Genome-Wide Analysis of Oxidosqualene Cyclase Genes in *Artemisia annua*: Evolution, Expression, and Potential Roles in Triterpenoid Biosynthesis

Changfeng Guo , Si Xu , Xiaoyun Guo *

Posted Date: 9 June 2025

doi: 10.20944/preprints202506.0611.v1

Keywords: *Artemisia annua*; oxidosqualene cyclase (OSC); triterpenoids; gene family; genome-wide analysis

 doi:10.20944/preprints202506.0611.v1

Article

Genome-Wide Analysis of Oxidosqualene Cyclase Genes in *Artemisia annua*: Evolution, Expression, and Potential Roles in Triterpenoid Biosynthesis

Changfeng Guo [1], Si Xu [2] and Xiaoyun Guo [1,*]

[1] Guangxi Botanical Garden of Medicinal Plants, Nanning 530012, Guangxi, China
[2] School of Life Science and Bioengineering, Jining University, Qufu 273155, Shandong, China
* Correspondence: guoxy@gxyyzwy.com

Abstract: Plant triterpenoids are structurally diverse specialized metabolites with significant ecological, medicinal, and agricultural importance. Oxidosqualene cyclases (OSCs) catalyze the crucial cyclization step in triterpenoid biosynthesis, generating the fundamental carbon skeletons that determine their structural diversity and biological functions. Genome-wide identification of OSC genes was performed using bioinformatics tools, including HMMER and BLASTP, followed by phylogenetic analysis, gene structure analysis, conserved domain and motifs identification, cis-regulatory element prediction, protein-protein interaction analysis, and expression profiling using publicly available transcriptome data from UV-B treated *A. annua* six-week-old seedlings. We identified 24 AaOSC genes, classified into CAS, LAS, LUP, and unknown subfamilies. Phylogenetic analysis revealed evolutionary relationships with OSCs from other plant species. Gene structure analysis showed variations in exon-intron organization. Promoter analysis identified cis-regulatory elements related to light responsiveness, plant growth and development, hormone signaling, and stress response. Expression profiling revealed differential expression patterns of AaOSC genes under UV-B irradiation. This genome-wide characterization provides insights into the evolution and functional diversification of the OSC gene family in *A. annua*. The identified AaOSC genes and their regulatory elements lay the foundation for future studies aimed at manipulating triterpenoid biosynthesis for medicinal and biotechnological applications, particularly focusing on enhancing stress tolerance and artemisinin production.

Keywords: *Artemisia annua*; oxidosqualene cyclase (OSC); triterpenoids; gene family; genome-wide analysis

1. Introduction

Oxidosqualene cyclase (OSC) constitutes a family of enzymes that play pivotal roles in the biosynthesis of triterpenoids and sterols in plants. These enzymes catalyze the cyclization of the linear precursor 2,3-oxidosqualene into a diverse array of triterpene skeletons, which serve as the building blocks for various bioactive compounds. The importance of OSCs in plant metabolism is underscored by their involvement in the production of compounds with significant pharmacological activities, such as anti-inflammatory, anti-cancer, and antimicrobial properties [1–4].

Artemisia annua L., commonly known as sweet wormwood, is a medicinal plant renowned for its ability to produce artemisinin, a potent anti-malarial sesquiterpenoid lactone [5–7]. In recent years, there has been a growing interest in understanding the biosynthetic pathways and regulatory mechanisms underlying artemisinin production in *A. annua* [8–10]. While significant progress has been made in elucidating the enzymes involved in artemisinin biosynthesis, the role of OSCs in this process remains largely unexplored.

Triterpenoids are abundant in *A. annua* and contribute significantly to the plant's defense mechanisms and adaptation to environmental stress [11]. Despite their importance, the genome-wide

 doi:10.20944/preprints202506.0611.v1

identification, evolution, and expression patterns of OSC genes in A. annua have not been systematically studied. This knowledge gap limits our understanding of the triterpenoid biosynthetic pathways in this plant and hinders efforts to enhance the production of valuable triterpenoids through genetic engineering.

To date, more than 170 OSC genes have been biochemically characterized in plants, resulting in the identification of over 100 triterpene skeletons and a total of 61 distinct products[1,12]. The first β-amyrin synthase and cycloartenol synthase were isolated from Pisum sativum [13]. The cDNA of *Arabidopsis thaliana* was successfully transformed into yeast cells, leading to the identification of the first cycloartenol synthase gene in plants [14]. To date, a total of 13 OSC genes have been characterized in *Arabidopsis* [15,16]. The gene encoding β-amyrin synthase was isolated from ginseng roots through homologous cloning [17]. *PvOSC6* encoded β-amyrin synthase from *Prunella vulgaris*, while PvOSC2 functioned as a multifunctional synthase [18]. β-amyrin synthase genes have been identified in a variety of dicotyledonous plants [19–22], constituting the largest class of triterpenoid synthase genes discovered in plants to date. Moreover, β-amyrin synthase genes have also been cloned from monocotyledonous species [23,24]. Another significant OSC in plants is lupane synthase, which catalyzes the conversion of 2,3-oxidosqualene into lupeol, a pentacyclic triterpenoid. This enzyme was first isolated from *Olea europaea* and *Taraxacum officinale* [25]. Since then, lupane synthase genes have been identified and cloned from various plant species, including *Betula platyphylla*, *Glycyrrhiza glabra*, *Lotus japonicus*, *Bruguiera gymnorrhiza* and *Prunella vulgaris* [18,22,26,27]. Additionally, a multifunctional lupeol synthase, primarily producing lupeol with minor by-products, has been isolated from *Arabidopsis thaliana* and *Ricinus communis* [16,28,29].

James et al. successfully engineered triterpene production in *Saccharomyces cerevisiae* by utilizing a β-amyrin synthase derived from *A. annua*. The predicted protein showed a high degree of similarity to other β-amyrin synthases, sharing up to 86% amino acid sequence identity. Expression of this gene in *S. cerevisiae*, followed by GC/MS analysis, confirmed its function as a β-amyrin synthase. By manipulating key enzymes in the sterol pathway, such as 3-hydroxy-3-methylglutaryl-CoA reductase and lanosterol synthase, the researchers achieved a 50% increase in β-amyrin production, reaching levels of 6 mg/L in culture [30].

In a related study, Tessa et al. identified three OSC genes in *A. annua* that play crucial roles in the biosynthesis of specialized triterpenoids for the cuticle of aerial organs. Focusing on the plant's aerial organs, the authors identified two key enzymes, OSC2 and CYP716A14v2, through comparative transcriptome analysis of glandular and filamentous trichomes. OSC2, a multifunctional oxidosqualene cyclase, catalyzes the cyclization of 2,3-oxidosqualene into α-amyrin, β-amyrin, and δ-amyrin, with a preference for α-amyrin. CYP716A14v2, a cytochrome P450 oxidase, oxidizes the C-3 hydroxyl group of these triterpenes to ketones, yielding α-amyrone and β-amyrone. The other two OSC genes encoded cycloartenol synthase and lupeol synthase, respectively. This finding underscores the functional diversity within the OSC gene family. The identified OSC genes encode enzymes that produce triterpenoids with unique structures and functions, contributing to the formation of the wax layer on the cuticle of *A. annua*. This wax layer likely serves as a protective barrier against biotic and abiotic stress [11].

While these studies have provided valuable insights into the biochemical functions of individual OSC genes, a comprehensive genome-wide analysis of the OSC gene family in A. annua is still lacking. Such an analysis would enable a more thorough understanding of the genetic basis for triterpene biosynthesis in this medicinal plant, potentially leading to the discovery of novel OSC genes and the development of related biotechnological applications.

In this study, we aim to address this knowledge gap by conducting a genome-wide identification and characterization of the OSC gene family in *A. annua*. We will employ bioinformatics tools to identify all putative OSC genes in the *A. annua* genome and analyze their phylogenetic relationships, gene structures, conserved domain, motifs, amino acid sequences and protein 3D structure. Furthermore, we will investigate the expression patterns of these OSC genes across different tissues of *A. annua* using publicly available transcriptome data. This comprehensive analysis will offer

 doi:10.20944/preprints202506.0611.v1

valuable insights into the evolution and functional diversification of the OSC gene family in *A. annua*. It will lay the foundation for future studies focused on manipulating triterpenoid biosynthesis in this plant.

2. Materials and Methods

2.1. Genome Data Acquisition and OSC Gene Identification

The genome data and General Feature Format (GFF) of *Artemisia annua* were re-trieved from the Ensembl Plants database (https://plants.ensembl.org/index.html) and the Phytozome database (https://phytozome-next.jgi.doe.gov/). The OSC protein sequences of *Arabidopsis thaliana* and rice were obtained from the TAIR database (https://www.arabidopsis.org/) and the Rice Genome Annotation Project database (https://rice.uga.edu/), respectively. Hidden Markov Model (HMM) profiles for PF13243 and PF13249 were downloaded from the InterPro database (https://www.ebi.ac.uk/interpro/). Subsequently, HMMER 3.0 software was employed to identify AaOSC proteins in *Artemisia annua* with an *E*-value threshold of $\leqslant 1 \times 10^{-5}$ and a sequence similarity exceeding 50%. Additionally, the BLASTP algorithm was utilized to search for homologous protein sequences in *Artemisia annua*, using the OSC protein sequences of *Arabidopsis thaliana* and rice as queries, under the same *E*-value and similarity thresholds. All candidate AaOSC protein sequences were further analyzed to confirm the presence of the OSC domain. Finally, the longest transcript for each gene was selected using the R package seqfinder (https://github.com/yueliu1115/seqfinder).

