



Article

Cloning and Functional Characterization of Two Germacrene A Oxidases Isolated from *Xanthium sibiricum*

Dong-Mei Xie ^{1,2}, Qiang Zhang ², Ling-Kai Xin ², Guo-Kai Wang ², Cong-Bin Liu ^{2,*} and Min-Jian Qin ^{1,*}

¹ Department of Resources Science of Traditional Chinese Medicines, China Pharmaceutical University, Nanjing 210009, China; xiedongmei@ahtcm.edu.cn

² Anhui Province Key Laboratory of Research & Development of Chinese Medicine, Anhui University of Chinese Medicine, Hefei 230012, China; zhangqiang@stu.ahtcm.edu.cn (Q.Z.); xinglingkai@stu.ahtcm.edu.cn (L.-K.X.); wanggk@ahtcm.edu.cn (G.-K.W.)

* Correspondence: liucongbin@ahtcm.edu.cn (C.-B.L.); qmj@cpcu.edu.cn (M.-J.Q.)

Abstract: Sesquiterpene lactones (STLs) from the cocklebur *Xanthium sibiricum* exhibit significant anti-tumor activity. Although germacrene A oxidase (GAO), which catalyzes the production of Germacrene A acid (GAA) from germacrene A, an important precursor of germacrene-type STLs, has been reported, the remaining GAOs corresponding to various STLs' biosynthesis pathways remain unidentified. In this study, 68,199 unigenes were studied in a de novo transcriptome assembly of *X. sibiricum* fruits. By comparison with previously published GAO sequences, two candidate *X. sibiricum* GAO gene sequences, XsGAO1 (1467 bp) and XsGAO2 (1527 bp), were identified, cloned, and predicted to encode 488 and 508 amino acids, respectively. Their protein structure, motifs, sequence similarity, and phylogenetic position were similar to those of other GAO proteins. They were most strongly expressed in fruits, according to a quantitative real-time polymerase chain reaction (qRT-PCR), and both XsGAO proteins were localized in the mitochondria of tobacco leaf epidermal cells. The two XsGAO genes were cloned into the expression vector for eukaryotic expression in *Saccharomyces cerevisiae*, and the enzyme reaction products were detected by gas chromatography–mass spectrometry (GC-MS) and liquid chromatography–mass spectrometry (LC-MS) methods. The results indicated that both XsGAO1 and XsGAO2 catalyzed the two-step conversion of germacrene A (GA) to GAA, meaning they are unlike classical GAO enzymes, which catalyze a three-step conversion of GA to GAA. This cloning and functional study of two GAO genes from *X. sibiricum* provides a useful basis for further elucidation of the STL biosynthesis pathway in *X. sibiricum*.

Keywords: *Xanthium sibiricum*; sesquiterpene lactones; biosynthesis; functional analysis; germacrene A



Citation: Xie, D.-M.; Zhang, Q.; Xin, L.-K.; Wang, G.-K.; Liu, C.-B.; Qin, M.-J. Cloning and Functional Characterization of Two Germacrene A Oxidases Isolated from *Xanthium sibiricum*. *Molecules* **2022**, *27*, 3322. <https://doi.org/10.3390/molecules27103322>

Academic Editors: Weishuo Fang and Yu Zhang

Received: 24 March 2022

Accepted: 15 May 2022

Published: 22 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sesquiterpene lactones (STLs) are widely distributed in nature and have a broad range of beneficial biological activities, including anti-bacterial, anti-inflammatory, and anti-cancer effects [1–5]. Two specific STLs, xanthatin and xanthinosin, are produced in the burs and leaves of *Xanthium L.* plants [6–8]. Many studies have been conducted on the quality and pharmacological activities of *X. sibiricum*. However, the details of the biological pathways associated with the anti-cancer effects of STLs in *Xanthium* species remain unclear.

Based on the carbon skeleton, STLs can be classified into multiple types, including germacrene, guaiane, xanthane, pseudoguaiane, eudesmane, and elemene lactones [9]. The molecular mechanisms of STLs differ among types. For example, eudesmane-type STLs synthesize the core carbon skeleton 10-epi-junenol before lactone ring synthesis [10], while guaiane-type STLs are produced from germacrene-type STLs and are induced by protonation [11]. The STL synthesis pathway is usually divided into three main processes: The first process is the synthesis of intermediates, including isopentenyl pyrophosphate (IPP)

