scientific reports

Check for updates

OPEN

Integrated transcriptomic and targeted triterpenoid profiling reveals key enzymes in triterpenoid biosynthesis of *Oplopanax elatus*

Hong Ju Choi[1], Ji Won Seo[1], Jiu Park[1], Won Hyeok Choi[1], Myong Jo Kim[2] & Eun Soo Seong[2]✉

Oplopanax elatus **(O. elatus) is an endangered medicinal plant containing pharmacologically important triterpenoids; however, the biosynthetic mechanisms underlying their accumulation remain largely unknown. This study aimed to elucidate the metabolic and genetic basis for triterpenoid production by comparing root tissues at 0 weeks (0 W) and regenerated plantlets cultured for 8 weeks (8 W). Quantitative HPLC analysis revealed that the contents of lupeol, oleanolic acid, and betulin were markedly higher in 8 W regenerated tissues than in 0 W roots. To investigate the molecular basis of this increase, we performed comparative transcriptomic analysis and identified differentially expressed genes associated with triterpenoid biosynthesis. Among oxidosqualene cyclases (OSCs), beta-amyrin synthase (Gene_22342T) and lupeol synthase (Gene_05624T) showed threefold and 30-fold higher expression at 8 W, respectively, which was further validated by qRT-PCR. Phylogenetic analysis, conserved motif characterization, and protein–ligand docking confirmed that these genes function as key enzymes in triterpenoid skeleton formation. Our findings provide the first integrated biochemical and transcriptomic evidence explaining triterpenoid accumulation in regenerated O. elatus tissues and offer foundational insights for metabolic engineering and conservation-based utilization of this endangered plant.**

Keywords *Oplopanax elatus*, Triterpenoid biosynthesis, Lupeol synthase, Beta-amyrin synthase, Transcriptome analysis

Triterpenoids are among the most diverse groups of natural compounds, with over 20,000 triterpenoids identified in nature to date[1]. More than 200 unique triterpene skeletons derived from natural sources or enzymatic reactions have been reported, with tetracyclic and pentacyclic triterpenoids being the most common[2]. Beta-amyrin and lupeol belong to the pentacyclic triterpene family[3]. Beta-amyrin is a low-abundance secondary metabolite; its extraction from plants typically results in low yields, necessitating significant consumption of natural resources[4]. Additionally, because of its complex structure and the need for complex synthetic pathways for its production, chemical synthesis is not recommended[5]. Beta-amyrin has pharmacological potential as an antitumor, anti-inflammatory, anxiolytic agent, and hepatoprotective agents[6]. Lupeol also has a low yield, limiting its commercial application, and its extraction and separation methods are challenging[7]. Lupeol has protective effects against various types of cancer, diabetes, obesity, cardiovascular disease, kidney and liver diseases, skin diseases, and neurological diseases, attracting the attention of researchers[8].

Triterpenoid biosynthesis begins with the production of isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). IPP and DMAPP are synthesized via two pathways: the mevalonate (MVA) pathway in the cytoplasm and the 2-C-methyl D-erythritol-4-phosphate (MEP) pathway. The MVA pathway uses acetyl-CoA as an initial substrate and undergoes six condensation reactions to produce IPP[9]. Subsequently, IPP and DMAPP are condensed by geranyl pyrophosphate synthase (GPS) to produce geranyl pyrophosphate (GPP), which is catalyzed by farnesyl pyrophosphate synthase (FPS) to produce farnesyl pyrophosphate (FPP) with the addition of a second IPP unit[10]. The condensation of two FPP molecules by squalene synthase (SQS) results in the formation of squalene, which is converted to 2,3-oxidosqualene via epoxidation. Lastly, 2,3-oxidosqualene undergoes cyclization mediated by oxidosqualene cyclase (OSC) followed by oxidation mediated by cytochrome P450 enzymes[11]. The first diversification step in triterpenoid biosynthesis is the OSC-catalyzed cyclization of 2,

[1]Interdisciplinary Program in Smart Agriculture, Kangwon National University, Chuncheon 24341, Korea. [2]Department of Applied Plant Sciences, Division of Bioresource Sciences, Kangwon National University, Chuncheon 24341, Korea. ✉email: esseong@kangwon.ac.kr

3-oxidosqualene[12]. Approximately 152 OSCs have been identified from various plants that catalyze the formation of triterpene scaffolds. The most abundant functionally confirmed OSCs are beta-amyrin synthase, cycloartenol synthase, and lupeol synthase, which represent the majority of the triterpenoid skeletons[13].

Previous studies have generated transcriptomic resources for *O. elatus* and reported global expression patterns or candidate genes associated with secondary metabolism. Eom et al.[14] performed the first de novo transcriptome assembly for this species and proposed genes potentially involved in triterpenoid saponin biosynthesis. More recently, Seo et al.[15] conducted a comparative transcriptome analysis of root tissues and regenerated plantlets, providing an assembled dataset that expanded the available genomic resources for this endangered plant.

Olapanax elatus Nakai (Araliaceae) is an endangered medicinal plant found in sparse populations in southeastern Russia, northeastern China, and northern Korea[16]. *O. elatus* is a deciduous shrub 1.0–1.8 m tall, with thick, thorny stems arising from the base of the rootstock and large, palmately lobed leaves[17]. Seed propagation of *O. elatus* is extremely limited. Most seeds develop poorly, and those that do germinate typically do so two years after falling to the ground, with most seedlings subsequently lost[18]. *O. elatus* is a far-eastern plant recommended for research as a source of herbal preparations similar to ginseng[19]. *O. elatus* has been used to treat nervous exhaustion, hypothermia, schizophrenia, cardiovascular disease, diabetes, and rheumatism and exhibits antifungal, antipyretic, analgesic, and anti-aging activities[20]. *O. elatus* contains various compounds with pharmacological effects, such as saponins, flavonoids, anthraquinones, and terpenes[21]. However, to date, little research has been conducted on the biosynthetic mechanisms of the major compounds that exhibit pharmacological activity in *O. elatus* or related transcriptomic studies.

The main goal of transcriptomics is to identify all types of transcripts and quantify changes in their expression levels[22]. RNA-seq and transcript comparison analyses were used to identify the DEGs involved in triterpenoid biosynthesis[23]. DEGs are genes whose expression levels differ significantly between two or more compared conditions[24]. In triterpenoid research, the identification of DEGs allows us to understand how gene expression changes in response to various treatments that affect triterpenoid biosynthesis process[25].

Previous transcriptome studies of *O. elatus* have reported global expression patterns and listed candidate genes potentially associated with secondary metabolite biosynthesis[14,15]. However, these studies did not investigate functional triterpenoid biosynthetic enzymes, nor did they examine the metabolic consequences of differential gene expression. Specifically, no study has linked triterpenoid accumulation with the expression of oxidosqualene cyclases (OSCs), nor performed motif characterization, phylogenetic validation, or molecular docking to confirm enzyme function. Moreover, the regeneration-induced increase in triterpenoid content and its underlying molecular mechanism have not been elucidated. Therefore, in this study, we re-analyzed previously generated transcriptome data of *O. elatus*[15] to identify key genes associated with triterpenoid biosynthesis and integrated this re-analysis with new experimental validation, including HPLC-based metabolite profiling, differential gene-expression confirmation by qRT-PCR, conserved motif analysis, phylogenetic reconstruction, and protein–ligand docking. By combining transcriptome re-analysis with multi-level functional characterization, we present the first functional evidence of OSC genes contributing to triterpenoid biosynthesis in regenerated *O. elatus* tissues.

Materials and methods

Preparation of plant materials and extracts

The initial mother plant used to establish the in vitro stock culture originated from a legally acquired nursery-grown individual maintained at Kangwon National University. As *O. elatus* is an endangered species, no wild individuals were collected, and all procedures followed national guidelines for the handling and non-destructive propagation of endangered plant materials. The *O. elatus* plants used in this study were obtained from in vitro cultures maintained at the Plant Biotechnology Laboratory, Kangwon National University, Chuncheon, South Korea. A voucher specimen (Voucher No.: KNU-OE-2023–01) has been prepared and stored in the Plant Biotechnology Laboratory, Kangwon National University. For in vitro culture, root explants were first maintained on 1/3 Murashige and Skoog (MS) solid medium supplemented with 1% (w/v) sucrose and solidified with 0.8% (w/v) agar. The pH of all media was adjusted to 5.8 prior to autoclaving at 121 °C for 20 min. Explants were incubated at 24 ± 1 °C under a 16/8 h light/dark photoperiod (40–50 μmol m^{-2} s^{-1}) until newly differentiated roots appeared. Newly formed roots were excised into 0.4–0.5 cm segments and transferred into 250 mL Erlenmeyer flasks containing 100 mL MS liquid medium supplemented with 3% (w/v) sucrose. Liquid cultures were maintained on a rotary shaker (HK-SK86-2, HANKUK S&I Co., Ltd., Hwaseong, Republic of Korea) at 120 rpm in the dark at 24 ± 1 °C for 8 weeks. All procedures were performed under aseptic conditions in a laminar-flow hood (Fig. 1). The plant material was dried at 60 °C, ground, and extracted using 100% methanol at room temperature for 24 h. The extract was concentrated under reduced pressure at 45 °C using a rotary vacuum concentrator (EYELA N-1110, Tokyo, Japan) and used in the experiment at a concentration of 10,000 μg/mL in 100% methanol. All experimental research and plant material handling in this study complied with institutional, national, and international guidelines and legislation. As the plant material used was derived from in vitro cultures maintained at Kangwon National University, no specific permissions or licenses for wild collection were required.

Detection of marker compounds using HPLC

As this study focused on the quantitative determination of three triterpenoid standards, the analysis represents a targeted metabolite profiling rather than a metabolome-wide investigation. Triterpenoid content was analyzed using high-performance liquid chromatography with diode-array detection (HPLC–DAD) (Agilent 1260 series) with a C18 column (4.6 × 250 mm, 5 μm; Agilent Technologies Inc., Santa Clara, CA, USA). For quantitative analysis, standard solutions of oleanolic acid, lupeol, and betulin were prepared at five concentrations (10, 25, 50, 100, and 200 μg/ml), and calibration curves were prepared. All compounds showed linearity with $R^2 \geq 0.99$.