2.2. Gene Structure, Protein Feature Analysis, and Phylogenetic Analysis

The domain architectures of gene family members were predicted using Pfam-scan (v1.6) against the Pfam database. For motif discovery, the MEME Suite (v5.5.2, http://memesuite.org/) was utilized with a maximum of 10 motifs. Gene structural features (exon–intron organization) were retrieved from the species' genome annotation file in GFF3 format. The SeaView software (v5.1) was utilized to conduct multiple sequence alignment. A phylogenetic tree was constructed in MEGA-X (v11.0.13) using the neighbor-joining method with 1,000 bootstrap replicates. The visualization of the phylogenetic tree was performed using the FigTree (v1.4.4).

2.3. Promoter Cis-Element and Functional Annotation

The 2000 bp promoter sequences of AaOSC family genes were extracted from the genomic files (GFF). The website of PlantCARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was utilized to conduct a statistical analysis of cis-element prediction for each gene promoter. Gene Ontology (GO) functional annotation of the plant species examined in this study was performed using the eggNOG-mapper database [31]. Subsequent statistical analyses and data visualization were carried out in R. To explore the regulatory network of AaOSC protein, we used AraNet2 database to predict the interaction network of AaOSC protein. In addition, we annotated the GO and KEGG results of *Artemisia annua* AaOSC using eggNOG-mapper database and used R software (4.3.0) for statistics and visualization.

2.4. Expression Profiling

To explore the expression pattern of the AaOSC gene during photomorphogenesis, RNA-seq datasets of *Artemisia annua* exposed to UV-B radiation were retrieved from the NCBI SRA database (accession PRJNA669851). Six-week-old seedlings were subjected to UV-B treatment under standard growth conditions using a narrow-band UV-B lamp (Philips TL20W/01RS; 1.5 $\mu mol \cdot m^{-2} \cdot s^{-1}$). Leaf samples were collected from the youngest fully expanded leaf at 0, 2, 4, and 6 h post-treatment for downstream expression analysis. The transcriptome data were aligned and quantified using the genome alignment software HISAT2 (v2.2.0) and featureCounts (v1.6.4) for subread regions. The R package Pheatmap (v1.0.12) was employed to visualize the resulting data.

 doi:10.20944/preprints202506.0611.v1

3. Results

3.1. Genome-Wide Identification of AaOSCs and Physical and Chemical Characteristics

The physicochemical properties of the oxidosqualene cyclase (OSC) gene family were analyzed across five plant species: *Artemisia annua*, *Arabidopsis thaliana*, *Helianthus annuus*, *Oryza sativa Japonica*, and *Vitis vinifera*. The analysis encompassed gene length, molecular weight (MW), hydrophobicity, and isoelectric point (pI) (Figure 1). The distribution of gene lengths varied among the species. *Artemisia annua* exhibited the largest range in gene length, from approximately 300 to nearly 1000 amino acids. *Arabidopsis thaliana* showed a relatively constrained distribution, clustered around 700 amino acids. *Helianthus annuus*, *Oryza sativa Japonica*, and *Vitis vinifera* displayed intermediate ranges and median values. Similar to gene length, the MW distribution was broadest in *Artemisia annua*, ranging from approximately 30 kDa to >100 kDa. The remaining species exhibited narrower MW distributions, with medians between approximately 60 kDa and 80 kDa. The predicted hydrophobicity of the OSC proteins was generally negative across all species, indicating a predominantly hydrophobic nature. *Vitis vinifera* displayed the widest range of hydrophobicity values, extending to approximately 0.0, whereas *Arabidopsis thaliana* showed the narrowest and most negative range. The predicted pI values of the OSC proteins ranged from approximately 5.5 to 8.5. *Artemisia annua* showed the widest distribution of pI values, with some members approaching a pI of 9. *Vitis vinifera* exhibited a subset of members with pI values exceeding 8. The other species exhibited more constrained pI ranges centered around a pI of 6.5.

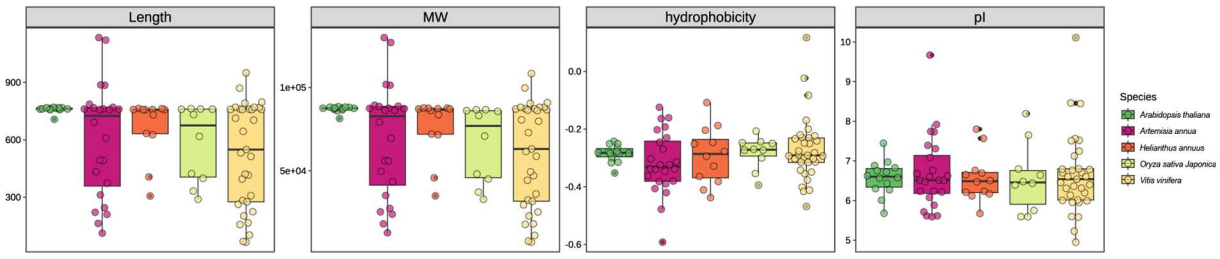

Figure 1. Comparative analysis of physicochemical properties of oxidosqualene cyclase (OSC) genes across five plant species. Gene length distribution (amino acid count). Molecular weight (MW, kDa) of OSC proteins. Hydrophobicity (GRAVY index). Isoelectric point (pI). Species: *Arabidopsis thaliana*, *Artemisia annua*, *Helianthus annuus*, *Oryza sativa Japonica*, and *Vitis vinifera*.

3.2. Gene Structure, Conserved Domain and Motif Analysis

The phylogenetic relationships, conserved domains, motif compositions, and gene structures of 24 AaOSC genes were analyzed (Figure 2). Phylogenetic analysis grouped the AaOSC genes into four clades: CAS (Cycloartenol Synthase), LAS (Lanosterol Synthase), LUP (Lupeol Synthase), and a group of genes with unknown functions. Bootstrap values greater than or equal to 75% were observed for most nodes. Nodes with bootstrap values between 45% and 75% were also indicated. Domain analysis revealed the presence of characteristic OSC domains within the predicted protein sequences. Specifically, the SQHop cyclase N-terminal domain and SQHop cyclase C-terminal domain were commonly observed. The length of the predicted protein varied, CTI12_AA467020 from the unknown clade possesses an F-box domain and an FBA 3 at the C-terminus. This domain architecture is distinct from other members. The distribution of motifs varied across different clades. Genes within the CAS clade shared a similar motif composition. However, the number and order of these motifs can differ within the clades. Gene lengths ranged from 1.5 kb (CTI12_AA425640, encoding 493 residues) to >3.3 kb (CTI12_AA2077660, encoding 1,121 residues). Exon numbers varied between 12 (CTI12_AA467020) and 26 (CTI12_AA207660), with intron lengths showing significant divergence, particularly in CTI12_AA475450 and CTI12_AA207660.

 doi:10.20944/preprints202506.0611.v1

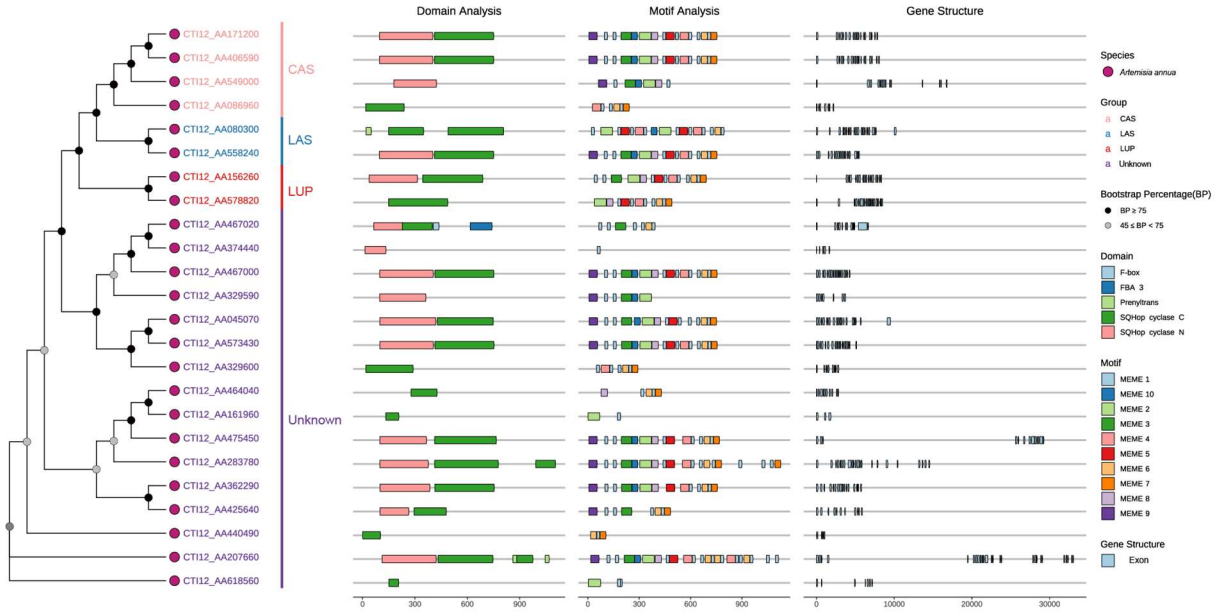

Figure 2. Phylogenetic relationships, domain architecture, conserved motifs, and gene structure of AaOSC genes.

