes against six databases.

**Additional file 5:** Figure S5. Venn diagram indicating the numbers of common and specific expressed unigenes among six databases.

**Additional file 6:** Figure S6. KEGG enrichment results of the DEGs among negative control group and different treatment groups.

**Additional file 7:** Figure S7. The expression pattern of key related genes involved with triterpenoids biosynthesis were shown using a heatmap.

**Additional file 8:** Figure S8. Three dimension structure of target gene coding proteins.

**Additional file 9:** Figure S9. Protein expression analysis of HaOSC2.

**Additional file 10:** Figure S10. Analysis of amino acid sequences of TATS (Triterpenoid acetyltransferase) and TBT (Triterpenoid benzoyl transferase).

**Additional file 11:** Figure S11. SDS-PAGE analysis.

**Additional file 12:** Table S1. Primers for qRT-PCR.

**Additional file 13:** Table S2. Primers for amplification of ten putative genes.

**Additional file 14:** Table S3. Gene-specific forward and reverse primers.

**Additional file 15:** Table S4. Effect of different methyl jasmonate and salicylic acid concentrations and on total triterpenoids content.

**Additional file 16:** Table S5. Efects of different metabolic stimulation treatments on total triterpenoids content.

**Additional file 17:** Table S6. Transcriptome sequencing data of the *H. angustifolia*.

**Additional file 18:** Table S7. The overview statistics of annotation.

**Additional file 19:** Table S8. KEGG enrichment of DEGs from the NC vs EtOH comparison.

**Additional file 20:** Table S9. KEGG enrichment of DEGs from the NC vs SA comparison.
Additional file 21: Table S10. KEGG enrichment of DEGs from the NC vs MeJA comparison.

Additional file 22: Table S11. KEGG enrichment of DEGs from the NC vs MD comparison.

Additional file 23: Table S12. Relative expression of genes involved in triterpenoids biosynthesis.

Additional file 24: Table S13. The relative expression levels of nine related genes in different treatment groups.

Additional file 25: Table S14. Nucleotide sequence of ten candidate genes.

Additional file 26: Table S15. Analysis of physicochemical properties of target gene coding proteins.

Additional file 27: Table S16. Two-dimensional structure information of target gene coding proteins.

Additional file 28: Table S17. Summary table of OSC genes.

Additional file 29: Table S18. Summary table of CP450 genes.

Additional file 30: Table S19. Summary table of BAHD genes.

Acknowledgements

We would like to thank Professor Aijia Ji from Guangzhou University of Chinese Medicine and Professor Caixia Wang from Institute of Chinese Materia Medica China Academy of Chinese Medical Sciences for the sharing of multiple plasmids used in this study.

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. ZBY participated in bioinformatics analysis, SSL, CZX and TZ participated in experiments and wrote the manuscript.

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.

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/).

Declarations

Ethics approval and consent to participate

H. angustifolia were collected in Guanzhu Town, Maoming City, Guangdong Province in 2015, and then planted in the medicinal botanical garden of Guangzhou University of Chinese Medicine (GZUCM). H. angustifolia used in this experiment were all collected from the medicinal botanical garden of GZUCM. Permission was not necessary for collecting these species, which have not been included in the list of national key protected species. Voucher specimen of H. angustifolia was deposited at the Chinese medicine herbarium of GZUCM. Experimental research on H. angustifolia complies with GZUCM guidelines (https://zcsy.gzucm.edu.cn/index.html), and does not include the genetic transformation of the same, preserving the genetic background of the species used. It does not require ethical approval.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflict of interest.

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;172:61–9.

2. Li K, Yu S, Sun S, Liu Y, Garg S, Kaul SC, et al. 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;109:262–70.

5. Pan MH, Chen CM, Lee SW, Chen ZT. Cytotoxic triterpenoids from the root bark of Helicteres angustifolia. Chem Biodivers. 2008;5(4):565–74.

6. Su D, Gao YQ. Heliceric 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;2017:518070.

7. Wang GC, Li T, Wei YR, Zhang YB, Li YL, Sze SC, et al. Two pregne derivatives and a quinolone alkaloid from Helicteres angustifolia. Fitoterapia. 2012;83(8):1643–7.

8. Chang YS, Ku YR, Lin JH, Hu KL, Ho HK. Analysis of three Lupane type triterpenoids in Helicteres angustifolia by high-performance liquid chromatography. J Pharm Biomed Anal. 2001;265(6–5):849–55.

9. Chen L, Liu J, Ge X, Xu W, Chen Y, Li F, et al. Simulated digestion and fermentation in vitro by human gut microbiota of polysaccharides from Helicteres angustifolia L. Int J Biol Macromol. 2019;141:1065–71.

10. Huang Q, Huang R, Wei L, Chen Y, Lv S, Liang C, et al. Antiviral activity of methyl helicterate isolated from Helicteres angustifolia (Sterculiaceae) against hepatitis B virus. Antivir Res. 2013;100(2):373–81.

11. Biggs BW, Lim CG, Sagliani K, Shankar S, Stephanopoulos G, De Mey M, et al. Overcoming heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in Escherichia coli. Proc Natl Acad Sci U S A. 2016;113(12):3209–14.

12. Zhang XL, Chen ZN, Huang QF, Bai FC, Nie JL, Lu SJ, et al. Methyl Helicerate inhibits hepatic stellate cell activation through modulation of apoptosis and autophagy. Cell Physiol Biochem. 2018;51(2):897–908.