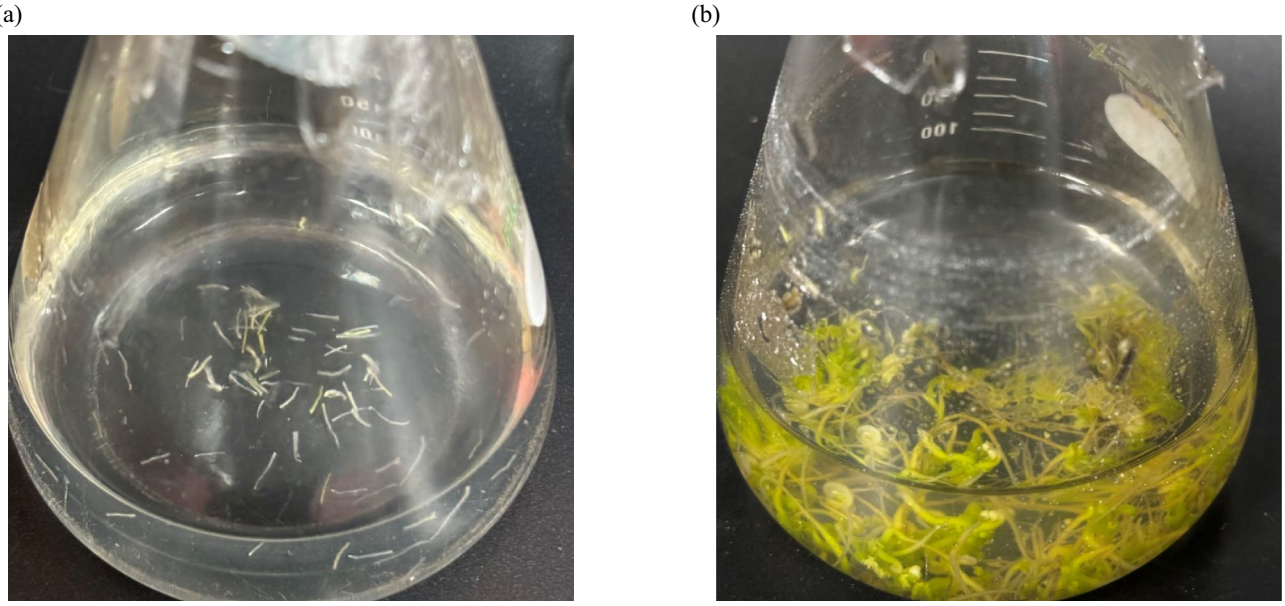

Fig. 1. Liquid culture of *O. elatus* (**a**) Root tissues at 0 weeks (**b**) regenerated plants from root tissues at 8 weeks.

The two samples were analyzed by repeating the experiment three times. HPLC analysis of lupeol was performed as previously described by Rabbani et al.[26]. The mobile phase was a mixture of acetonitrile and water at a 90:10 (v/v) ratio, and the flow rate was set to 0.9 mL/min. The HPLC analysis of oleanolic acid was performed as previously described by Tyszczuk-Rotko et al.[27]. The mobile phase was a mixture of acetonitrile, water, and 1% phosphoric acid in a ratio of 80:20:0.5 (v/v/v). The HPLC analysis of betulin was performed as described by Cho et al.[28]. The mobile phase consisted of water and acetonitrile, and the gradient conditions were as follows: at the start (0 min), A:B = 85:15; A was gradually changed to 0% and B to 100% over 20 min and maintained until 25 min, then returned to the initial conditions at 30 min and maintained until 40 min. All three compounds were detected at 210 nm.

Selection of candidate genes related to triterpenoid biosynthesis

Transcriptome data used in this study were obtained from the previously published RNA sequencing (RNA-seq) dataset of regenerated plantlets and root tissues of *O. elatus*[15]. No new RNA-seq libraries were generated. Instead, we performed a focused re-analysis of this dataset to identify candidate genes involved in triterpenoid biosynthesis. Heat maps were generated using TBtools[29,30]. Heatmaps were generated using TBtools (v1.098), applying hierarchical clustering with the Euclidean distance method and complete linkage. Gene expression values were normalized using $\log_2(\text{FPKM} + 1)$ transformation. The raw and assembled RNA-seq reads analyzed in this study were obtained from the publicly available transcriptome dataset of *O. elatus* deposited in the NCBI Sequence Read Archive under BioProject accession number PRJNA1136030.

RT-qPCR analysis

Total RNA was extracted from 0 W roots and 8 W seedlings cultured in liquid medium using TRIzol reagent (Invitrogen Scientific, Inc., USA), and the purity of the total RNA was confirmed using a microvolume spectrophotometer (Optizen microQ; Keen Innovative Solutions, Daejeon, Korea). cDNA was synthesized using PrimeScript™ RT Master Mix (Perfect Real Time) (Takara Korea Biomedical Inc., Seoul, Korea) in a 20-µL volume. A 25-µL mixture was prepared using TB Green® Premix Ex Taq™ II (Tli RNaseH Plus) (Takara Korea Biomedical Inc., Seoul, Republic of Korea), and reverse transcription quantitative PCR (RT-qPCR) analysis was performed using the CronoSTAR™ 96 Real-Time PCR System (Takara Korea Biomedical Inc., Seoul, Republic of Korea). Genes encoding enzymes related to triterpenoid biosynthesis were analyzed to measure transcript expression levels. Specific primers for qPCR analysis were designed using Primer3Plus (Table 1). *OeActin* was used as a reference gene. The qPCR analysis conditions included an initial denaturation step at 95 °C for 30 s, followed by 40 cycles of two-step amplification (denaturation at 95 °C for 5 s and annealing at 60 °C for 30 s). The dissociation stage was performed under the following conditions: 95 °C for 1 min, 60 °C for 15 s, and 98 °C for 5 s.

Phylogenetic relationship and sequence analysis of candidate enzyme proteins

A phylogenetic tree of putative genes encoding OSCs involved in triterpenoid biosynthesis was generated using MEGA X software with the neighbor-joining method and 1000 bootstraps[31]. The Poisson substitution model was applied, and gaps/missing data were treated with the pairwise-deletion option. To analyze the phylogenetic relationships of OSCs, the sequences of each protein were collected from the National Center for Biotechnology

Reference genes	Primer sequences (5'→3')
Gene_43800T	Forward: CTGGCACCCACCTATAAGCA Reverse: GAACTGGGAAAAACTGCAGGT
Gene_13767T	Forward: GCCCTCGATTCAGCTTTCCT Reverse: GCTACTCAACGACCCTGCTT
Gene_10233T	Forward: ACCATGAGAATGGGTGCAGGC Reverse: ATTTCCACCCCAGAGGCATG
Gene_11115T	Forward: GGGGGATGCCTTCAACATGA Reverse: CACTGGCTTACGAAGGGTGT
Gene_22342T	Forward: TCCCCACCCTCCATCTTCAT Reverse: TGCCGGACCTCTCTTTTTCC
Gene_05624T	Forward: CATACAAGCGCCATCTTGCC Reverse: TGCTTTCCTAAGAGCTGGCC
OeActin	Forward: AACGTTGTAGCCGACCTTCT Reverse: GTCTGCACGAACAAGATGGA

Table 1. Primer sequence for reference genes to qPCR analysis.

Information (NCBI) (https://www.ncbi.nlm.nih.gov/), and trEvolView (http://www.evolgenius.info/evolview/) was used for tree editing. Multiple Expectation–Maximization for Motif Elicitation (MEME) (http://meme-suite.org/tools/meme) was used to search for conserved motifs in the lupeol synthase and beta-amyrin synthase protein sequences. Motif discovery was performed using MEME Suite (v5.5.4) with the following parameters: maximum number of motifs = 10, motif width of 6–50 amino acids, and site distribution set to ZOOPS (zero or one occurrence per sequence). Additionally, the identified motifs were compared with protein sequences from other plants to predict conserved motifs in candidate enzymes. For protein motif analysis, the InterPro database (https://www.ebi.ac.uk/interpro/), provided by the European Bioinformatics Institute (EBI) was used. Protein domain annotations were performed using InterProScan (v5.62–94.0) with default parameters, including Pfam, SMART, SUPERFAMILY, and Gene Ontology databases. The e-value threshold was set to 1e-5. Protein sequences were analyzed using the InterProScan tool, which identifies protein motifs, including repetitive sequences and functional domains. The analysis results were further interpreted using sub-databases such as Gene Ontology (GO) and Pfam. Multiple amino acid sequence alignments were performed using DNAMAN[32].

Molecular docking

Protein 3D structure modeling was performed using SWISS-MODEL (https://swissmodel.expasy.org/). SWISS-MODEL was run with automatic template selection. Molecular docking in SWISS-Dock was performed using the default CHARMM-based scoring function, and the top-ranked clusters were selected for visualization. Molecular docking analysis was performed using SWISS-Dock software (https://www.swissdock.ch/). Default settings were used to predict the binding affinity between the ligand and target protein, and the docking results and protein–ligand complexes were visualized and analyzed using Discovery Studio. In addition, distance analysis between the ligand and key residues of the target protein was performed to evaluate specific interactions. This allowed us to confirm the binding modes, residues, and hydrogen bonds of the complexes.

Results

Comparison of triterpenoid marker substance content through HPLC analysis of *O. elatus*

The triterpenoid-type marker content in *O. elatus* was analyzed using HPLC–DAD (Fig. 2). First, the oleanolic acid content was 8.28 ± 0.09 μg/mL at 0 W and 15.85 ± 0.12 μg/mL at 8 W. The lupeol content was 70.11 ± 1.74 μg/mL at 0 W and 97.60 ± 0.37 μg/mL at 8 W. The betulin content was 9.22 ± 0.28 μg/mL at 0 W and 20.89 ± 0.25 μg/mL at 8 W (Table 2). Betulin content at 8 W was more than twice that at 0 W (Table 2). The concentrations of all marker compounds were higher in the re-differentiated *O. elatus* after 8 W than in the root tissues.