3.3. Promoter Cis-Elements

The cis-regulatory elements (CREs) present within the promoter regions of 24 AaOSC genes were analyzed to predict potential regulatory mechanisms (Figure 3). The analysis focused on CREs associated with light responsiveness, plant growth and development, plant hormone signaling, and stress response. The number of each type of CRE was quantified for each AaOSC gene promoter. Genes within the CAS clade generally had a higher proportion of CREs associated with light responsiveness compared to other clades. In contrast, genes within the LUP clade showed a relatively high proportion of CREs related to plant growth and development. Notably, the LAS clade demonstrated heightened enrichment of stress response associated. The promoter of CTI12_AA329600 (unknown clade) harbored the highest number of TATA-box elements (n=297). CTI12_AA161960, belonging to the unknown clade, contains the highest number of AAGAA-motif (n=5). The number of plant hormone-related elements, including ABRE, CGTCA-motif, and TGACG-motif, was assessed. CTI12_AA467020 (unknown clade) has 11 ARE elements within its promoter region. Several CREs associated with stress responses were identified, including ARE, as-1, TC-rich repeats, A-box, LTR, and WUN-motif. CTI12_AA549000 (CAS clade) has the highest number of ABRE elements (n=6). The gene CTI12_AA080300 and CTI12_558240 (LAS clade) has a remarkable number of ARE elements (n=5).

 doi:10.20944/preprints202506.0611.v1

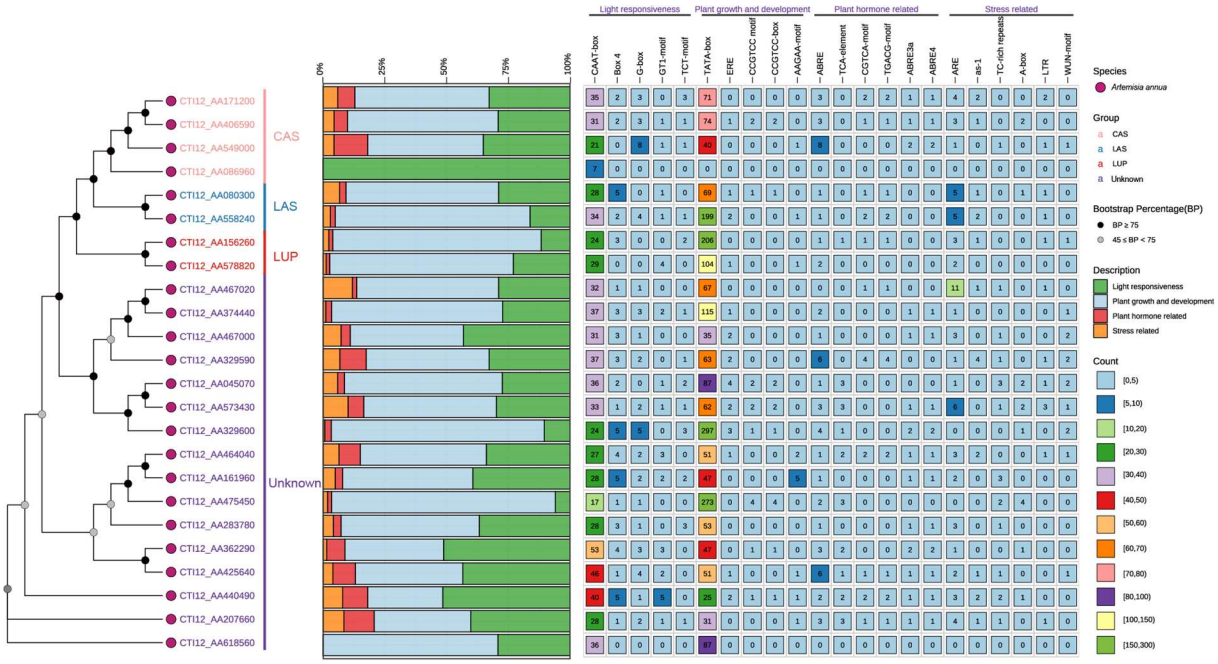

Figure 3. Cis-regulatory element analysis of AaOSC gene promoters. The phylogenetic tree on the left indicates the relationships between the 24 AaOSC genes analyzed. The stacked bar chart displays the relative proportion of CRE categories (light responsiveness, plant growth and development, plant hormone related, and stress related). The heatmap represents the number of each CRE type found in the promoter region of each gene. The color intensity in the heatmap corresponds to the number of elements as indicated in the legend.

3.4. Phylogenetic Analysis

A phylogenetic tree was constructed to investigate evolutionary relationships and other functionally characterized OSCs from diverse plant species (Figure 4). In the phylogenetic reconstruction, putative pseudogenes encoding shorter proteins were excluded through stringent quality control, with 12 members ultimately incorporated into the final phylogenetic tree. The analysis revealed distinct clades corresponding to known OSC functionalities, as well as subclades containing AaOSC genes. The tree exhibited clear separation of OSCs based on their known enzymatic activity. The AaOSC genes CTI12_AA578820 and CTI12_AA156260 cluster within the lupeol synthase clade. These AaOSCs exhibited strong bootstrap support for their placement within this clade, indicating high confidence in their relatedness to known lupeol synthases. CTI12_AA171200 and CTI12_AA406590 are grouped within the cycloartenol synthase clade, sharing a close evolutionary relationship with cycloartenol synthases from other plant species. CTI12_AA558240 falls within the lanosterol synthase clade, which is known to produce lanosterol, a precursor to steroidal compounds. CTI12_AA283780, CTI12_AA475450, CTI12_AA362290, CTI12_AA207660, CTI12_AA045070, CTI12_AA573430, and CTI12_AA467000 are positioned among clades producing diverse products without clustering strongly with any functionally defined groups, suggesting a potential for neofunctionalization.

 doi:10.20944/preprints202506.0611.v1

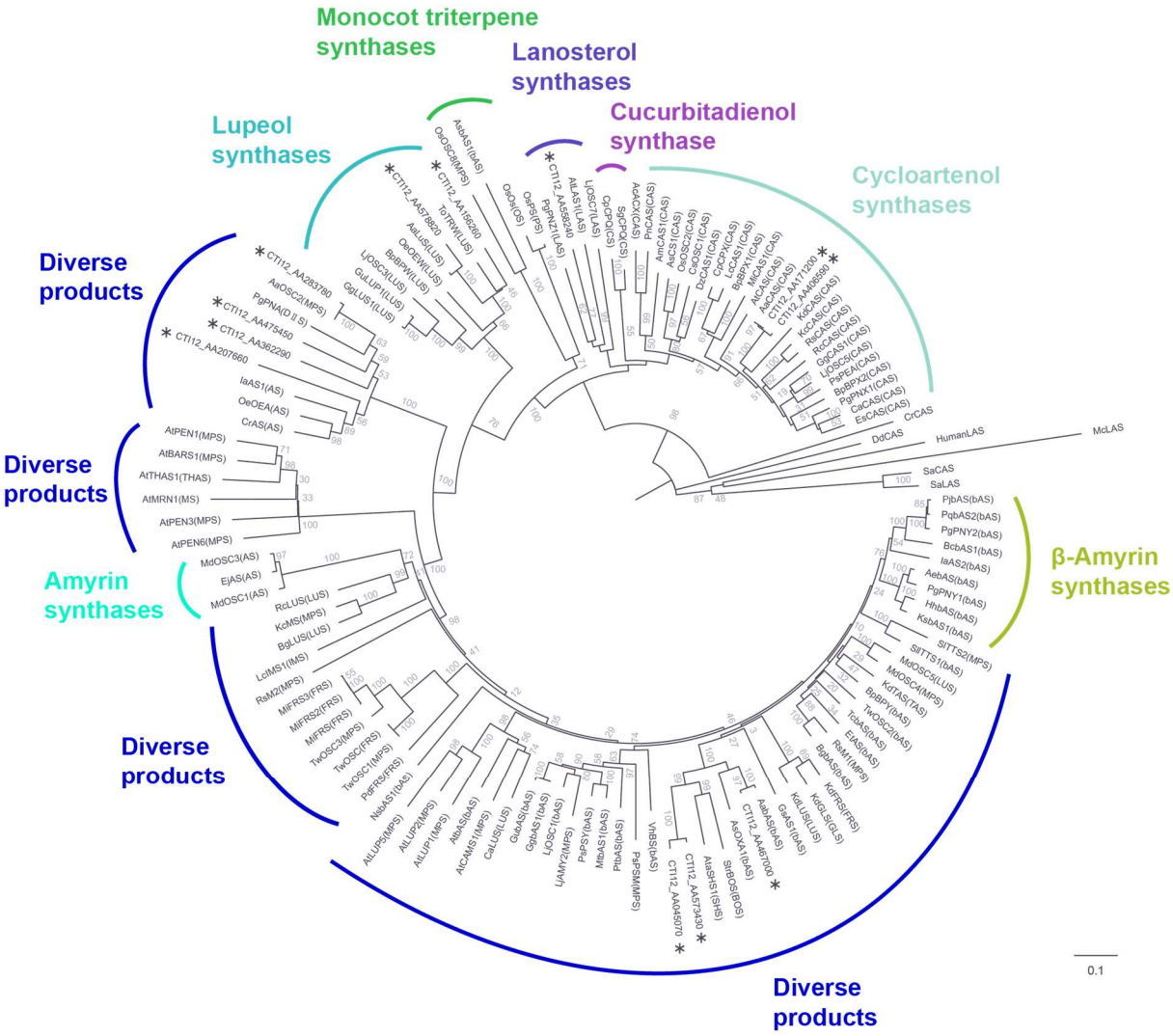

Figure 4. Phylogenetic relationships of AaOSC genes. The tree was constructed using neigh-bor-joining method based on the amino acid sequences of 12 AaOSC genes and other OSCs from various plant species. AaOSC genes are marked with asterisks. Numbers at the nodes indicate bootstrap support values (%).