and dimethylallyl pyrophosphate (DMAPP). The second step involves the formation of the sesquiterpene skeleton, which is preceded by the formation of farnesyl pyrophosphate (FPP) from IPP and DMAPP. Sesquiterpene synthase (STP) then catalyzes FPP to produce the sesquiterpene skeleton. The final step is the formation of the STL end-product, which involves a variety of structural modifications to the carbon skeleton. In particular, STP is critical for the structural transformation of FPP to STL, which subsequently catalyzes the formation of multiple types of sesquiterpenes through a series of chemical processes such as intermediate cyclization of carbenium ion, deprotonation, and hydrogen transfer [12–17]. Importantly, some STPs, the cytochrome P450 enzymes, play a modulatory role in STL biosynthesis, participating in the addition of functional groups to the sesquiterpene backbone [18]. For example, the cytochrome P450 GAO, isolated from an Asteraceae plant, catalyzes the three-step sequential oxidation of germacrene A to GAA [19]. In addition, in common chicory (*Cichorium intybus* L.), the most critical modification enzyme in the biosynthetic pathway of the 6 α -type STL-myrcene lactone C12 is the cytochrome P450 enzyme [20]. It was also shown that the cytochrome P450 enzymes parthenolide synthase (PTS) and kauniolide synthase (KLS), cloned from the aster *Chrysanthemum paludosum*, catalyzed the oxidation reaction of the C4–5 double-bonds of costunolide, which in turn, generated parthenolide [11,21]. Since germacrene-derived STLs are the simplest, a number of studies have investigated these STLs using synthetic biology and related techniques. Three STP genes were cloned from *X. stramonium*: XSTPSS1 catalyzed the production of germacrene D, XSTPSS2 catalyzed the formation of Guaia-4,6-diene, and XSTPSS3 catalyzed the production of germacrene A (Figure 1) [22]. However, it is not clear how different biologically active STLs are produced in *Xanthium* species after sesquiterpene skeleton formation.

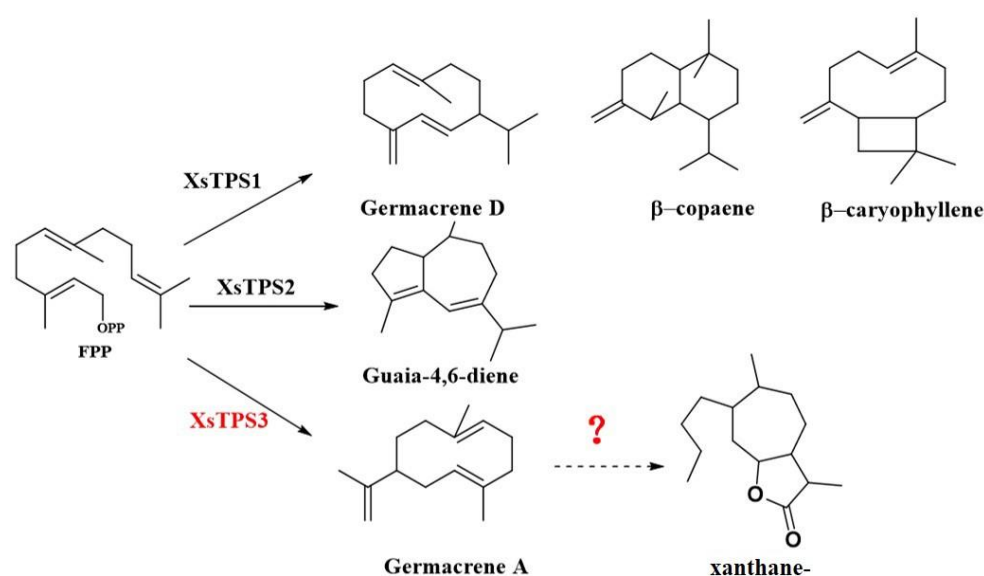


Figure 1. Overview of STL synthesis in *X. stramonium*. Solid lines are actual biosynthetic steps, and the dashed line indicates the hypothesized synthetic step.

Studies have shown that GAA is an important precursor substance in the biosynthesis pathway of germacrene-derived STLs [19]. Based on the findings outlined above, we hypothesized that, in *X. sibiricum*, GAO would catalyze the production of GAA from germacrene A to produce GAA in two consecutive steps (Figure 2).

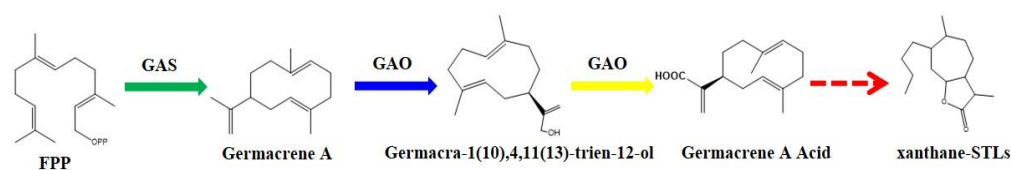


Figure 2. Hypothesized STL biosynthetic pathway in *X. sibiricum*.