13. Lin X, Huang R, Zhang S, Zheng L, Wei L, He M, et al. Methyl helicerate protects against CCl4-induced liver injury in rats by inhibiting oxidative stress, NF-κB activation, Fas/Fasl pathway and cytochrome P450E1 level. Food Chem Toxicol. 2012;50(10):3413–20.

14. Wen S, Wei Y, Zhang X, Bai F, Tan S, Nie J, et al. Methyl heliceritate 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, Jiang 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):817–37.

17. Heathner JM, Chain B. The sequence of sequencers: the history of droplet-based ultra-high-throughput single-cell RNA-Seq systems. Mol Cell. 2019;73(1):130–142.e135.

18. Zhang X, Li T, Liu F, Chen Y, Yao J, Li Z, et al. Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-Seq systems. Mol Cell. 2019;73(1):130–142.e135.

19. Haralampidis K, Trojanowska M, Osbourn AE. Biosynthesis of triterpenoid saponins: biosynthesis, in vitro production, and pharmacological relevance. Protoplasma. 2019;256(6):1463–86.
21. Yang C, Gao X, Jiang Y, Sun B, Gao F, Yang S. Synergy between methy- 
lerithiol phosphate pathway and mevalonate pathway for isoprene 
production in Escherichia coli. Metab Eng. 2016;37:79–91.
22. Kobayashi K, Suzulo M, Muraoka T, Nagata N. The mevalonate pathway 
but not the methylerithiol phosphate pathway is critical for elaioplast 
and pollen coat development in Arabidopsis. Plant Biotechnol (Tokyo, 
Japan). 2018;35(4):381–5.
23. Wolwev-Rick U, May B, Lankes C, Wüst M. Methylerythritol and meva-
lonate pathway contributions to biosynthesis of mono-, sesqui-, and dit-
erpenes in glandular trichomes and leaves of Stevia rebaudiana Bertoni. J 
Agric Food Chem. 2014;62(11):2428–35.
24. Giestz RD, Schiestl RH. High-efficiency yeast transformation using the 
Cry1A subfamily members are multifunctional oxidases in triterpenoid 
Biosynthesis. Plant Cell Physiol. 2011;52(12):2050–61.
25. Xue Z, Duan L, Liu D, Guo J, Ge S, Dicks J, et al. Divergent evolution of 
odiosqualene cyclases in plants. N Phytol. 2012;193(4):1022–38.
26. Wan Q, Tang J, Han Y, Wang D. Co-expression modules construction by 
expression analysis of the DCTA motif, which is highly conserved among oxidos-
qualene cyclases. FEBS J. 2013;280(5):1267–80.
27. 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, Sanghuang-
porus sanghuang. Sci Rep. 2019;9(1):7418.
28. Oludemi T, Barros L, Prieto MA, Heleno SA, Barreiro MF, Ferreira I. Extrac-
tion of triterpenoids and phenolic compounds from Ganoderma luci-
dum: optimization study using the response surface methodology. Food 
Funct. 2018;9(1):209–26.
29. Wang Q, Tang J, Han Y, Wang D. Co-expression modules construction by 
WGCNA and identify potential prognostic markers of uveal melanoma. 
Exp Eye Res. 2018;166:13–20.
30. Yin L, Cai Z, Xu B, Xu C. Identification of key pathways and genes in the 
jasmonate pathway contributions to biosynthesis of mono-, sesqui- and dit-
erpenoid acid biosynthesis in Glycyrrhiza uralensis Fisch adventitious roots. J 
Biol Chem. 2018;293(19):7089–103.
31. de Costa F, Vendo AC, Fleck JD, Gossmann G, Fett-Neto AG. Accumulation 
of a bioactive triterpene saponin fraction of Quillaja brasiliensis leaves is 
associated with abiotic and biotic stresses. Plant Physiol Biochem. 
2015;94:66–74.
32. Ito R, Masukawa Y, Hoshino T. Purification, kinetics, inhibitors and CD for 
recombinant β-amin synthase from Euphorbia tirucalli. I and functional 
analysis of the DCTA motif, which is highly conserved among oxidos-
qualene cyclases. FEBS J. 2013;280(5):1267–90.
33. Vishwakarma RK, Sonawane P, Singh S, Kumar U, Ruby, Khan BM. 
Molecular characterization and differential expression studies of an 
odiosqualene cyclase (OSC) gene of Brahmi (Bacopa monniera). Physiol 
Mol Biol Plants. 2013;19(4):547–53.
34. 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(2):2050–61.
35. Suzuki H, Fukushima EO, Umemoto N, Ohyama K, Seki H, Muranaka T. 
Comparative analysis of CYP716A subfamily enzymes for the heterolo-
gous production of C-28 oxidized triterpenoids in transgenic yeast. Plant 
Biotechnol (Tokyo, Japan). 2018;35(2):131–9.
Publisher’s Note 
Springer Nature remains neutral with regard to jurisdictional claims in published 
maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from: 
• fast, convenient online submission 
• thorough peer review by experienced researchers in your field 
• rapid publication on acceptance 
• support for research data, including large and complex data types 
• gold Open Access which fosters wider collaboration and increased citations 
• maximum visibility for your research: over 100M website views per year 

At BMC, research is always in progress. 
Learn more: biomedcentral.com/submissions