Identification of genes involved in triterpenoid biosynthesis

In this study, we used transcriptomic analyses to identify several candidate genes that may be involved in triterpenoid biosynthesis. The prediction and expression levels of enzyme-encoding genes were analyzed based on $\log_2$(FPKM) values and visualized as a heatmap (Fig. 3). Additionally, the identified unigenes were found to be involved in triterpenoid backbone biosynthesis. The major enzymes analyzed in this study are summarized as follows. The enzymes involved in the MVA pathway include acetyl-CoA C-acetyltransferase (AACT), hydroxymethylglutaryl-CoA synthase (HMGS), hydroxymethylglutaryl-CoA reductase (HMGR), mevalonate kinase (MVK), phosphomevalonate kinase (PMK), and diphosphomevalonate decarboxylase (MVD) Mevalonate pyrophosphate decarboxylase. The enzymes involved in the MEP pathway are as follows: 1-deoxy-D-xylulose-5-phosphate synthase (DXS), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (MCT), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MCS), (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (HDS), 4-hydroxy 3-methylbut-2-enyl diphosphate reductase (HDR), Isopentenyl-diphosphate delta-isomerase (IDI). In addition, isoprenoid precursors synthesized via the MVA and MEP pathways are converted to oxidosqualene, a key intermediate in triterpenoid biosynthesis, through the squalene biosynthesis pathway. This pathway includes geranyl diphosphate synthase (GPS), farnesyl diphosphate synthase (FPS), squalene synthase (SQS), squalene epoxidase (SE), and other enzymes that play important

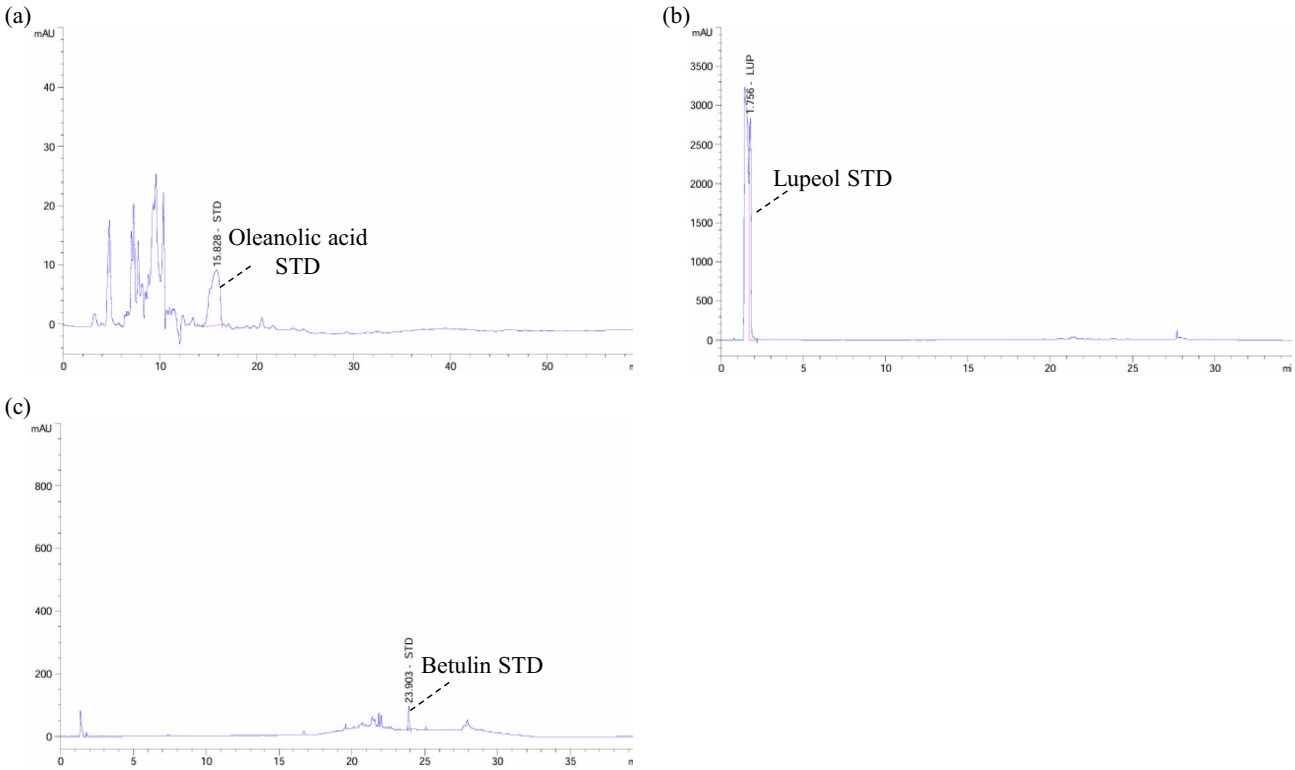

Fig. 2. HPLC analysis of triterpenoid content in regenerated O. elatus at 0 and 8 weeks. (**a**) oleanolic acid standard, (**b**) lupeol standard, (**c**) betulin standard, (**d**,**e**) oleanolic acid at 0 and 8 weeks, (**f**,**g**) lupeol at 0 and 8 weeks and (**h**,**i**) betulin at 0 and 8 weeks.

roles in triterpenoid skeleton formation. Transcriptome analysis revealed two unigenes encoding GPS and one unigene each encoding FPS, SQS, and SE. In addition, four unigenes for beta-amyrin synthase and three unigenes for lupeol synthase were identified. In this study, we identified several candidate genes potentially involved in triterpenoid biosynthesis through transcriptomic analysis and performed qRT-PCR to verify the expression patterns of genes with high expression levels (Fig. 4). All genes showed high expression levels at 8 W, with Gene_22342T exhibiting more than three-fold higher expression at 8 W than at 0 W. Furthermore, Gene_05624T showed a more than 30-fold higher expression at 8 W than at 0 W.

Phylogenetic and motif analysis of oxidosqualene cyclases

To determine the phylogenetic relationships between lupeol synthase, beta-amyrin synthase, and other OSC enzymes, we constructed a phylogenetic tree based on protein sequences. Therefore, we performed an analysis that included various OSC sequences reported from other plant species. Consequently, Gene_22342T and Gene_05624T were classified into the beta-amyrin synthase and lupeol synthase groups, respectively. Furthermore, these genes showed strong homology with OSC from specific plant species, suggesting that the genes identified in this study likely play an important role in the triterpenoid biosynthetic pathway (Fig. 5). To confirm whether this phylogenetic similarity was reflected in protein sequences, we performed conserved motif comparisons and amino acid alignment analyses. Gene_22342T shared motif patterns similar to those of beta-amyrin synthase sequences from other plants (Fig. 6a). InterPro analysis was performed to examine whether the Gene_22342T motif sequence matched known protein domains. This analysis confirmed that Gene_22342T belongs to the squalene cyclase family and performs triterpenoid biosynthetic process functions (Table 3). Additionally, protein alignment was performed between Gene_22342T and beta-amyrin synthase sequences from other plants to analyze the conservation of the major functional motifs (Fig. 6b). The analysis revealed that gene_22342T contained two QW motifs and one MWCYCR motif, which are characteristic of beta-amyrin synthase. Gene_05624T, which was presumed to be a lupeol synthase, was also analyzed using the same method, and the results showed that it shared 10 motif patterns with high similarity to lupeol synthase sequences from other plants (Fig. 7a). Furthermore, InterPro analysis confirmed that the motif sequences of Gene_05624T matched previously reported protein domains, indicating that this gene belongs to the squalene cyclase family and is involved in triterpenoid biosynthesis (Table 4). Furthermore, to evaluate the functional conservation of Gene_05624T, protein alignment was performed using lupeol synthase sequences from other plants, and the conservation of major functional motifs was analyzed (Fig. 7b). The analysis revealed that Gene_05624T contained four QW motifs, one MLCYCR motif, and one DCTAE motif, which are characteristic of lupeol synthase. In conclusion, considering the phylogenetic analysis, motif comparison, and InterPro functional

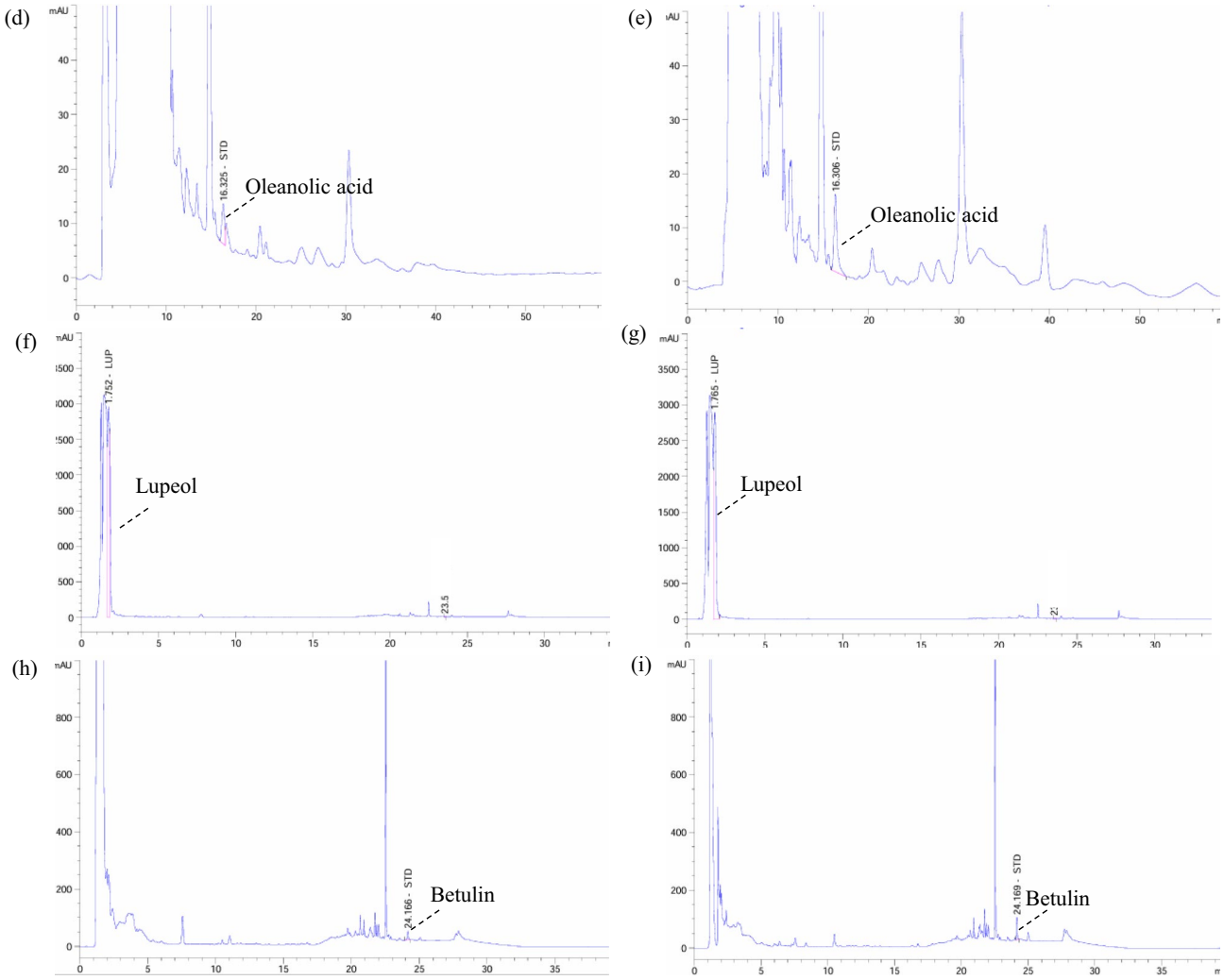

Fig. 2. (continued)

Compound	Name	Amount (µg/ml)
Oleanolic acid	0W	8.28 ± 0.09
	8W	15.85 ± 0.12
Lupeol	0W	70.11 ± 1.74
	8W	97.60 ± 0.37
Betulin	0W	9.22 ± 0.28
	8W	20.89 ± 0.25

Table 2. Triterpenoid contents in the root tissues of *O. elatus* at 0 week and in plants regenerated from root tissues after 8 weeks of cultivation.

analysis, Gene_05624T is likely to function as a lupeol synthase and Gene_22342T as a beta-amyrin synthase. Both genes encode OSC enzymes that play important roles in the triterpenoid biosynthetic pathway.