To test this hypothesis, we identified the genes homologous to GAO by searching known GAO gene sequences in the National Center for Biotechnology Information (NCBI) database against a transcriptome library of *X. sibiricum* established therein and cloned the GAO gene using complementary DNA (cDNA). Subsequently, bioinformatic analyses of the predicted amino acid and protein structures, gene expression patterns, and subcellular localization were carried out. The enzyme reaction products were detected by GC-MS and LC-MS methods, where GAO genes were cloned into the expression vector for eukaryotic expression in *S. cerevisiae*. The results clarify the downstream STL synthesis pathway in *X. sibiricum* in future work.

2. Results

2.1. Establishment of a Transcriptome Library and Gene Annotation of *X. sibiricum*

The de novo transcriptome library of *X. sibiricum* included 5,989,562,311 nucleotides (nt), and the transcriptome Q20, N, and GC percentages were 97.57%, 0.01%, and 45.36%, respectively. After low-quality reads and filtering out those containing duplicates or junctions, 49,957,916 valid clean reads remained. The clean reads were assembled de novo using Trinity assembly software [23], and a total of 68,199 unigenes were obtained, with an average length of 639 nt and an N50 of 954 nt.

Protein function annotation information for all unigenes was obtained using BLAST [24]. Of the 68,199 unigenes, 19,129, 13,845, 24,721, 37,205, and 14,001 unigenes were successfully annotated with the Pfam, Kyoto Encyclopedia of Genes and Genome (KEGG), SwissProt, non-redundant protein sequence database (NR), and string libraries, respectively. A total of 13,619 unigenes from the transcriptome were successfully annotated to the Cluster of Orthologous Groups of proteins (COG) database, corresponding to the 25 functional categories. A total of 990 unigenes mainly focused on function prediction, with the highest percentage focusing on STLs, and a further 250 unigenes were annotated to secondary metabolite biosynthesis, transport, and catabolism (Appendix A, Figure A1).

A total of 13,845 unigenes were annotated to 128 metabolic pathways in the KEGG database, and 342 were annotated to “metabolism of terpenoids and polyketides” (Appendix A, Figure A2). Among these 342 unigenes, 89 were related to “terpenoid backbone biosynthesis”, and 22 were involved in “sesquiterpenoid and triterpenoid biosynthesis.”

In plants, the biosynthesis of STLs primarily occurs through the mevalonate (MVA) or methyl-D-erythritol phosphate (MEP) pathways, which synthesize DMAPP and IPP precursors [14]. KEGG pathway analysis showed that a total of 31 transcripts in *X. sibiricum* encoded six enzymes of the MVA pathway (acetyl-CoA C-acetyltransferase [ACCT, E.C.No:2.3.1.9], 3-hydroxy-3-methylglutaryl coenzyme A synthetase [HMGS, E.C.No:2.3.1.10], 3-hydroxy-3-methylglutaryl coenzyme A reductase [HMGR, E.C.No:1.1.3.34], mitogen-activated protein kinase/extracellular signal-regulated kinase [MEK, E.C.No:2.7.1.36], phosphomevalonate kinase [PMK, E.C.No:2.7.4.2], and pyrophosphomevalonate decarboxylase [MVD, E.C.No:4.1.1.33]), and 21 transcripts encoded seven enzymes in the MEP pathway (1-deoxy-D-xylulose-5-phosphate synthase [DXS, E.C.No:2.2.1.7], 1-deoxy-D-xylulose 5-phosphate reductoisomerase [DXR, E.C.No:1.1.1.26], 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase [MCT, E.C.No:2.7.7.60], 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase [CMK, E.C.No:2.7.7.148], 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase [MDS, E.C.No:4.6.1.12], 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase [HDS, E.C.No:1.17.7.1], and 4-hydroxy-3-methylbut-2-enyl-diphosphate reductase [HDR, E.C.No:1.17.1.2]).