3.5. GO Annotation

To elucidate the functional implications of the 24 AaOSC genes, Gene Ontology (GO) enrichment analysis was performed (Figure 5). Enriched GO terms were categorized into three main ontologies: Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). As illustrated in Figure 5, the size of the circles corresponds to the number of genes associated with each GO term, indicating the relative enrichment. Within the biological process category, “pentacyclic triterpenoid biosynthetic process” showed the highest enrichment. Other significantly enriched biological processes included “pollen development” and “thylakoid membrane organization.” The most enriched cellular component term was “vacuole.” In the molecular function category, several OSC activities were enriched, with “beta-amyrin synthase activity” being the most prominent. Other notable molecular functions identified included “lupeol synthase activity”, “camelliol c synthase activity”, “cycloartenol synthase activity”, “lanosterol synthase activity”, “lyase activity”, “thalianol synthase activity”, and “baruol synthase activity”.

 doi:10.20944/preprints202506.0611.v1

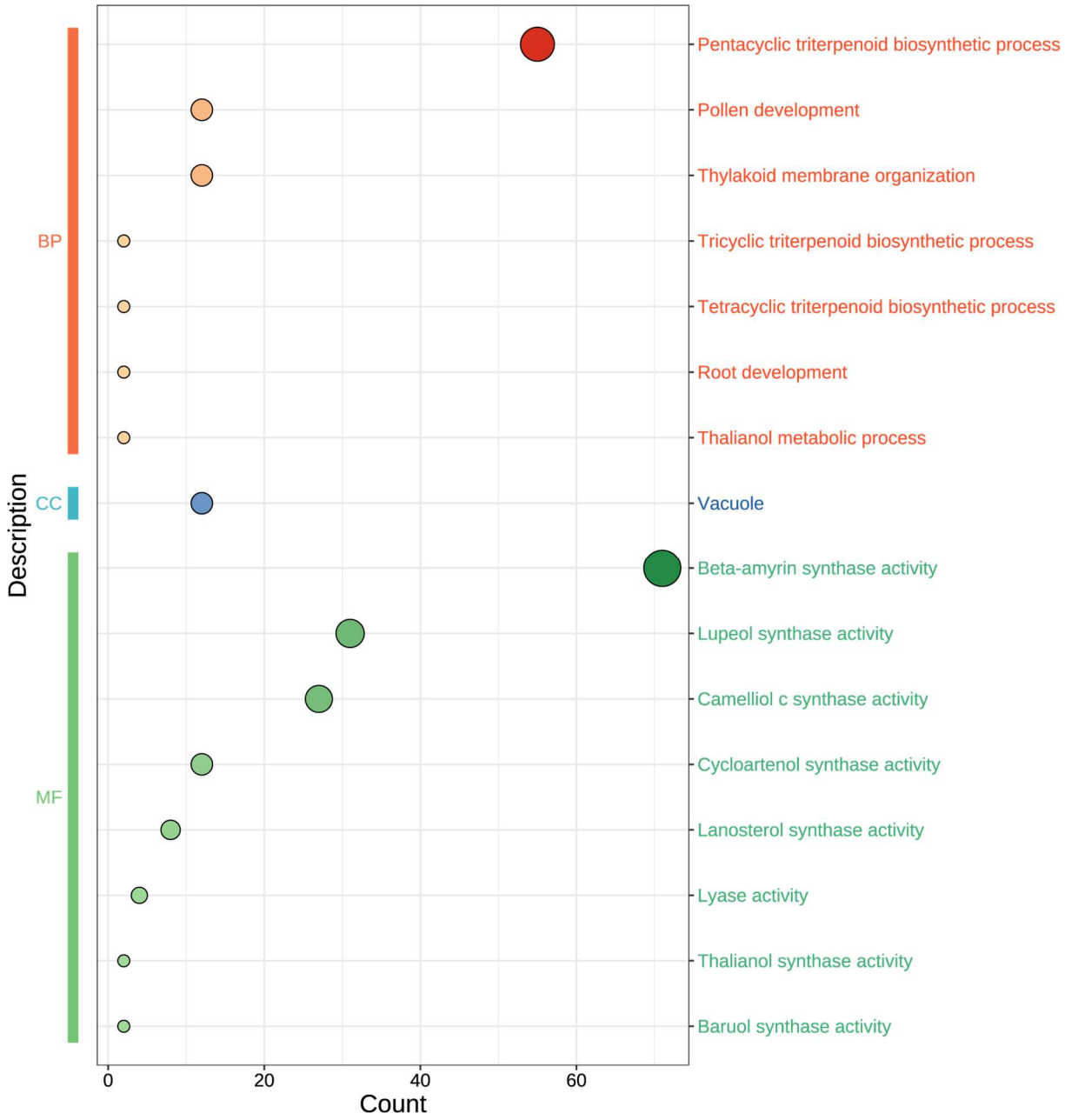

Figure 5. Gene Ontology (GO) enrichment analysis of AaOSC genes. GO terms are categorized into Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). The size of each circle represents the number of AaOSC genes associated with the corresponding GO term, and the x-axis indicates the number of genes associated with each GO term.

3.6. Protein-Protein Intreraction

The PPI network displays significant interconnectivity among AaOSC family members. Proteins containing SQHop cyclase domains (specifically SQHop cyclase N, SQHop cyclase C, and combinations) constitute a major interacting group within the network. Nodes representing proteins with both SQHop cyclase N and SQHop cyclase C domains display consistently high popularity. OSC proteins annotated with both SQHop cyclase N and SQHop cyclase C domains are observed with interaction weights ranging from 4.25 to 4.30, connecting them to proteins with similar domain architectures. A distinct subnetwork is formed by proteins containing the Prenyltrans domain. This subnetwork is closely associated with SQHop cyclase domain-containing proteins. One notable observation is the strong interaction (weight = 4.30) between a protein containing both Prenyltrans and SQHop cyclase C domains and proteins containing only the Prenyltrans domain. The ABC

 doi:10.20944/preprints202506.0611.v1

membrane domain-containing proteins show peripheral locations in the network, linked to the SQHop cyclase cluster through lower-weight interactions.