Molecular interaction analysis between OSCs and their substrate

First, analysis of the molecular interaction between beta-amyrin and Gene_22342T in 3D structure revealed that beta-amyrin binds to the active site of the enzyme (Fig. 8a). When visualized in 2D, beta-amyrin was found to form van der Waals interactions with ARG88 (arginine), GLU53, GLU85 (glutamic acid), ILE86 (isoleucine), PHE60 (phenylalanine) of Gene_Oe22342T, a π-σ stacking interaction with TYR89 (tyrosine), and alkyl interactions with ALA55 (alanine) and LEU61 (Leucine) (Fig. 8b). The calculated affinity for this binding was − 8.682 kcal/mol, indicating a relatively stable interaction (Table 5). Analysis of the molecular interactions between lupeol and Gene_05624T in the 3D structure revealed that lupeol also binds to the active site of the enzyme (Fig. 8c).

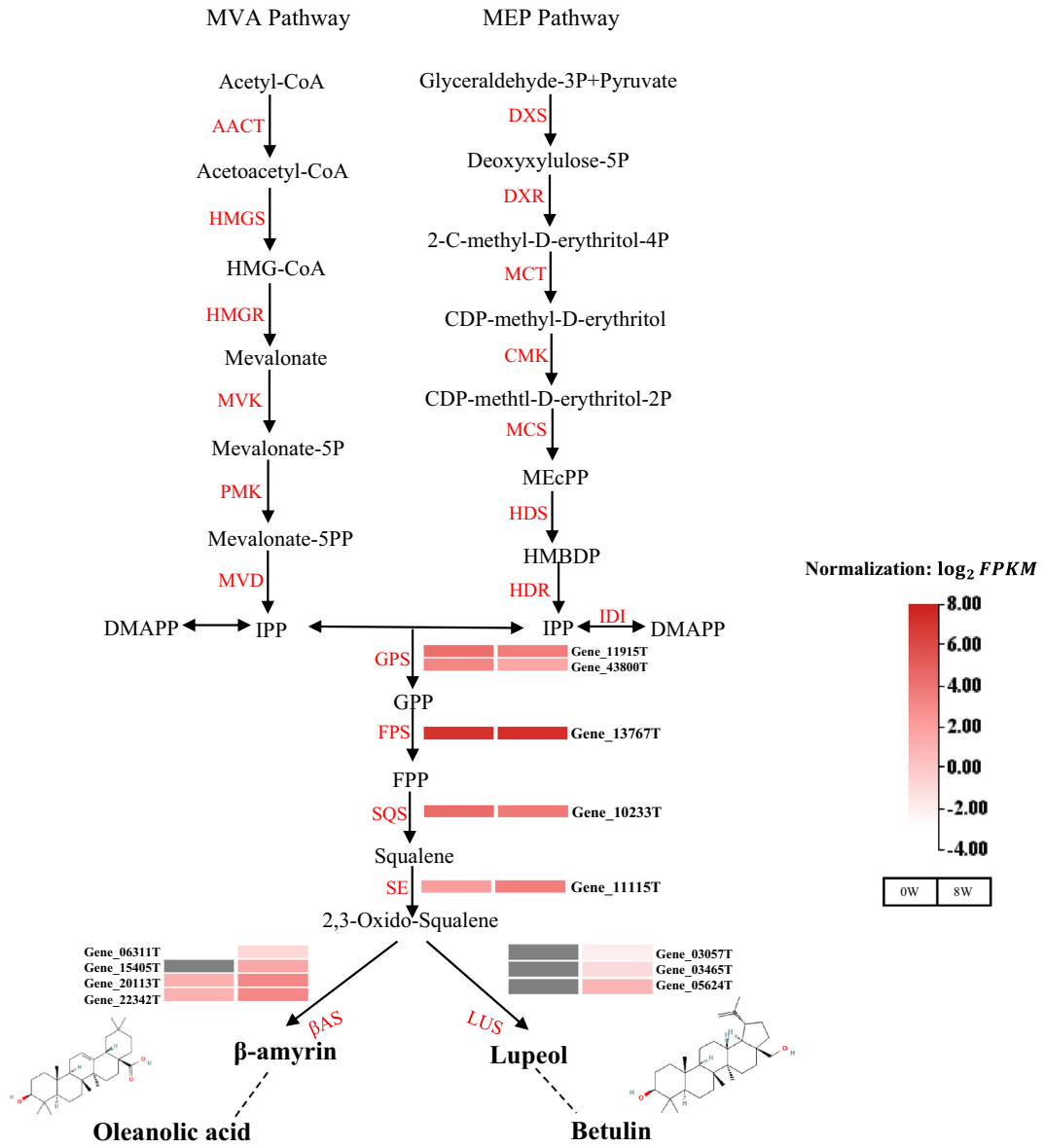

Fig. 3. The heatmap compares the expression profiles of candidate genes associated with triterpenoid biosynthesis in the roots at 0 week and the regenerated plants at 8 weeks of *O. elatus*, normalized by $\log_2$ FPKM values.

When visualized in 2D, lupeol forms van der Waals interactions with ALA486 (alanine), ASP488 (aspartic acid), CYS373, CYS489 (cysteine), LEU416 (leucine), PHE129, PHE416 (phenylalanine), SER415 (serine), TRP536 (tryptophan), TYR738 (tyrosine), VAL486 (valine), and van der Waals interactions, TRP421 (tryptophan) and π-σ stacking interactions, and PHE477, PHE730 (phenylalanine), TRP614 (tryptophan), TYR263 (tyrosine) through alkyl interactions (Fig. 8d). Their calculated affinity was − 9.249 kcal/mol, indicating slightly higher stability than the binding between Gene_22342T and beta-amyrin (Table 6). These results suggest that Gene_22342T and Gene_Oe05624T are likely function as beta-amyrin synthase and lupeol synthase, respectively, and are interpreted to play important roles in the selective recognition and conversion of substrates by these two enzymes. These results represent strong functional indications of the catalytic roles of Gene_22342T and Gene_05624T, although direct biochemical validation remains to be performed.

Discussion

In previous studies of *O. elatus*, transcriptome analysis has revealed the expression of numerous genes associated with the synthesis and biosynthesis of various secondary metabolites[15]). It has been reported that *O. elatus* stems contain substances such as uracil, adenosine, protocatechuic acid, syringin, and scoparone[33]. Although some analyses of secondary metabolites in *O. elatus* have been reported, information on the biosynthetic pathways and gene-level details of triterpenoid compounds remains largely unknown. In the present study, HPLC analysis detected lupeol, betulin, and oleanolic acid in *O. elatus* 0 W and 8 W samples, with a higher triterpenoid content

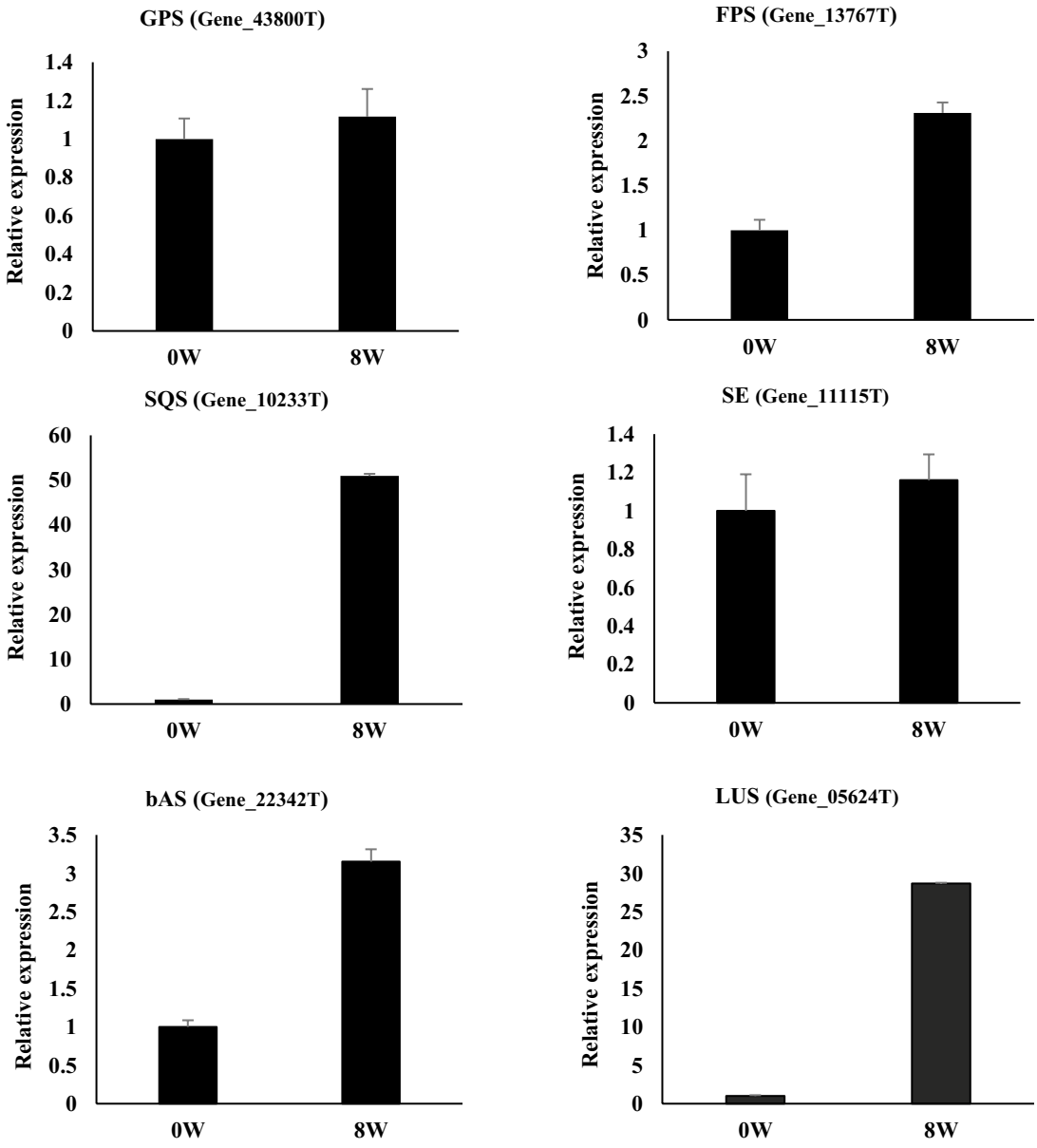

Fig. 4. Comparison of gene expression level related to triterpenoid biosynthesis using RT-qPCR. Enzyme names are abbreviated as follows: *GPS* geranyl pyrophosphate synthase, *FPS* farnesyl diphosphate synthase, *SQS* squalene synthase, *SE* squalene monooxygenase, *BAS* beta-amyrin synthase, *LUS* lupeol synthase.

at 8 W than at 0 W (Fig. 2, Table 2). Based on previous transcriptomic analysis of *O. elatus* by tissue type, the expression of genes related to triterpenoid biosynthesis was reported to be most active in leaves[14]. These results serve as foundational data to explain the phenomenon observed in this study, in which the triterpenoid content was higher in the shoot tissue than in the root tissue, suggesting that increased gene expression in the shoot tissue is closely associated with the accumulation of biosynthetic compounds. This study did not aim to perform metabolomics-level profiling; instead, we focused on three representative triterpenoid metabolites to investigate their correlation with transcript expression.