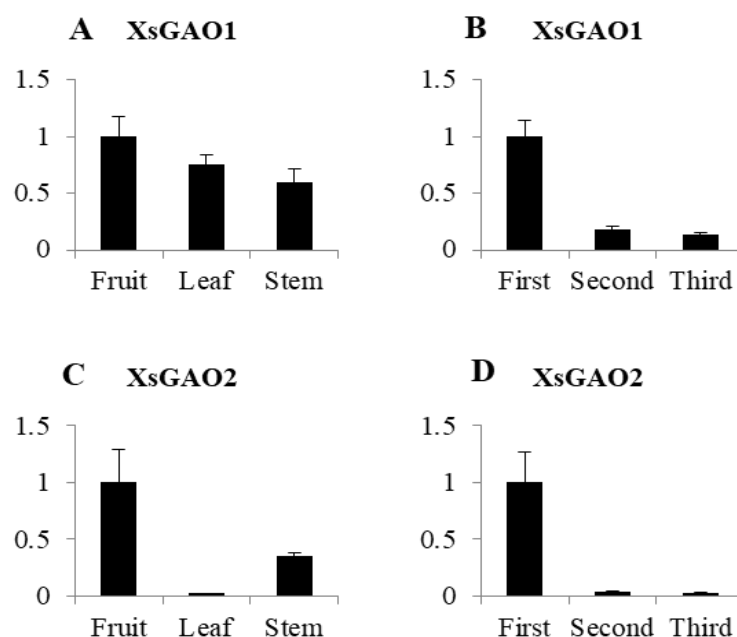


Figure 4. Expression of XsGAO1 and XsGAO2 in different organizations and periods ((A,C) are the expressions in different organs; (B,D) are the expressions at different times).

To localize XsGAO1 and XsGAO2 in the cell, tobacco leaf transformation was performed. Confocal laser scanning microscopy (CLSM) examination of the transformed tobacco leaves identified XsGAO1 and XsGAO2 signals in the mitochondria (Figure 5). This was consistent with predictions based on their sequence features.

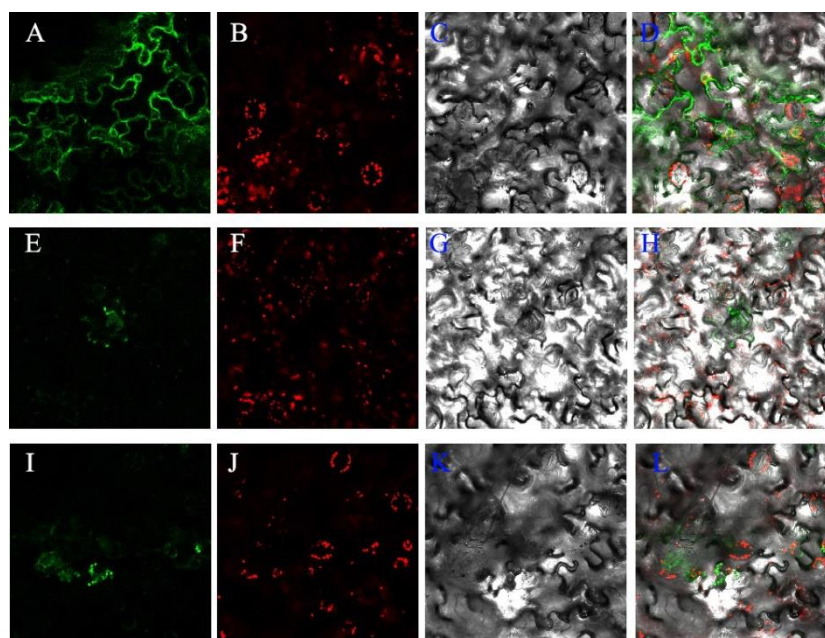


Figure 5. Subcellular localization of XsGAO1 and XsGAO2 fusion proteins in tobacco leaves. Images show tobacco leaves transformed with (A–D) the 1300-GFP empty plasmid, (E–H) the 35S: GAO1-GFP plasmid, and (I–L) the 35S: GAO2-GFP plasmid under various lights. Green signals correspond to the target gene fused with green fluorescent protein (GFP) after excitation at 488 nm; red signals correspond to the chloroplasts' autofluorescence after excitation at 488 nm.

2.4. Functional Study of XsGAO1 and XsGAO2

The full-length XsGAO genes were cloned into the yeast expression vector pYeDP60 and co-transferred into *S. cerevisiae* WAT11. Compared with the expression of LsGAS alone or the control yeast bearing the empty vector, germacrene A was obtained through prokaryotic expression and recombinant protein enzyme activity assays and was used as the substrate for XsGAO1 and XsGAO2 in the enzymatic activity reaction. Additionally, the LsGAO gene was cloned from *Lactuca sativa*, and the microsomal protein expressed in this gene was used as a positive control. The inactivated microsomes were used as a negative control to verify whether the microsomal protein expressed in XsGAO1 and XsGAO2 showed catalytic activity through a two-step enzyme activity catalytic assay (Appendix A, Figures A6–A9).