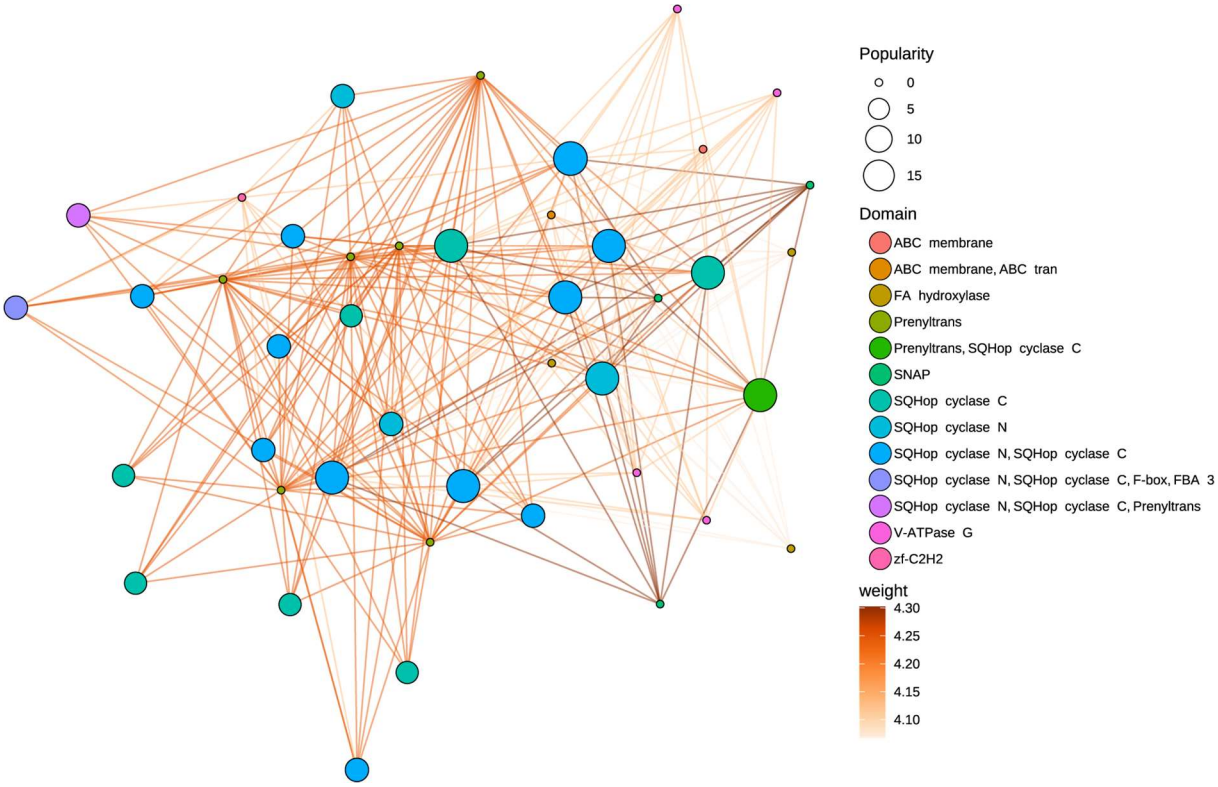

Figure 6. Predicted protein-protein interaction (PPI) network of AaOSC genes. Nodes represent proteins, and edges represent predicted interactions. Node size indicates 'popularity' (number of interactions). Node color indicates the protein domain. Edge color indicates the interaction weight based on a gradient from light orange (lower weight) to dark orange (higher weight).

3.7. Expression Pattern in Tissues and Under UV Pressure

The expression patterns of AaOSC genes were analyzed under various UV irradiation treatments (0 hours, 2 hours, 4 hours, and 6 hours of UV exposure) using heatmap visualization (Figure 7). A phylogenetic tree showing the relationships among the analyzed AaOSC genes is displayed alongside the heatmap. The heatmap reveals distinct expression patterns across the different AaOSC gene family members and UV treatments. Most of the genes show differences in gene expression between the different UV exposure levels. Genes belonging to the CAS and LAS subfamilies generally exhibited relatively low expression levels across all UV treatments compared to the other groups. Notably, CTI12_AA406590 exhibited relatively higher expression in the UV4 condition compared to other CAS family members. The expression of the LUS subfamily showed variation under different UV treatments. CTI12_AA156260 displayed relatively high expression in the UV0 condition, while CTI12_AA578820 appeared to be upregulated in the UV2 condition. Genes from the unknown subfamily showed distinct expression patterns. Notably, CTI12_AA467020 was upregulated in the UV4 conditions. Several genes within the unknow subfamily showed minimal changes in expression. However, CTI12_AA440490 expression notably increased in UV4.

 doi:10.20944/preprints202506.0611.v1

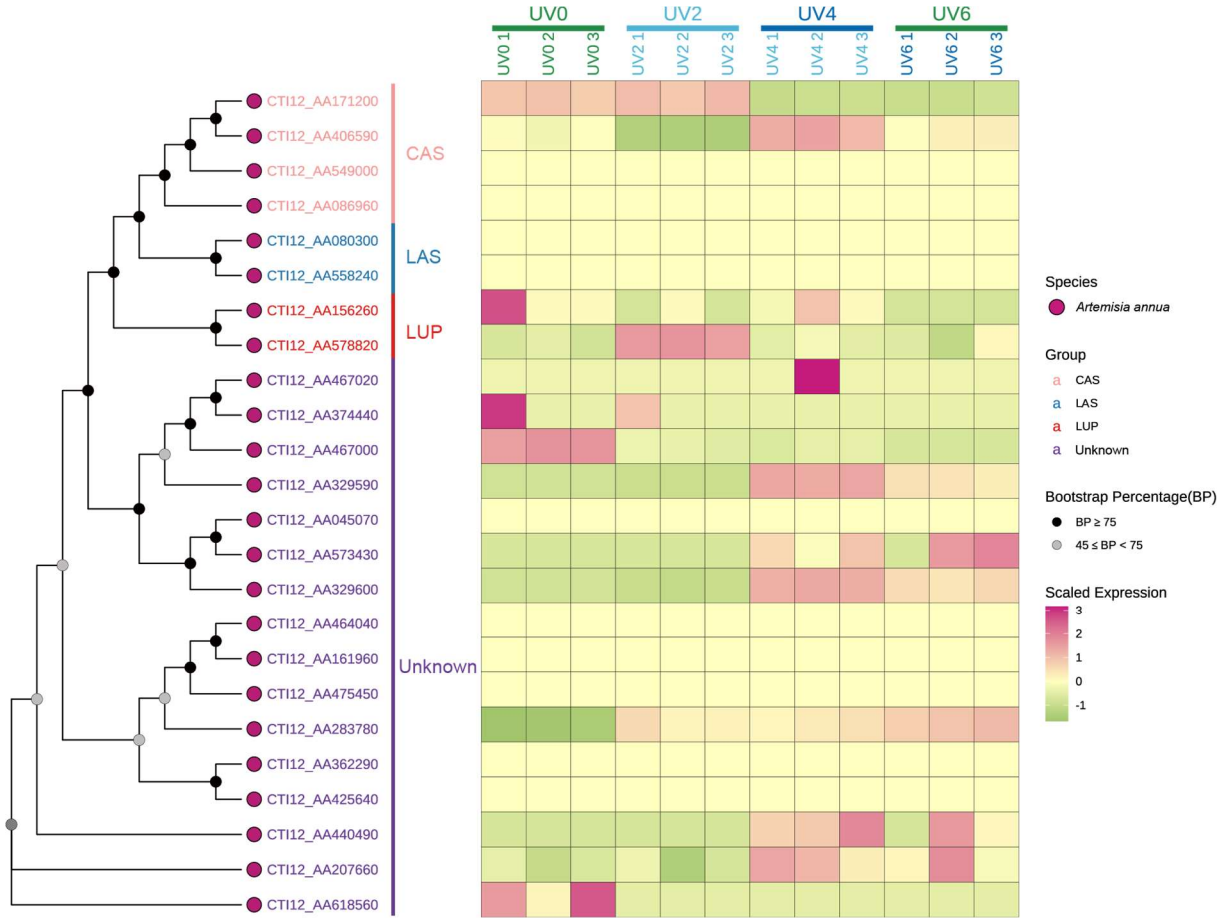

Figure 7. Heatmap representation of the expression profiles of AaOSC genes under UV irradiation. Columns correspond to different UV treatment conditions (0 hours, 2 hours, 4 hours, and 6 hours of UV exposure), with numbers indicating replicates. Bootstrap values for the phylogenetic tree are indicated by node shading.

4. Discussion

Oxidosqualene cyclase (OSC) is a key enzyme in plant secondary metabolism that controls triterpene and sterol biosynthesis. As a membrane-associated enzyme, OSC catalyzes the stereospecific cyclization of 2,3-oxidosqualene into triterpenoid scaffolds (e.g., α-amyrin, β-amyrin, lupeol, taraxerol, friedelin) or sterol precursors (e.g., cycloartenol, lanosterol), depending on its isoform specific activity [3,32,33]. This step regulates metabolic flux toward defense-related triterpenoids (e.g., saponins) or essential phytosterols (e.g., sitosterol, stigmasterol), balancing plant growth and environmental adaptation [32]. Our genome-wide analysis identified 24 OSC genes (AaOSCs) in *Artemisia annua*, revealing a metabolic toolkit more expansive than model species like *Arabidopsis thaliana* [34–36]. This expansion, particularly in clades associated with specialized triterpenoids (e.g., cycloartenol, lanosterol, lupeol), aligns with *A. annua*'s ecological adaptability and its production of pharmacologically active compounds, including artemisinin precursors [34,37–39]. Notably, the LUP clade encompasses evolutionarily conserved genes that encode enzymes responsible for synthesizing lupeol and its derivatives [25,32]. These bioactive triterpenoids serve critical defensive roles in plants, such as antimicrobial protection, ROS scavenging, and herbivore deterrence [32,40,41]. The multifunctionality of lupeol derivatives highlights their ecological significance in plant stress adaptation. These findings underscore the evolutionary diversification of OSCs in *A. annua*, potentially driven by selective pressures to optimize defense metabolites while maintaining sterol homeostasis [32,41].

The phylogenetic segregation of AaOSCs into CAS, LAS, LUP, and unknown subfamilies mirrors catalytic specialization observed in other plants but reveals unique features. The potential presence of LAS-like enzymes in *A. annua* raises intriguing questions about horizontal gene transfer

 doi:10.20944/preprints202506.0611.v1

(HGT) from fungi or animals, a mechanism documented in terpene synthase evolution [3]. This HGT event may have conferred adaptive advantages under abiotic stress, as lanosterol-derived sterols enhance membrane stability in fluctuating environments [42–44]. This is a hypothesis supported by the enrichment of abiotic stress-responsive cis-elements in CTI12_AA080300 and CTI12_AA558240 promoters. This evolutionary plasticity aligns with the "gene toolkit" model, where HGT expands metabolic repertoires in sessile organisms facing environmental pressures [34,35,43]. Experimental validation of CTI12_AA558240's activity could redefine our understanding of sterol pathway plasticity in plants. Additionally, exon-intron variability among AaOSCs suggests regulatory divergence, such as tissue-specific splicing, which may fine-tune triterpenoid production under stress [38,45,46]. The discovery of an F-box domain in CTI12_AA467020 further hints at post-translational regulation via ubiquitination, a mechanism known to modulate secondary metabolism in response to environmental cues [47–49]. The structural plasticity of unknown clade members, positioned diverse products clades, highlights a "transitional zone" for catalytic promiscuity [35,50]. Minor active-site variations, including residue substitutions in the QW motif, DCTAE loop, or other regions, could enable product switching, a mechanism proposed to drive metabolic diversity in *Arabidopsis*, *Avena strigose* and rice [3,35,51–53]. This plasticity may explain the expansion of triterpenoid profiles in *A. annua*, a species renowned for its ecological resilience [35,38].