Although previous transcriptome datasets of *O. elatus* are available, they primarily describe gene catalogs or broad expression landscapes without functional validation of triterpenoid-pathway genes. Notably, no prior study has characterized β-amyrin synthase or lupeol synthase at the sequence-motif level, nor demonstrated substrate–enzyme interactions through molecular docking. Our study is the first to connect increased triterpenoid accumulation in regenerated tissues with the upregulation and functional confirmation of specific OSC enzymes, thereby elucidating a mechanism that was not addressed in earlier transcriptomic analyses. This study did not generate new transcriptome libraries; rather, it applied a functional, pathway-oriented re-analysis to an existing dataset, followed by experimental validation of key biosynthetic genes.

In this study, we identified genes involved in the biosynthesis of beta-amyrin and lupeol, which are triterpenoid compounds found in the endangered plant *O. elatus*, using comparative transcriptomic analysis.

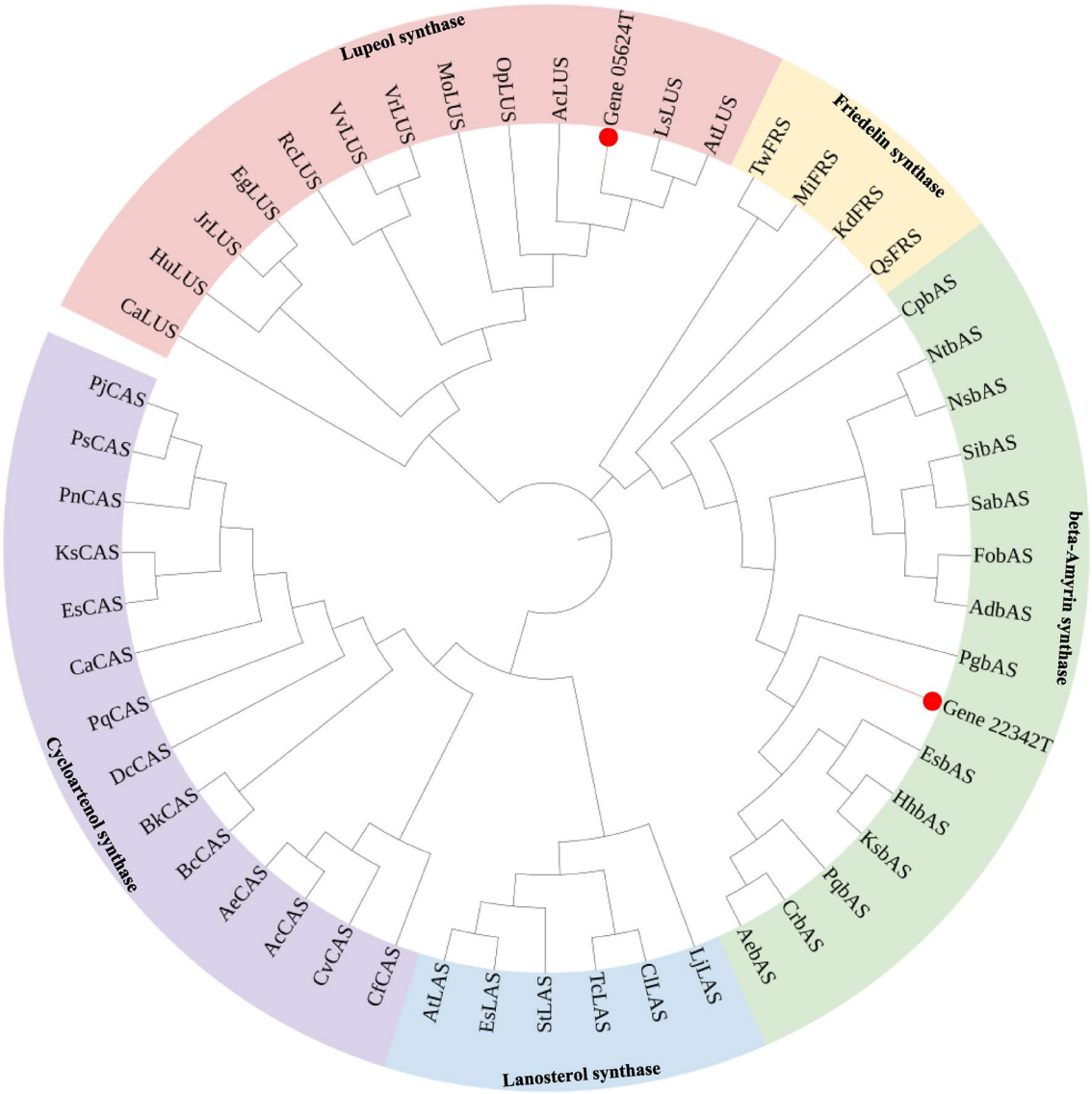

Fig. 5. Phylogenetic tree constructed from amino acid sequences of *O. elatus* OSCs and characterized OSCs from other plants. Red circles mark Gene_22342T and Gene_05624T, respectively.

The biosynthesis of triterpenoids is a crucial pathway for producing natural compounds with various medicinal effects, such as cycloartenol, lupeol, beta-amyrin, and lanosterol[34]. In the present study, transcriptomic analysis revealed that genes associated with the synthesis of triterpenoid precursors (GPS, FPS, SQS, and SE) were highly expressed in *O. elatus* (Fig. 3). These genes accumulate 2,3-oxidosqualene, and after accumulation, the OSCs mentioned earlier determine the secondary metabolites of other groups[35]. OSCs catalyze the cyclization of 2,3-oxidosqualene into sterols and triterpenes with various skeletons[36]. OSCs are widely distributed across various plant lineages, with beta-amyrin synthase and lupeol synthase accounting for a significant proportion of the OSC family[29,30]. In this study, comparative transcriptomic analysis of root tissues and re-differentiated somatic plants revealed that OSC family genes were highly expressed in somatic plants, with Gene_22342T showing a threefold difference and Gene_ 05624 T showing a 30-fold difference in expression levels (Fig. 4). Some of these genes were phylogenetically similar to the previously reported genes, Gene_22342T and Gene_ 05624 T (Fig. 5).

OSCs contain three conserved motifs, OSC-specific DCTAE, MXCYCR, and QW repeats[37]. The DCTAE motif is associated with substrate binding[38]. Additionally, the QW repeat sequence is an aromatic region with a negative charge that may stabilize the carbocations generated during reduction reactions and contribute to maintaining the structural stability of the protein[39]. MXCYCR is a conserved motif important for triterpenoid production that varies depending on OSCs. Beta-amyrin synthase primarily has the MWCYCR sequence, whereas lupeol synthase primarily has the MLCYCR form[40]. In this study, Gene_22342T was found to have the MWCYCR and QW motifs, whereas Gene_ 05624Twas had three motifs: DCTAE, MXCYCR, and QW (Figs. 6b, 7b). However, the sequence of Gene_22342T obtained in this study was partial rather than full length, which