The gas chromatography–mass spectrometry (GC-MS) chromatogram of enzymatic activity experiments on the LsGAS recombinant protein showed two distinct peaks (Figure 6A). The second peak exhibited fragment ion peaks at m/z 53, 67, 79, 93, 107, 119, 133, 147, 161, 175, 189, and 204 (Figure 6B), which corresponded to the characteristic ions of germacrene A. However, the first peak exhibited fragment ion peaks at m/z 77, 81, 93, 107, 121, 133, 147, 161, and 189, which corresponded to the characteristic ions of β -elemene (Figure 6C) [10,25]. Therefore, we assumed that peak 1 corresponded to β -elemene and peak 2 corresponded to germacrene A.

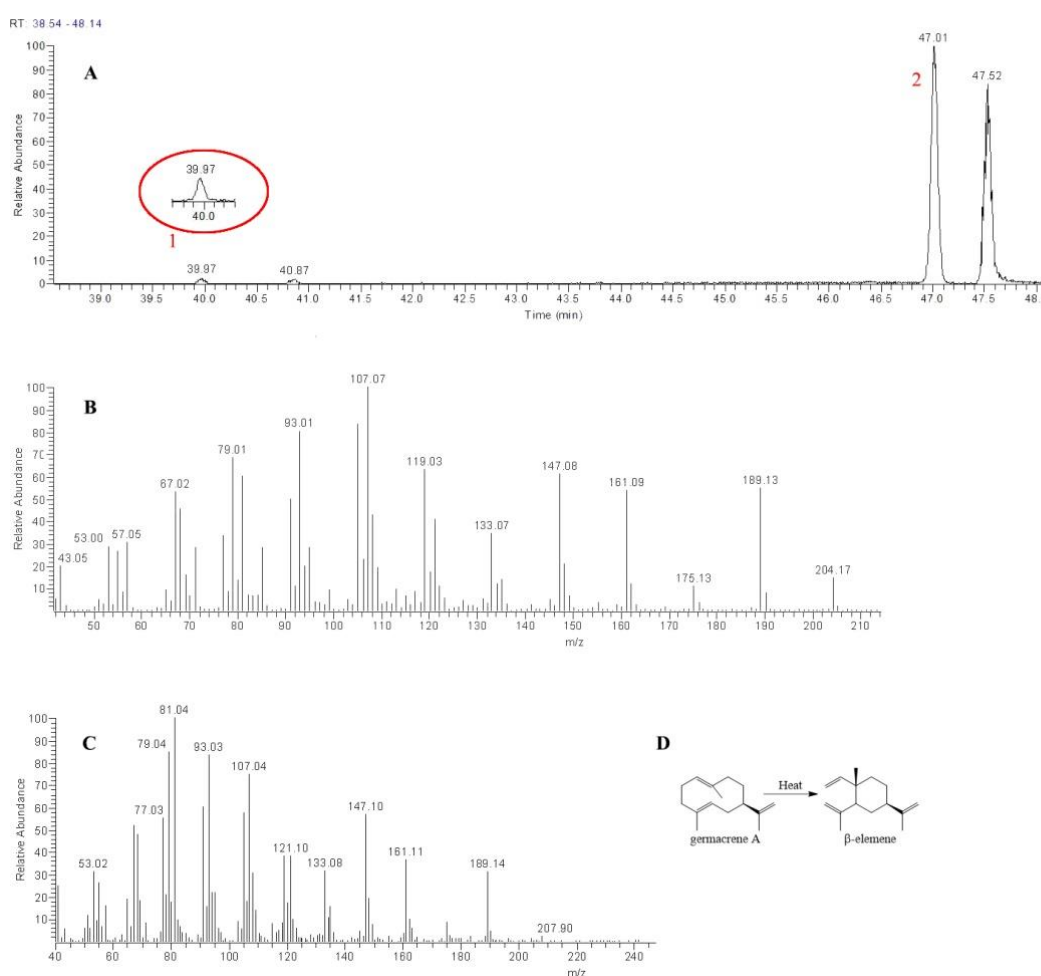


Figure 6. LsGAS gene catalyzed the production of GA from FPP. (A) A GC-MS ion chromatogram, in which peak 1 corresponds to β -elemene and peak 2 corresponds to GA. (B) Fragment ion of GA. (C) Fragment ion of β -elemene. (D) Rearrangement of germacrene A to β -elemene through the application of heat.

An analysis of the enzymatic activity of XsGAO1 and XsGAO2 in microsomes identified peaks in both liquid chromatography–mass spectrometry (LC-MS) chromatograms with the same retention time as the positive control LsGAO (tR = 12.54 min; Figure 7A). Further analysis showed that the peaks produced in all three assays (i.e., XsGAO1, XsGAO2, and LsGAO) had the same m/z in the positive ion mode ($[M+H]^+ = 235.17$; Figure 7B), suggesting that this peak corresponded to GAA. The results indicated that the XsGAO1 and XsGAO2 proteins both catalyzed the production of GAA from germacrene A in yeast microsomes.