Promoter analysis uncovered clade-specific cis-elements, linking transcriptional regulation to physiological roles. Promoter cis-element analysis revealed that AaOSCs are tightly regulated by hormonal and environmental cues. These findings echo studies in *Glycyrrhiza uralensis*, where ABA and JA signaling co-regulate triterpenoid biosynthesis under stress [54]. The enrichment of ABA/JA-responsive motifs (TGACG/CGTCA motif) in LUP promoters aligns with lupeol's role in biotic stress responses [55], while light-responsive elements in CAS promoters may synchronize sterol biosynthesis with photosynthetic activity [42,56]. Strikingly, UV-B-responsive HY5-binding sites (G-box) prevalent across AaOSCs suggest co-regulation with photoprotective pathways, akin to UV-induced flavonoid biosynthesis in grapevines [57,58]. These regulatory signatures, combined with the abiotic stress motifs in LAS promoters (e.g., MYC, LTR), position AaOSCs at the nexus of environmental adaptation, enabling *A. annua* to dynamically allocate resources between growth and defense.

The low basal expression of most CAS subfamily members aligns with observations in *Arabidopsis* and *Medicago*, where CAS-like enzymes are predominantly regulated by developmental cues rather than abiotic stress [59–61]. The transient upregulation of CTI12_AA171200 and CTI12_AA406590 at UV2 suggests a stress-responsive role distinct from canonical CAS functions, possibly linked to triterpenoid phytoalexin biosynthesis hypothesized in plants [3,60]. The upregulation of LUP subfamily members, specifically CTI12_AA578820 in UV4 aligns with the known role of lupane-type triterpenoids in UV-B shielding and membrane stabilization [62]. This response mirrors the UV-inducible β-amyrin synthase activity in *Glycyrrhiza uralensis* [63], suggesting conserved adaptive strategies across eudicots [33,40]. Notably, the unknown subfamily member CTI12_AA440490 exhibits UV4-specific upregulation. Phylogenetic divergence from characterized OSCs suggests this isoform may catalyze novel triterpenoid scaffolds, analogous to the recently discovered marneral synthase in *A. thaliana* [64,65]. Such cryptic metabolic pathways could represent untapped reservoirs for bioactive compound discovery, particularly in stress-adapted species like *A. annua* [37].

While our study provides a genomic foundation, key gaps remain. First, functional characterization of the unknown AaOSCs is critical; heterologous expression in yeast or *Nicotiana benthamiana* could elucidate their products, though in planta assays may better capture substrate availability and regulatory interactions [66–70]. Second, cis-element predictions do not account for epigenetic modifications or transcription factor availability; chromatin immunoprecipitation (ChIP-seq) or CRISPR-edited promoter studies could validate motif functionality [71–73]. Third, expression profiling under diverse stresses (e.g., herbivory, drought) and tissues (e.g., glandular trichomes, roots) is needed to resolve spatiotemporal roles, particularly for clades with weak phylogenetic clustering

 doi:10.20944/preprints202506.0611.v1

[34,74,75]. Finally, protein structure prediction of AaOSCs (e.g., via AlphaFold2) could reveal catalytic residues and substrate-binding pockets, enabling rational engineering to enhance triterpenoid yields [76,77].

The AaOSC repertoire offers a roadmap for metabolic engineering. CRISPR-Cas9-mediated knockout of CAS or LAS genes could redirect flux toward artemisinin precursors or novel triterpenoids, while overexpression of LUP clade genes might bolster stress tolerance. Furthermore, resolving the evolutionary origin of LAS-like enzymes through comparative genomics could uncover HGT mechanisms applicable to crop improvement. Beyond A. annua, this study underscores the value of integrating phylogenomics, regulatory element analysis, and structural modeling to unravel metabolic diversity in medicinal plants.

5. Conclusions

This study provides a comprehensive genome-wide analysis of the OSC gene family in *Artemisia annua*, revealing a repertoire of 24 AaOSC genes with diverse phylogenetic relationships, structural features, and expression patterns. The classification of these genes into CAS, LAS, LUP, and unknown subfamilies suggests functional specialization in sterol and triterpenoid biosynthesis. The identification of cis -regulatory elements in AaOSC promoters highlights their responsiveness to light, hormones, and stress signals, indicating a critical role in environmental adaptation. Expression analysis under UV-B irradiation revealed differential regulation of AaOSC genes, further supporting their involvement in stress responses. The presence of LAS-like enzymes raises the possibility of horizontal gene transfer and highlights evolutionary plasticity in this medicinal plant. The AaOSC repertoire offers a roadmap for metabolic engineering strategies aimed at enhancing triterpenoid production and stress tolerance in *A. annua*. Future research should focus on functional characterization of the unknown AaOSCs, validation of cis-element functionality, and expression profiling under diverse stress conditions to fully elucidate the spatiotemporal roles of these genes.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, Table S1: Information on the OSC genes of other plants within the OSC phylogenetic tree; Table S2: Expression data for AaOSC genes under UV stress; Data 1: 24 CDS sequences of AaOSC genes in Artemisia annua L.; Data 2: Physical and chemical characteristics of the AaOSCs.

Author Contributions: Conceptualization, C.G. and X.G.; methodology, C.G. and S.X.; formal analysis, C.G. and S.X.; writing—original draft preparation, C.G.; writing—review and editing, X.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Guangxi Science and Technology Base and Talent Special Project (Grant No. Guike AD22080016), the Guangxi Qihuang Scholars Training Program (Grant No. GXQH202402), and the Guangxi Major Science and Technology Special Project (Grant No. Guike AA22096021).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and materials that support the findings of this study were available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, Y.; Wang, J.; Li, L.; Song, W.; Li, M.; Hua, X.; Wang, Y.; Yuan, J.; Xue, Z. Natural Products of Pentacyclic Triterpenoids: from Discovery to Heterologous Biosynthesis. *Nat. Prod. Rep.* **2023,** *40*, 1303-1353.
2. Dahlin, P.; Srivastava, V.; Bulone, V.; McKee, L. S. The Oxidosqualene Cyclase from the Oomycete Saprolegnia parasitica Synthesizes Lanosterol as a Single Product. *Front. Microbiol.* **2016,** *7*, 1802.