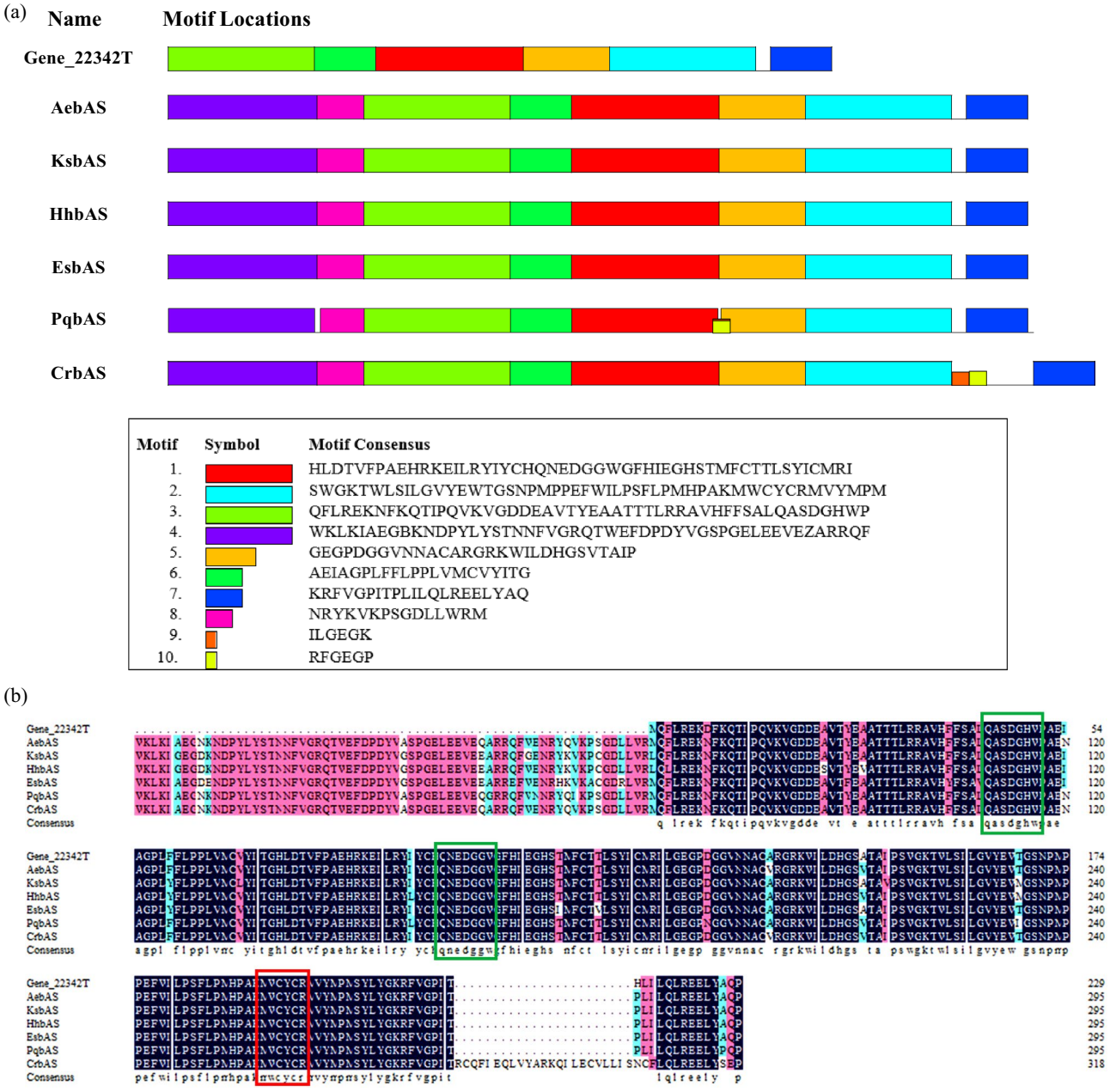

Fig. 6. The conserved motifs of beta-amyrin synthase (**a**) The distribution patterns of ten motifs in beta-amyrin synthase from various plants are shown. Each motif is represented with specific colors at the bottom. (**b**) Alignment of the deduced amino acid sequence of beta-amyrin synthase from *O. elatus* with those of other species. Red boxes highlight the **MWCYCR** motif, while green boxes indicate the **QW** motifs.

may limit our understanding of the overall functional characteristics of the protein. Nevertheless, it showed high sequence similarity to beta-amyrin synthases from other plants and included major conserved motifs, suggesting a high likelihood of functioning as a beta-amyrin synthase. The full sequence of Gene_05624T was obtained, and a sequence comparison with lupeol synthases from other plant species revealed shared conserved motifs, leading to the conclusion that it is a functional gene involved in lupeol biosynthesis. These findings expand upon previous transcriptome analyses of *O. elatus* by providing functional validation—through motif comparison, phylogenetic relationships, and sequence conservation—that was not addressed in earlier genomic studies.

In addition, with this study we further verified the possibility that Gene_22342T and Gene_05624T function as beta-amyrin synthase and lupeol synthase, respectively, through protein–ligand molecular docking and interaction analysis (Fig. 8). Protein–ligand docking is a molecular modeling method that predicts the binding mode between a ligand and a protein, with the primary goal of achieving optimal binding between the ligand and protein[41]. Docking can also be used to understand how a ligand affects protein function and identify potential protein–ligand interactions[42]. Docking algorithms evaluate the quality of the predicted binding poses

Motif	Sequence	Width	Protein family	InterPro GO terms
1	HLDTVFPAEHRKEILRYIYCHQNEDGGWGFHIEGHSTMFCTTLSYICMRI	50	Squalene cyclase	Biological Process: triterpenoid biosynthetic process (GO:0016104) Molecular Function: intramolecular transferase activity (GO:0016866) Celluar Component: lipid droplet (GO:0005811)
2	SWGKTWLSILGVYEWTGSNPMPPEFWILPSFLPMHPAKMWCYCRMVYMPM	50	Squalene cyclase	Biological Process: triterpenoid biosynthetic process (GO:0016104) Molecular Function: intramolecular transferase activity (GO:0016866) Celluar Component: lipid droplet (GO:0005811)
3	QFLREKNFKQTIPQVKVGDDEAVTYEAATTTLRRAVHFFSALQASDGHWP	50	Squalene cyclase	Biological Process: triterpenoid biosynthetic process (GO:0016104) Molecular Function: intramolecular transferase activity (GO:0016866) Celluar Component: lipid droplet (GO:0005811)
4	WKLKIAEGBKNDPYLYSTNNFVGRQTWEFDPDYVGSPGELEEVEZARRQF	50	Non predicted	
5	GEGPDGGVNNACARGRKWILDHGSVTAIP	29	Non predicted	
6	AEIAGPLFFLPPLVMCVYITG	21	Non predicted	
7	KRFVGPITPLILQLREELYAQ	21	Non predicted	
8	NRYKVKPSGDLLWRM	15	Non predicted	
9	ILGEGK	6	Non predicted	
10	RFGEGP	6	Non predicted	

Table 3. Conserved motifs and functional annotations of beta-amyrin synthase.

using a scoring function[43]. In molecular docking, "kcal/mol" refers to the unit used to express the predicted binding energy or affinity between the ligand and the protein[44]. A negative docking score indicates that the binding process is energetically favorable; the more negative the score, the stronger the binding affinity[45]. In the molecular docking conducted in this study, the binding affinity between Gene_22342T and beta amyrin was − 8.682 kcal/mol, and the binding affinity between Gene_05624T and lupeol was − 9.249 kcal/mol. This indicates that the binding between the substance and the enzyme is relatively stable (Tables 5 and 6).

In plants, oleanolic acid is produced through oxidation and diversification processes at the c-28 position of beta amyrin via cytochrome P450 monooxygenase activity[46]. CYP716A enzymes, such as CYP716A, CYP716C67, and CYP716A48 play important roles in this process[47,48]. In the present study, oleanolic acid was detected in *O. elatus*, suggesting that oxidative metabolic pathways may be active. However, the transcriptomic comparison analysis did not identify candidate genes from the CYP716 family that mediate such reactions. Further in-depth analysis of the CYP gene family is required to fully understand the biosynthetic pathways of triterpenoid compounds. Although further analysis of subsequent oxidative enzymes, particularly CYP family genes, is required, the results of the expression and functional analyses of OSCs family genes in *O. elatus* obtained in this study can contribute to a deeper understanding of the triterpenoid biosynthetic pathway in this species and expand our understanding of the biosynthetic network as a whole. Overall, our work complements prior transcriptome-based gene surveys in *O. elatus*[14,15] and advances the field by linking transcriptomic signatures with experimentally validated triterpenoid biosynthetic functions.

Although direct functional assays such as heterologous expression, enzyme activity measurements, or transient expression assays were beyond the scope of the present study, the combined evidence from (i) expression profiling, (ii) conserved motif analysis, (iii) phylogenetic clustering, and (iv) molecular docking strongly supports the functional identity of Gene_22342T and Gene_05624T as β-amyrin synthase and lupeol synthase, respectively. Future functional characterization using heterologous systems (e.g., *Nicotiana benthamiana* or yeast) will be essential to experimentally confirm their catalytic activity and to elucidate their direct contribution to triterpenoid accumulation in *O. elatus*.

Conclusion

In this study, we quantitatively analyzed the contents of lupeol, oleanolic acid, and betulin, which are the major triterpenoid compounds in the endangered plant *O. elatus*, and found that the 8-week-old regenerated plant (8 W) sample had a higher content than the 0-week-old root tissue (0 W) sample. Through transcriptomic analysis, genes involved in the biosynthesis of beta-amyrin and lupeol, the major triterpenoid compounds, were identified, and their functional potential was further validated. Gene_22342T and Gene_05624T showed the potential to function as beta-amyrin synthase and lupeol synthase, respectively. Their expression patterns, conserved motifs, and binding affinity results strongly suggested a role for these enzymes. Although the functional analysis of the subsequent oxidases and conversion pathways remains challenging, this study represents the first comprehensive gene-level analysis of OSCs in the endangered plant *O. elatus*, suggesting the possibility of previously unknown metabolic pathways. These findings provide foundational data for future metabolic engineering applications. While experimental functional validation was not performed, the multi-level integrative analyses presented here provide a strong foundation for identifying key OSC genes in *O. elatus*. Follow-up studies incorporating functional assays will help clarify the mechanistic relationship between these OSCs and the increased levels of lupeol, oleanolic acid, and betulin observed in regenerated tissues.

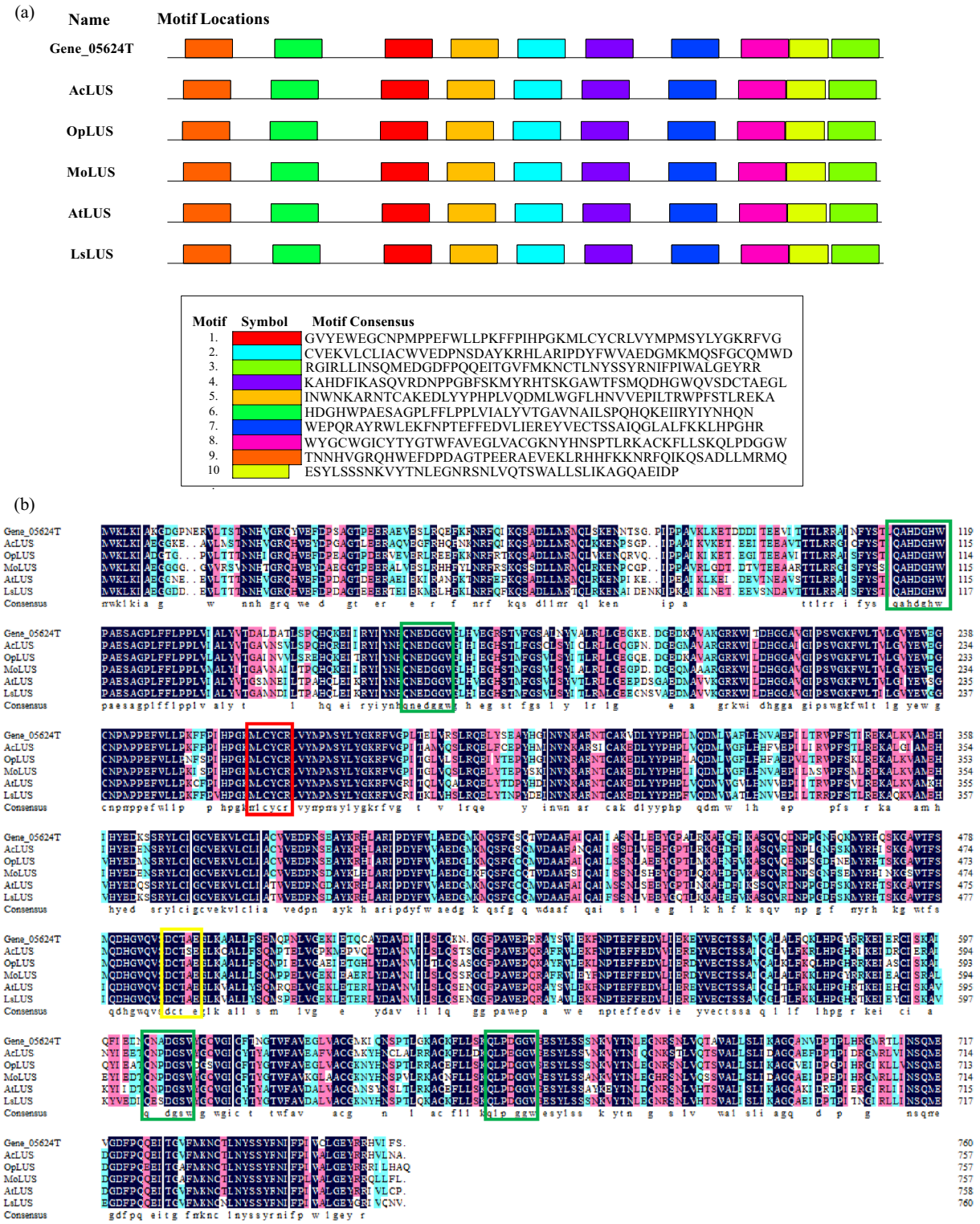

Fig. 7. The conserved motifs of lupeol synthase (**a**) The distribution patterns of ten motifs in lupeol synthase from various plants are shown. Each motif is represented with specific colors at the bottom. (**b**) Alignment of the deduced amino acid sequence of lupeol synthase from *O. elatus* with those of other species. Red box highlight the MLCYCR motif, green boxes indicate the QW motifs and yellow box represents the DCTAE motif.