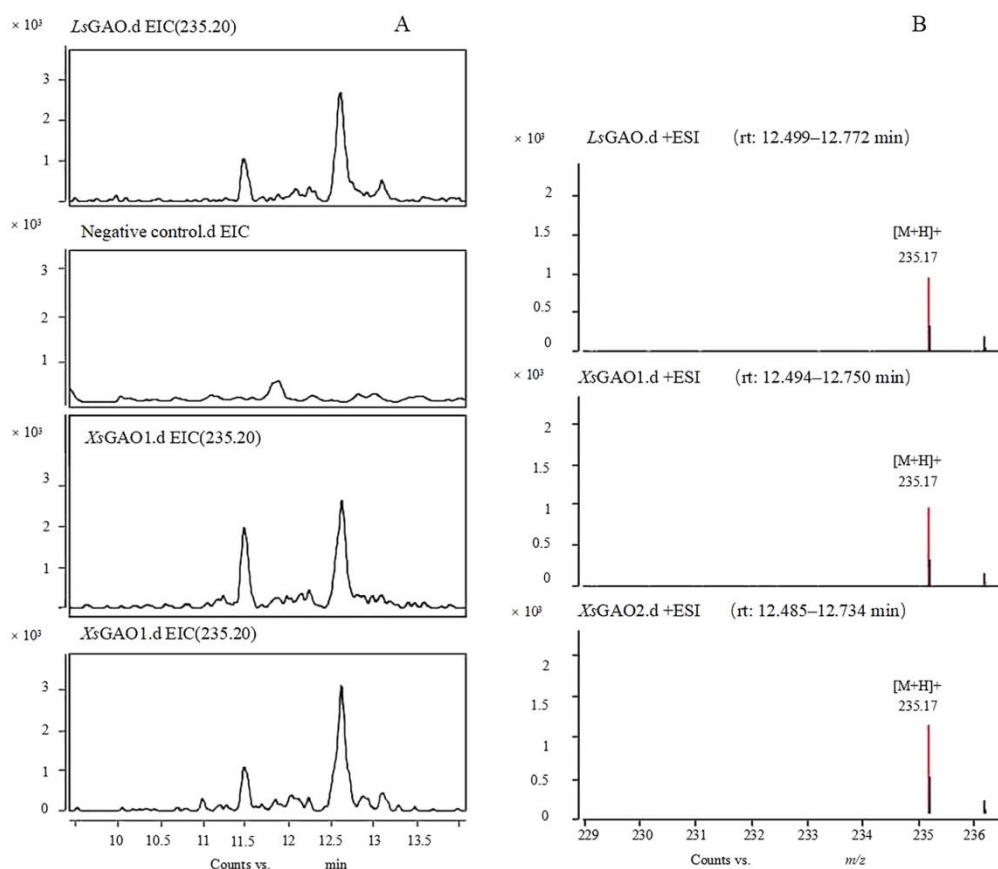


Figure 7. LsGAO, XsGAO1, and XsGAO2 catalyzed germacrene A. (A) Renal chromatogram and (B) mass spectrum.

3. Discussion

To investigate plant gene expression and analyze its function, transcriptome sequencing is an important molecular method that can provide genetic information in the absence of genomic data [26]. The de novo assembly platform greatly contributes to finding new genes, providing databases of sesquiterpene synthases and cytochrome P450s for cloning in *X. strumarium* [27]. *X. sibiricum* is a traditional plant containing unique secondary metabolites, of which STLs have various pharmaceutical properties. Although sesquiterpene synthase (STP) has only been cloned from *X. strumarium* glandular trichomes [28], transcriptome databases established from fruits of *X. sibiricum* provided cDNAs of two GAO genes that were cloned accurately in this study.

As an enzyme that produces an important precursor substance for the synthesis of STLs, the GAO gene is conserved in Asteraceae [25,29]. To investigate the function of this gene, the full-length cDNA sequences of XsGAO1 and XsGAO2 were successfully cloned. Phylogenetic analysis indicated that XsGAO1 may have a similar function to the GAO in *H. annuus*. However, XsGAO2 was distinct from other GAOs, representing a separate branch that needs further investigation. Multiple comparisons showed that the predicted

XsGAO1 protein had high homology with other redox-like proteins, and its protein sequence contained conserved amino acid residues that are expected in the cytochrome P450 enzyme family [30]. Analysis of the 3D structure predicted that XsGAO1 and XsGAO2 had functions similar to those corresponding with ferruginol synthase (CYP71 family).