 doi:10.20944/preprints202506.0611.v1

3. Thimmappa, R.; Geisler, K.; Louveau, T.; O'Maille, P.; Osbourn, A. Triterpene Biosynthesis in Plants. *Annu. Rev. Plant Biol.* **2014,** *65*, 225-257.
4. Thoma, R.; Schulz-Gasch, T.; D'Arcy, B.; Benz, J.; Aebi, J.; Dehmlow, H.; Hennig, M.; Stihle, M.; Ruf, A. Insight into Steroid Scaffold Formation from the Structure of Human Oxidosqualene Cyclase. *Nature* **2004,** *432*, 118-122.
5. Wang, J.; Xu, C.; Lun, Z.-R.; Meshnick, S. R. Unpacking 'Artemisinin Resistance'. *Trends Pharmacol. Sci.* **2017,** *38*, 506-511.
6. Tu, Y. The Discovery of Artemisinin (qinghaosu) and Gifts from Chinese Medicine. *Nat. Med.* **2011,** *17*, 1217-1220.
7. White, N. J. Qinghaosu (Artemisinin): The Price of Success. *Science* **2008,** *320*, 330-334.
8. Li, Y.; Yang, Y.; Li, L.; Tang, K.; Hao, X.; Kai, G., Advanced Metabolic Engineering Strategies for Increasing Artemisinin Yield in Artemisia annua L. *Hortic. Res.* **2024,** *11*, uhad292.
9. Zhang, F.; Lu, X.; Lv, Z.; Zhang, L.; Zhu, M.; Jiang, W.; Wang, G.; Sun, X.; Tang, K. Overexpression of the Artemisia Orthologue of ABA Receptor, AaPYL9, Enhances ABA Sensitivity and Improves Artemisinin Content in Artemisia annua L. *PLoS ONE* **2013,** *8*, e56697.
10. Yu, Z.-X.; Li, J.-X.; Yang, C.-Q.; Hu, W.-L.; Wang, L.-J.; Chen, X.-Y. The Jasmonate-Responsive AP2/ERF Transcription Factors AaERF1 and AaERF2 Positively Regulate Artemisinin Biosynthesis in Artemisia annua L. *Mol. Plant* **2012,** *5*, 353-365.
11. Moses, T.; Pollier, J.; Shen, Q.; Soetaert, S.; Reed, J.; Erffelinck, M.-L.; Van Nieuwerburgh, F. C. W.; Vanden Bossche, R.; Osbourn, A.; Thevelein, J. M.; Deforce, D.; Tang, K.; Goossens, A. OSC2 and CYP716A14v2 Catalyze the Biosynthesis of Triterpenoids for the Cuticle of Aerial Organs of Artemisia annua. *The Plant Cell* **2015,** *27*, 286-301.
12. Chen, K.; Zhang, M.; Ye, M.; Qiao, X. Site-directed Mutagenesis and Substrate Compatibility to Reveal the Structure–function Relationships of Plant Oxidosqualene Cyclases. *Nat. Prod. Rep.* **2021,** *38*, 2261-2275.
13. Abe, I.; Sankawa, U. Purification and Properties of Squalene-2, 3-epoxide Cyclases from Pea Seedlings. *CHEM PHARM BULL* **1992**, 40, 1755-1760.
14. E J, C.; S P, M.; B, B. Isolation of an Arabidopsis thaliana Gene Encoding Cycloartenol Synthase by Functional Expression in a Yeast Mutant Lacking Lanosterol Synthase by the Use of a Chromatographic Screen. *Proc. Natl. Acad.Sci. USA* **1993**, *90*, 11628-11632.
15. Morlacchi, P.; Wilson, W. K.; Xiong, Q.; Bhaduri, A.; Sttivend, D.; Kolesnikova, M. D.; Matsuda, S. P. T. Product Profile of PEN3: The Last Unexamined Oxidosqualene Cyclase in Arabidopsis thaliana. *Org. Lett.* **2009**, *11*, 2627-2630.
16. Segura, M. J. R.; Meyer, M. M.; Matsuda, S. P. T. Arabidopsis thaliana LUP1 Converts Oxidosqualene to Multiple Triterpene Alcohols and a Triterpene Diol. *Org. Lett.* **2000,** *2*, 2257-2259.
17. Zhang, H.; Hua, X.; Zheng, D.; Wu, H.; Li, C.; Rao, P.; Wen, M.; Choi, Y.-E.; Xue, Z.; Wang, Y.; Li, Y. De Novo Biosynthesis of Oleanane-Type Ginsenosides in Saccharomyces cerevisiae Using Two Types of Glycosyltransferases from Panax ginseng. *J. Agric. Food. Chem.* **2022,** *70*, 2231-2240.
18. Zhang, S.; Meng, F.; Pan, X.; Qiu, X.; Li, C.; Lu, S. Chromosome-level Genome Assembly of Prunella vulgaris L. Provides Insights into Pentacyclic Triterpenoid Biosynthesis. *Plant J.* **2024,** *118*, 731-752.
19. Vernoud, V. β-Amyrin Synthase1 Controls the Accumulation of the Major Saponins Present in Pea (Pisum sativum). *Plant Cell Physiol.* **2021**, *62*, 784-797..
20. Xue, Z.; Duan, L.; Liu, D.; Guo, J.; Ge, S.; Dicks, J.; ÓMáille, P.; Osbourn, A.; Qi, X. Divergent Evolution of Oxidosqualene Cyclases in Plants. *New Phytol.* **2011,** *193*, 1022-1038.

 doi:10.20944/preprints202506.0611.v1

21. Meesapyodsuk, D.; Balsevich, J.; Reed, D. W.; Covello, P. S. Saponin Biosynthesis in Saponaria vaccaria. cDNAs Encoding β-Amyrin Synthase and a Triterpene Carboxylic Acid Glucosyltransferase. *Plant Physiol.* **2007,** *143*, 959-969.
22. Sawai, S.; Akashi, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Ayabe, S.-i.; Aoki, T. Plant Lanosterol Synthase: Divergence of the Sterol and Triterpene Biosynthetic Pathways in Eukaryotes. *Plant Cell Physiol.* **2006,** *47*, 673-677.
23. Sun, J.; Xu, X.; Xue, Z.; Snyder, J. H.; Qi, X. Functional Analysis of a Rice Oxidosqualene Cyclase through Total Gene Synthesis. *Mol. Plant* **2013,** *6*, 1726-1729.
24. Haralampidis, K., Bryan, G., Qi, X., Papadopoulou, K., Bakht, S., Melton, R., Osbourn, A. A New Class of Oxidosqualene Cyclases Directs Synthesis of Antimicrobial Phytoprotectants in Monocots. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13431-13436.
25. Shibuya, M.; Zhang, H.; Endo, A.; Shishikura, K.; Kushiro, T.; Ebizuka, Y. Two Branches of the Lupeol Synthase Gene in the Molecular Evolution of Plant Oxidosqualene Cyclases. *Eur. J. Biochem.* 1999, 266, 302-307.
26. Basyuni, M.; Oku, H.; Tsujimoto, E.; Kinjo, K.; Baba, S.; Takara, K. Triterpene Synthases from the Okinawan mangrove Tribe, Rhizophoraceae. *FEBS J.* **2007,** *274*, 5028-5042.
27. Hayashi, H., Huang, P., Takada, S., Obinata, M., Inoue, K., Shibuya, M., & Ebizuka, Y. Differential Expression of Three Oxidosqualene Cyclase mRNAs in Glycyrrhiza glabra. *BIOL PHARM BULL* **2004**, *27*, 1086-1092.
28. Guhling, O.; Hobl, B.; Yeats, T.; Jetter, R. Cloning and Characterization of a Lupeol Synthase Involved in the Synthesis of Epicuticular Wax Crystals on Stem and Hypocotyl Surfaces of Ricinus communis. *Arch. Biochem. Biophys.* **2006,** *448*, 60-72.
29. R. Herrera, J. B.; Bartel, B.; Wilson, W. K.; Matsuda, S. P. T. Cloning and Characterization of the Arabidopsis thaliana Lupeol Synthase Gene. *Phytochemistry* **1998,** *49*, 1905-1911.
30. Kirby, J.; Romanini, D. W.; Paradise, E. M.; Keasling, J. D. Engineering Triterpene Production in Saccharomyces cerevisiae–β-Amyrin Synthase from Artemisia annua. *FEBS J.* **2008,** *275*, 1852-1859.
31. Cantalapiedra, C. P.; Hernández-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J.; Tamura, K. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **2021,** *38*, 5825-5829.
32. Saito, K.; Sawai, S. Triterpenoid Biosynthesis and Engineering in Plants. *Front. Plant Sci.* **2011,** *2*, 25.
33. Moses, T.; Pollier, J.; Thevelein, J. M.; Goossens, A. Bioengineering of Plant (Tri)Terpenoids: from Metabolic Engineering of Plants to Synthetic Biology in Vivo and in Vitro. *New Phytol.* **2013,** *200*, 27-43.
34. Pichersky, E.; Lewinsohn, E. Convergent Evolution in Plant Specialized Metabolism. *Annu. Rev. Plant Biol.* **2011,** *62*, 549-566.
35. Weng, J.-K.; Philippe, R. N.; Noel, J. P. The Rise of Chemodiversity in Plants. *Science* **2012,** *336*, 1667-1670.
36. Osbourn, A. Secondary Metabolic Gene Clusters: Evolutionary Toolkits for Chemical Innovation. *Trends Genet.* **2010,** *26*, 449-457.
37. Graham, I. A.; Besser, K.; Blumer, S.; Branigan, C. A.; Czechowski, T.; Elias, L.; Guterman, I.; Harvey, D.; Isaac, P. G.; Khan, A. M.; Larson, T. R.; Li, Y.; Pawson, T.; Penfield, T.; Rae, A. M.; Rathbone, D. A.; Reid, S.; Ross, J.; Smallwood, M. F.; Segura, V.; Townsend, T.; Vyas, D.; Winzer, T.; Bowles, D. The Genetic Map of Artemisia annua L. Identifies Loci Affecting Yield of the Antimalarial Drug Artemisinin. *Science* **2010,** *327*, 328-331.
38. Shen, Q.; Zhang, L.; Liao, Z.; Wang, S.; Yan, T.; Shi, P.; Liu, M.; Fu, X.; Pan, Q.; Wang, Y.; Lv, Z.; Lu, X.; Zhang, F.; Jiang, W.; Ma, Y.; Chen, M.; Hao, X.; Li, L.; Tang, Y.; Lv, G.; Zhou, Y.; Sun, X.; Brodelius, P. E.;

 doi:10.20944/preprints202506.0611.v1

Rose, J. K. C.; Tang, K. The Genome of Artemisia annua Provides Insight into the Evolution of Asteraceae Family and Artemisinin Biosynthesis. *Mol. Plant* **2018,** *11*, 776-788.