Motif	Sequence	Width	Protein family	InterPro GO terms
1	GVYEWEGCNPMPPEFWLLPKFFPIHPGKMLCYCRLVYMPMSYLYGKRFVG	50	Squalene cyclase	Biological Process: triterpenoid biosynthetic process (GO:0016104) Molecular Function: intramolecular transferase activity (GO:0016866) Celluar Component: lipid droplet (GO:0005811)
2	CVEKVLCLIACWVEDPNSDAYKRHLARIPDYFWVAEDGMKMQSFGCQMWD	50	Squalene cyclase	Biological Process: triterpenoid biosynthetic process (GO:0016104) Molecular Function: intramolecular transferase activity (GO:0016866) Celluar Component: lipid droplet (GO:0005811)
3	RGIRLLINSQMEDGDFPQQEITGVFMKNCTLNYSSYRNIFPIWALGEYRR	50	Squalene cyclase	Biological Process: triterpenoid biosynthetic process (GO:0016104) Molecular Function: intramolecular transferase activity (GO:0016866) Celluar Component: lipid droplet (GO:0005811)
4	KAHDFIKASQVRDNPPGBFSKMYRHTSKGAWTFSMQDHGWQVSDCTAEGL	50	Squalene cyclase	Biological Process: triterpenoid biosynthetic process (GO:0016104) Molecular Function: intramolecular transferase activity (GO:0016866) Celluar Component: lipid droplet (GO:0005811)
5	INWNKARNTCAKEDLYYPHPLVQDMLWGFLHNVVEPILTRWPFSTLREKA	50	Squalene cyclase	Biological Process: triterpenoid biosynthetic process (GO:0016104) Molecular Function: intramolecular transferase activity (GO:0016866) Celluar Component: lipid droplet (GO:0005811)
6	HDGHWPAESAGPLFFLPPLVIALYVTGAVNAILSPQHQKEIIRYIYNHQN	50	Squalene cyclase	Biological Process: triterpenoid biosynthetic process (GO:0016104) Molecular Function: intramolecular transferase activity (GO:0016,866) Celluar Component: lipid droplet (GO:0005811)
7	WEPQRAYRWLEKFNPTEFFEDVLIEREYVECTSSAIQGLALFKKLHPGHR	50	Squalene cyclase	Biological Process: triterpenoid biosynthetic process (GO:0016104) Molecular Function: intramolecular transferase activity (GO:0016866) Celluar Component: lipid droplet (GO:0005811)
8	WYGCWGICYTYGTWFAVEGLVACGKNYHNSPTLRKACKFLLSKQLPDGGW	50	Squalene cyclase	Biological Process: triterpenoid biosynthetic process (GO:0016104) Molecular Function: intramolecular transferase activity (GO:0016866) Celluar Component: lipid droplet (GO:0005811)
9	TNNHVGRQHWEFDPDAGTPEERAEVEKLRHHFKKNRFQIKQSADLLMRMQ	50	Squalene cyclase	Biological Process: triterpenoid biosynthetic process (GO:0016104) Molecular Function: intramolecular transferase activity (GO:0016866) Celluar Component: lipid droplet (GO:0005811)
10	ESYLSSSNKVYTNLEGNRSNLVQTSWALLSLIKAGQAEIDP	41	Squalene cyclase	Biological Process: triterpenoid biosynthetic process (GO:0016104) Molecular Function: intramolecular transferase activity (GO:0016866) Celluar Component: lipid droplet (GO:0005811)

Table 4. Conserved motifs and functional annotations of lupeol synthase.

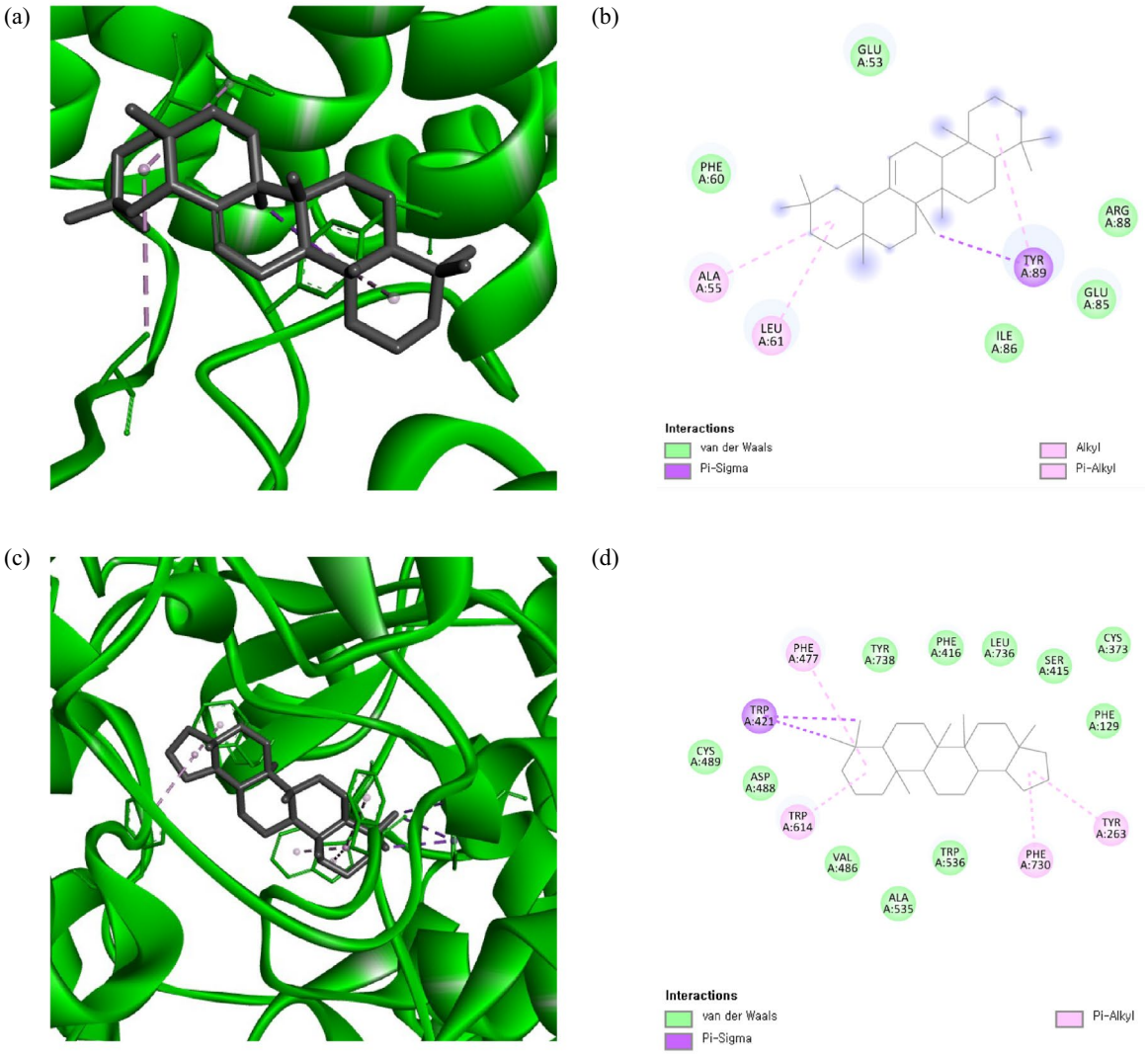

Fig. 8. Molecular docking of beta-amyrin and lupeol with their respective synthases. (**a**) 3D representation of the interaction between beta-amyrin and Gene_22342T. (**b**) 2D animated representation of beta-amyrin and Gene_22342T synthase, showing their interactions. (**c**) 3D representation of the interaction between lupeol and Gene_05624T. (**d**) 2D animated representation of lupeol and Gene_05624T, showing their interactions.

Model	Calculated affinity (kcal/mol)
1	−8.682
2	−8.331
3	−8.159
4	−7.712
5	−7.7549

Table 5. Calculated binding affinities between beta amyrin and Gene_22342T(bAS)models.

Model	Calculated affinity (kcal/mol)
1	−9.249
2	−8.629
3	−8.078
4	−7.654
5	−7.474

Table 6. Calculated binding affinities between lupeol and Gene_05624T(LUS)models.

Data availability

The RNA-seq data used in this study are available in the NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA1136030. No new RNA-seq data were generated in this study.