The expression patterns of XsGAO1 and XsGAO2 in *X. sibiricum* leaves differed over time, with the highest expression level observed in young leaves, and expression levels decreasing with maturation. XsGAO1 and XsGAO2 were also differentially expressed among fruits and stems, presumably related to their functions. The expression of the two XsGAOs was the highest in fruits, which explains why fruits with higher contents of STLs are used in the traditional Chinese medicine Cang Er Zi.

XsGAO1 and XsGAO2 cDNAs with the correct sequence were successfully inserted into an expression vector and used in transient transformation assays of *Nicotiana benthamiana*. However, no fluorescence was observed in the protoplasts. This may be because the accumulated concentration of the product was below the detection limit of the instrument or because the GAA generated was intermediately transient in *N. benthamiana* [31]. Both XsGAO1 and XsGAO2 were localized in the mitochondria, which was consistent with terpene synthase in tomatoes (also localized in the mitochondria) [32,33].

GAO isolated from *L. sativa* was expressed in an engineered yeast to synthesize GAA de novo, and the classical GAO activity involved three-step oxidation of germacrene A (GA) to yield GAA and 12,6-guaianolides [25,34], similar to *N. benthamiana* [35]. Meanwhile, an XsGAO from *X. strumarium* catalyzed only one-step conversion of germacrene A to germacrene alcohol [36], but this study clearly shows that XsGAO1 and XsGAO2 catalyzed a second step of oxidation of the non-natural substrate germacrene A to germacrene A acid, which was not observed in yeast with a different GAO. Apparently, XsGAO2 is a unique enzyme, a functional adaptation of STLs' biosynthetic pathway diversification. Of course, a structural analysis of the XsGAO2 biochemical function, such as to identify the active center and crystal structure of oxidase, would help to examine whether it has unique GAO activity. In addition, with the advent of CRISPR-Cas genome editing, CRISPR-Cas-mediated gene knockout in tomatoes and the medicinal plant *Salvia miltiorrhiza* has been successfully performed [37,38]. This approach could be applied to verify the function of XsGAO2 in the future.

Beyond this, *X. sibiricum* contains a variety of biologically active STLs, mainly xanthane-STLs with anti-tumor activities [39,40]. As such, studying the genes in the STL synthesis pathway could provide new ideas for the investigation of xanthane-STL biosynthesis pathways.

4. Materials and Methods

4.1. Establishment of a Transcriptome Library and Gene Annotation

The fresh samples (fruit, leaf, and stem) of *X. sibiricum* used for the RNA extraction were collected from Chaoyang District (N: 40.0031, E: 116.5468, H: 113 m, Beijing, China) in August, wrapped in tinfoil, and frozen immediately in liquid nitrogen for storage at -80°C . Species verification was performed by Professor Dongmei Xie at the School of Pharmacy (Anhui University of Chinese Medicine). The total ribonucleic acid (RNA) was isolated using a TransZol Up Plus RNA kit (*TansGen* Biotech, Beijing, China), according to the manufacturer's protocol. The RNA extract was reverse-transcribed and then sequenced on an Illumina HiSeq 3000 platform at Shanghai Majorbio Bio-pharm Technology Corporation. After high-throughput sequencing, unigenes were assembled de novo from the clean reads obtained from the raw sequencing reads. To predict the biological function, all unigenes were annotated via a similarity search against the public databases, which contained Pfam, NR (<http://www.ncbi.nlm.nih.gov/>) (accessed on 5 May 2018), SwissProt (<http://www.uniprot.org/>) (accessed on 6 May 2018), KEGG (<http://www.genome.jp/kegg/>) (accessed on 8 May 2018), GO (Gene Ontology, <http://www.geneontology.org/>) (accessed on 10 May 2018), and BlastX (E value $< 1 \times 10^{-5}$) [24,26].

4.2. Cloning and Bioinformatics Analysis of XsGAO1 and XsGAO2

GAO gene sequences were searched for in the transcriptome database of *X. sibiricum* using local BLAST, with seed sequences downloaded from the NCBI. The full-length cDNA of two candidate genes, XsGAO1 and XsGAO2, was then cloned using reverse transcription polymerase chain reaction (RT-PCR) (primers are listed in Appendix A, Table A1).

The nucleotide sequences and their encoded amino acid sequences were analyzed using bioinformatics software, and the physicochemical properties of the encoding proteins were predicted using vector NTI, open reading frames (ORF), and amino acid sequence translation through ExPasy Translate (<http://web.expasy.org/translate/>) (accessed on 1 March 2020). Gene domain analysis was performed using SMART (<http://smart.embl-heidelberg.de/>) (accessed on 5 March 2020) [41].