39. Kliebenstein, D. J. Plant Defense Compounds: Systems Approaches to Metabolic Analysis. *Annu. Rev. Phytopathol.* **2012,** *50*, 155-173.
40. Osbourn, A. Saponins and Plant Defence-a Soap Story. *Trends Plant Sci.* **1996,** *1*, 4-9.
41. Mithöfer, A.; Boland, W. Plant Defense Against Herbivores: Chemical Aspects. *Annu. Rev. Plant Biol.* **2012,** *63*, 431-450.
42. Hartmann, M. Plant Sterols and the Membrane Environment. *Trends Plant Sci.* **1998,** *3*, 170-175.
43. Richards, T. A.; Soanes, D. M.; Jones, M. D. M.; Vasieva, O.; Leonard, G.; Paszkiewicz, K.; Foster, P. G.; Hall, N.; Talbot, N. J. Horizontal Gene Transfer Facilitated the Evolution of Plant Parasitic Mechanisms in the Oomycetes. *Proc. Natl. Acad. Sci. U.S.A.* **2011,** *108*, 15258-15263.
44. Daum, G.; Lees, N. D.; Bard, M.; Dickson, R. J. Y. Biochemistry, Cell Biology and Molecular Biology of Lipids of Saccharomyces cerevisiae. *Yeast* **1998,** *14*, 1471-1510.
45. William Roy, S.; Gilbert, W. The Evolution of Spliceosomal Introns: Patterns, Puzzles and Progress. *Nat. Rev. Genet.* **2006,** *7*, 211-221.
46. Filichkin, S. A.; Cumbie, J. S.; Dharmawardhana, P.; Jaiswal, P.; Chang, J. H.; Palusa, S. G.; Reddy, A.; Megraw, M.; Mockler, T. Environmental Stresses Modulate Abundance and Timing of Alternatively Spliced Circadian Transcripts in Arabidopsis. *Mol. Plant* **2015,** *8*, 207-227.
47. Hua, Z.; Vierstra, R. D. The Cullin-RING Ubiquitin-Protein Ligases. *Annu. Rev. Plant Biol.* **2011,** *62*, 299-334.
48. Tholl, D.; Lee, S. Terpene Specialized Metabolism in Arabidopsis thaliana. *The Arabidopsis Book/American Society of Plant Biologists* **2011,** 9, e0143.
49. Santner, A.; Estelle, M. The Ubiquitin-Proteasome System Regulates Plant Hormone Signaling. *Plant J.* **2010,** *61*, 1029-1040.
50. Tokuriki, N.; Tawfik, D. S. Protein Dynamism and Evolvability. *Science* **2009,** *324*, 203-207.
51. Xu, R.; Fazio, G. C.; Matsuda, S. P. T. On the Origins of Triterpenoid Skeletal Diversity. *Phytochemistry* **2004,** *65*, 261-291.
52. Kushiro, T.; Shibuya, M.; Masuda, K.; Ebizuka, Y. Mutational Studies on Triterpene Synthases: Engineering Lupeol Synthase into β-Amyrin Synthase. *J. Am. Chem. Soc.* **2000,** *122*, 6816-6824.
53. Miettinen, K.; Pollier, J.; Buyst, D.; Arendt, P.; Csuk, R.; Sommerwerk, S.; Moses, T.; Mertens, J.; Sonawane, P. D.; Pauwels, L.; Aharoni, A.; Martins, J.; Nelson, D. R.; Goossens, A. The Ancient CYP716 Family is a Major Contributor to the Diversification of Eudicot Triterpenoid Biosynthesis. *Nat. Commun.* **2017,** 8, 14153.
54. Seki, H.; Ohyama, K.; Sawai, S.; Mizutani, M.; Ohnishi, T.; Sudo, H.; Akashi, T.; Aoki, T.; Saito, K.; Muranaka, T. Licorice β-Amyrin 11-oxidase, a Cytochrome P450 with a Key Role in the Biosynthesis of the Triterpene Sweetener Glycyrrhizin. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105*, 14204-14209.
55. De Geyter, N.; Gholami, A.; Goormachtig, S.; Goossens, A. Transcriptional Machineries in Jasmonate-elicited Plant Secondary Metabolism. *Trends Plant Sci.* **2012,** *17*, 349-359.
56. Clark, S. E.; Running, M. P.; Meyerowitz, E. M. CLAVATA3 is a Specific Regulator of Shoot and Floral Meristem Development Affecting the Same Processes as CLAVATA1. *Development* **1995,** *121*, 2057-2067.
57. Jenkins, G. I., Photomorphogenic responses to ultraviolet-B light. *Plant Cell Environ.* **2017,** 40, 2544-2557.
58. Carbonell-Bejerano, P.; Diago, M.-P.; Martínez-Abaigar, J.; Martínez-Zapater, J. M.; Tardáguila, J.; Núñez-Olivera, E. Solar Ultraviolet Radiation is Necessary to Enhance Grapevine Fruit Ripening Transcriptional and Phenolic Responses. *BMC Plant Biol.* **2014,** *14*, 1-16.

 doi:10.20944/preprints202506.0611.v1

59. 69. Babiychuk, E.; Bouvier-Navé, P.; Compagnon, V.; Suzuki, M.; Muranaka, T.; Van Montagu, M.; Kushnir, S.; Schaller, H. Allelic Mutant Series Reveal Distinct Functions for Arabidopsis Cycloartenol Synthase 1 in Cell Viability and Plastid Biogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105*, 3163-3168.
60. Venkateshwaran, M.; Jayaraman, D.; Chabaud, M.; Genre, A.; Balloon, A. J.; Maeda, J.; Forshey, K.; den Os, D.; Kwiecien, N. W.; Coon, J. J.; Barker, D. G.; Ané, J.-M. A Role for the Mevalonate Pathway in Early Plant Symbiotic Signaling. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112*, 9781-9786.
61. Suzuki, M.; Kamide, Y.; Nagata, N.; Seki, H.; Ohyama, K.; Kato, H.; Masuda, K.; Sato, S.; Kato, T.; Tabata, S. J. T. P. J. Loss of Function of 3-hydroxy-3-methylglutaryl Coenzyme A Reductase 1 (HMG1) in Arabidopsis Leads to Dwarfing, Early Senescence and Male Sterility, and Reduced Sterol Levels. *Plant J.* **2004,** *37*, 750-761.
62. Szakiel, A.; Pączkowski, C.; Huttunen, S. Triterpenoid Content of Berries and Leaves of Bilberry Vaccinium myrtillus from Finland and Poland. *J. Agric. Food. Chem.* **2012,** *60*, 11839-11849.
63. Hayashi, H.; Huang, P.; Inoue, K. Up-regulation of Soyasaponin Biosynthesis by Methyl Jasmonate in Cultured Cells of Glycyrrhiza glabra. *Plant Cell Physiol.* **2003,** *44*, 404-411.
64. Huang, A. C.; Kautsar, S. A.; Hong, Y. J.; Medema, M. H.; Bond, A. D.; Tantillo, D. J.; Osbourn, A. Unearthing a Sesterterpene Biosynthetic Repertoire in the Brassicaceae Through Genome Mining Reveals Convergent Evolution. *Proc. Natl. Acad. Sci. U.S.A.* **2017,** 114, E6005-E6014.
65. Osbourn, A. Gene Clusters for Secondary Metabolic Pathways: An Emerging Theme in Plant Biology. *Plant Physiol.* **2010,** *154*, 531-535.
66. Ro, D.-K.; Paradise, E. M.; Ouellet, M.; Fisher, K. J.; Newman, K. L.; Ndungu, J. M.; Ho, K. A.; Eachus, R. A.; Ham, T. S.; Kirby, J.; Chang, M. C. Y.; Withers, S. T.; Shiba, Y.; Sarpong, R.; Keasling, J. D. Production of the Antimalarial Drug Precursor Artemisinic Acid in Engineered Yeast. *Nature* **2006,** *440*, 940-943.
67. Kirby, J.; Keasling, J. D. Biosynthesis of Plant Isoprenoids: Perspectives for Microbial Engineering. *Annu. Rev. Plant Biol.* **2009,** *60*, 335-355.
68. Sainsbury, F.; Thuenemann, E. C.; Lomonossoff, G. P. pEAQ: Versatile Expression Vectors for Easy and Quick Transient Expression of Heterologous Proteins in Plants. *Plant Biotechnol. J.* **2009,** *7*, 682-693.
69. Reed, J.; Osbourn, A. Engineering Terpenoid Production Through Transient Expression in Nicotiana benthamiana. *Plant Cell Rep.* **2018,** *37*, 1431-1441.
70. Goodin, M. M.; Zaitlin, D.; Naidu, R. A.; Lommel, S. A. Nicotiana benthamiana: Its History and Future as a Model for Plant–pathogen Interactions. *Mol. Plant-Microbe Interact.* **2008,** *21*, 1015-1026.
71. Pavesi, G.; Mereghetti, P.; Mauri, G.; Pesole, G. Weeder Web: Discovery of Transcription Factor Binding Sites in a Set of Sequences from Co-regulated Genes. *Nucleic Acids Res.* **2004,** *32*, W199-W203.
72. Johnson, D. S.; Mortazavi, A.; Myers, R. M.; Wold, B. Genome-Wide Mapping of in Vivo Protein-DNA Interactions. *Science* **2007,** *316*, 1497-1502.
73. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012,** 337, 816-821.
74. Bohlmann, J.; Meyer-Gauen, G.; Croteau, R. Plant Terpenoid Synthases: Molecular Biology and Phylogenetic Analysis. *Proc. Natl. Acad. Sci. U.S.A.* **1998,** *95*, 4126-4133.
75. Ober, D. Gene Duplications and the Time Thereafter - Examples from Plant Secondary Metabolism. *Plant Biol.* **2010,** *12*, 570-577.
76. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.;

 doi:10.20944/preprints202506.0611.v1

Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021,** 596, 583-589.

77. Karunanithi, P. S.; Zerbe, P. Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity. *Front. Plant Sci.* **2019,** 10, 1166.