Received: 10 September 2025; Accepted: 13 March 2026
Published online: 17 March 2026

References

1. Ma, A. et al. Functional diversity of oxidosqualene cyclases in genus *Oryza*. *New Phytol.* **244**, 2430–2441 (2024).
2. Du, Z. et al. Genome-wide investigation of oxidosqualene cyclase genes deciphers the genetic basis of triterpene biosynthesis in tea plants. *J. Agric. Food Chem.* **72**, 10584–10595 (2024).
3. Tamfu, A. N. et al. Synthesis of benzoyl esters of β-amyrin and lupeol and evaluation of their antibiofilm and antidiabetic activities. *Results. Chem.* **4**, 100322 (2022).
4. Liu, H., Fan, J., Wang, C., Li, C. & Zhou, X. Enhanced beta-amyrin synthesis in *Saccharomyces cerevisiae* by coupling an optimal acetyl-CoA supply pathway. *J. Agric. Food Chem.* **67**, 3723–3732 (2019).
5. Du, M. M. et al. Engineering *Saccharomyces cerevisiae* for hyperproduction of beta-amyrin by mitigating the inhibition effect of squalene on beta-amyrin synthase. *J. Agric. Food Chem.* **70**, 229–237 (2022).
6. Viet, T. D., Xuan, T. D. & Anh, H. Alpha-amyrin and beta-amyrin isolated from *Celastrus hindsii* leaves and their antioxidant, anti-xanthine oxidase, and anti-tyrosinase potentials. *Molecules* **26**, 7248 (2021).
7. Li, D. et al. Natural lupeol level variation among castor accessions and the upregulation of lupeol synthesis in response to light. *Indust Crops Pro* **192**, 116090 (2023).
8. Sohag, A. A. M. et al. Molecular pharmacology and therapeutic advances of the pentacyclic triterpene lupeol. *Phytomedicine* **99**, 154012 (2022).
9. Liu, Y. et al. Identification and functional characterization of squalene epoxidases and oxidosqualene cyclases from *Tripterygium wilfordii*. *Plant Cell Rep.* **39**, 409–418 (2020).
10. Zhou, M. et al. Identification and analysis of UGT genes associated with triterpenoid saponin in soapberry (*Sapindus mukorossi* Gaertn.). *BMC Plant Biol.* **24**, 588 (2024).
11. Lertphadungkit, P. et al. De novo transcriptome analysis and identification of candidate genes associated with triterpenoid biosynthesis in *Trichosanthes cucumerina* L. *Plant Cell Rep.* **40**, 1845–1858 (2021).
12. Busta, L. et al. Oxidosqualene cyclases involved in the biosynthesis of triterpenoids in *Quercus suber* cork. *Sci. Rep.* **10**, 8011 (2020).
13. Zhang, M. et al. Genome-wide identification and functional characterization of 2, 3-oxidosqualene cyclase genes in *Phellodendron amurense*. *Ind. Crops Prod.* **208**, 117721 (2024).
14. Eom, S. H., Lee, J. K., Kim, H. & Hyun, T. K. De novo transcriptomic analysis to reveal functional genes involved in triterpenoid saponin biosynthesis in *Oplopanax elatus* NAKAI. *J. Appl. Bot. Food Qual.* **90**, 25–31 (2017).
15. Seo, J. W. et al. Comparative analysis of the transcriptomes from regenerated plants and root explants of endangered *Oplopanax elatus*. *Genes. Genomics.* **46**, 1387–1398 (2024).
16. Park, S., Moon, H., Murthy, H. & Kim, Y. Improved growth and acclimatization of somatic embryo-derived *Oplopanax elatus* plantlets by ventilated photoautotrophic culture. *Biol. Plant.* **55**, 559–562 (2011).
17. Moon, H. K., Kim, J. A., Park, S. Y., Kim, Y. W. & Kang, H. D. Somatic embryogenesis and plantlet formation from a rare and endangered tree species, *Oplopanax elatus*. *J. Plant Biol.* **49**, 320–325 (2006).
18. Reunova, G. D., Kats, I. L. & Zhuravlev, Y. N. Genetic variation of *Oplopanax elatus* (Araliaceae) populations estimated using DNA molecular markers. *Dokl. Biol. Sci.* **433**, 252–256 (2010).
19. Moon, Y. K., Kim, Y. W., Hong, Y. P. & Park, S. Y. Improvement of somatic embryogenesis and plantlet conversion in *Oplopanax elatus*, an endangered medicinal woody plant. *Springerplus* **2**, 1–8 (2013).
20. Jiang, X. L. et al. Hepatoprotective effect of *Oplopanax elatus* Nakai adventitious roots extract by regulating CYP450 and PPAR signaling pathway. *Front. Pharmacol.* **13**, 761618 (2022).
21. Shikov, A. N., Pozharitskaya, O. N., Makarov, V. G., Yang, W. Z. & Guo, D. A. *Oplopanax elatus* (Nakai) Nakai: Chemistry, traditional use and pharmacology. *Chin. J. Nat. Med.* **12**, 721–729 (2014).
22. Zu, Y. et al. Transcriptome analysis of main roots of *Panax ginseng* identifies genes involved in saponin biosynthesis under arsenic stress. *Plant Gene* **16**, 1–7 (2018).
23. Huo, Y. et al. Transcriptional responses for biosynthesis of triterpenoids in exogenous inducers treated *Inonotus hispidus* using RNA-seq. *Molecules* **27**, 8541 (2022).
24. Zhao, Y. et al. Transcriptomics-metabolomics joint analysis: New highlight into the triterpenoid saponin biosynthesis in quinoa (*Chenopodium quinoa* Willd.). *Front. Plant Sci.* **13**, 964558 (2022).
25. Zhang, D. X. et al. Paclobutrazol induces triterpenoid biosynthesis via downregulation of the negative transcriptional regulator SlMYB in *Sanghuangporus lonicericola*. *Commun. Biol.* **8**, 551 (2025).
26. Rabbani, V., Garoosi, G. A., Haddad, R., Farjaminezhad, R. & Japelaghi, R. H. Improvement and prediction of the extraction parameters of lupeol and stigmasterol metabolites of *Melia azedarach* with response surface methodology. *BMC. Biotechnol.* **24**, 39 (2024).
27. Tyszczuk-Rotko, K., Domańska, K., Sadok, I., Wójciak-Kosior, M. & Sowa, I. Voltammetric procedure for the determination of oleanolic and ursolic acids in plant extracts. *Anal. Methods* **7**, 9435–9441 (2015).
28. Cho, N. K., Kim, D. H. & Sung, S. H. Simultaneous determination of platyphylloside, aceroside VIII and betulin in *Betula platyphylla* bark by HPLC-DAD. *Kor. J. Pharmacogn.* **45**, 294–299 (2014).
29. Wang, P., Wei, G. & Feng, L. Research advances in oxidosqualene cyclase in plants. *Forests* **13**, 1382 (2022).

30. Wang, Y. et al. Deletion and tandem duplications of biosynthetic genes drive the diversity of triterpenoids in *Aralia elata*. *Nat. Commun.* **13**, 2224 (2022).
31. Lertphadungkit, P., Qiao, X., Ye, M. & Bunsupa, S. Characterization of oxidosqualene cyclases from *Trichosanthes cucumerina* L. reveals key amino acids responsible for substrate specificity of isomultiflorenol synthase. *Planta* **256**, 58 (2022).
32. Yu, H. et al. Transcriptome analysis identifies putative genes involved in triterpenoid biosynthesis in *Platycodon grandifloras*. *Planta* **254**, 34 (2021).
33. Yoo, N. H., Kwon, Y. S. & Kim, M. J. Establishment of HPLC-UV analysis method validation for simultaneous analysis of standard compounds of *Oplopanax elatus* Nakai stem. *Kor. J. Pharm.* **50**, 133–140 (2019).
34. Mishra, S. et al. RNAi and homologous over-expression based functional approaches reveal triterpenoid synthase gene-cycloartenol synthase is involved in downstream withanolide biosynthesis in **Withania somnifera**. *PLoS ONE* **11**, e0149691 (2016).
35. Ramadan, A. M. et al. Control of β-sitosterol biosynthesis under light and watering in desert plant *Calotropis procera*. *Steroids* **141**, 1–8 (2019).
36. Haralampidis K, Trojanowska M, Osbourn AE. Biosynthesis of triterpenoid saponins in plants. His Trends Biopro Biotrans, 31–49. (2002)
37. Yu, Y. et al. Productive amyrin synthases for efficient alpha-amyrin synthesis in engineered *Saccharomyces cerevisiae*. *ACS Synth. Biol.* **7**, 2391–2402 (2018).
38. Lu, Y. et al. A multifunctional oxidosqualene cyclase from **Tripterygium regelii** that produces both alpha- and beta-amyrin. *RSC Adv.* **8**, 23516–23521 (2018).
39. Huang, Y. et al. Metabolic stimulation-elicited transcriptional responses and biosynthesis of acylated triterpenoids precursors in the medicinal plant *Helicteres angustifolia*. *BMC Plant Biol.* **22**, 86 (2022).
40. Huang, R. et al. Characterization of a group of 2,3-oxidosqualene cyclase genes involved in the biosynthesis of diverse triterpenoids of *Perilla frutescens*. *J. Agric. Food Chem.* **71**, 2523–2531 (2023).
41. Chen, T., Shu, X., Zhou, H., Beckford, F. A. & Misir, M. Algorithm selection for protein-ligand docking: Strategies and analysis on ACE. *Sci. Rep.* **13**, 8219 (2023).
42. Zhang, W., Bell, E. W., Yin, M. & Zhang, Y. EDock: Blind protein-ligand docking by replica-exchange Monte Carlo simulation. *J. Cheminform.* **12**, 37 (2020).
43. Yang, C., Chen, E. A. & Zhang, Y. Protein-ligand docking in the machine-learning era. *Molecules* **27**, 4568 (2022).
44. Kavitha, K., Revathi, K. & Tamiraikani, T. Exploring the potential of *Celtis aetnensis* compounds through molecular docking for *hepatocellular carcinoma*. *bioRxiv* **2024**, 2024.02. 28.582503 (2024).
45. Talukder, M. E. K. et al. Cheminformatics-based identification of phosphorylated RET tyrosine kinase inhibitors for human cancer. *Front. Chem.* **12**, 1407331 (2024).
46. Zhan, C. et al. Cytochrome P450 CYP716A254 catalyzes the formation of oleanolic acid from beta-amyrin during oleanane-type triterpenoid saponins biosynthesis in *Anemone flaccida*. *Biochem. Biophys. Res. Commun.* **495**, 1271–1277 (2018).
47. Alagna, F. et al. OeBAS and CYP716C67 catalyze the biosynthesis of health-beneficial triterpenoids in olive (*Olea europaea*) fruits. *New Phytol.* **238**, 2047–2063 (2023).
48. Fukushima, E. O. et al. CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. *Plant Cell Physiol.* **52**, 2050–2061 (2011).

Author contributions

Hong Ju Choi: Writing—original draft, Methodology, Formal analysis, Data curation. Ji Won Seo, Jiu Park, Won Hyeok Choi, Myong Jo Kim: Methodology. Eun Soo Seong: Writing—review & editing, Visualization, Project administration, Investigation, Conceptualization.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

As *O. elatus* is an endangered medicinal species, only in vitro–maintained mother plants were used, and no wild collection was performed.

Additional information

Correspondence and requests for materials should be addressed to E.S.S.

Reprints and permissions information is available at www.nature.com/reprints.

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