Phylogenetic relationships were constructed using the amino acid sequences of XsGAO1 and XsGAO2 with different reported GAO sequences. Nineteen sequences were aligned using ClustalX2, and the alignment was used to construct a phylogenetic tree using MEGA5.0 software [42].

Secondary protein structures were obtained using SOMPA online analysis software (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) (accessed on 20 March 2020), and 3D structures and peptide-sequence fragments of XsGAO proteins were predicted using the SWISS-MODEL web server (<https://swissmodel.expasy.org/interactive>) (accessed on 20 March 2020) and PyMOL software [43].

4.3. Examination of the Expression Patterns of the XsGAO1 and XsGAO2 Genes

To further understand the distribution characteristics of GAO genes in *X. sibiricum*, qRT-PCR (primers are listed in Appendix A, Table A2) was employed to determine the expression patterns of two GAO genes in different organs (leaf, stem, and fruit) at different stages of *X. sibiricum*. RNA was extracted according to protocol provided for the TRIzol reagent (Invitrogen), and this was then converted into cDNA via reverse transcription with the Transcriptor First-Strand cDNA Synthesis Supermix kit. Subcellular localization of the protein was observed using CLSM, and the consistency with prediction was determined using SLP-Local (<https://sunflower.kuicr.kyoto-u.ac.jp/~smatsuda/slplocal.html>) (accessed on 25 March 2018) online [27,43].

4.4. Functional Study of XsGAO1 and XsGAO2 Genes in Yeast

The ORFs of XsGAO1 and XsGAO2 were PCR-amplified (primers are listed in Appendix A, Table A2), and the amplicons were digested with BamHI/EcoRI and cloned into the respective sites in pYeDP60-XsGAO (pYeDp60 plasmid provided by the Department of Pharmacology, Second Military Medical University). To supply the substrate for XsGAOs, the germacrene A synthase gene (*LsGAS*; AF489965) from *L. sativa* (provided by the Department of Pharmacognosy, Second Military Medical University) was inserted into the *E. coli* expression vector pet28a-*LsGAS* at EcoRI-SacI sites. Germacrene A was produced by *LsGAS*, which was expressed through Transetta (DE3), and then germacrene A was catalyzed by the XsGAO gene expressed in *S. cerevisiae* WAT11 (WAT11 provided by the China Academy of Chinese Medical Sciences). For comparison, we used a classical GAO from *L. sativa* (*LsGAO*; GU198171) that is known to oxidize germacrene A in a three-step oxidation process. Transgenic yeast cells were cultivated in appropriate dropout media, and the expression of the transferred genes was induced by 2% galactose [25,34,41].

4.5. GC-MS and LC-MS Analyses

GC-MS analysis was performed [25] using a Shimadzu GC-MS TRACE GC Ultra/DSQ II instrument (Thermo Fisher Scientific, Waltham, MA, USA). A sample volume of 1 μ L was injected at an inlet temperature of 150 $^{\circ}$ C, and compounds were separated using a DB-5MS column (30 m \times 250 μ m \times 0.1 μ m) with helium as a carrier gas at a flow rate of 2 mL/min. The GC oven temperature program was as follows: 45 $^{\circ}$ C for 4 min, 45–170 $^{\circ}$ C for 67 min, and 170 $^{\circ}$ C for 72 min. The electron impact ionization of the mass spectrometric detector

was tuned to 70 eV and operated at 40–400 Da in full scan mode. LC-MS/MS analysis was performed [41] using a UHPLC-Q-TOF-MS system (Agilent Technologies, Santa Clara, CA, USA) equipped with an XBridge™ C₁₈ column (2.1 × 100 mm; Waters Corporation, Milford, MA, USA), with a mobile phase of 0.1% formic acid aqueous solution (A) and 0.1% acetonitrile (B) at a flow rate of 0.4 mL/min, injection volume of 1 µL, and column temperature of 40 °C. Mass spectrometry data were collected by Electrospray ion sources in the positive mode, and the collection range was 100–1700 *m/z*.

Author Contributions: M.-J.Q., D.-M.X., and C.-B.L. conceived and designed the experiments; Q.Z. and G.-K.W. performed the experiments and analyzed the data; G.-K.W. and D.-M.X. contributed the reagents, materials, and analysis tools; L.-K.X. and D.-M.X. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by the National Natural Science Foundation of China, grant number 81503190, and the National Natural Science Foundation of Anhui University of Chinese Medicine, grant number 2021hxts22.

Data Availability Statement: The data presented in this study are available in the article.

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

Sample Availability: Samples of the compounds GA and GAA are available from the authors.

Appendix A

The appendix provides data supplemental to the main text.

Table A1. Primers used in GAO gene cloning.

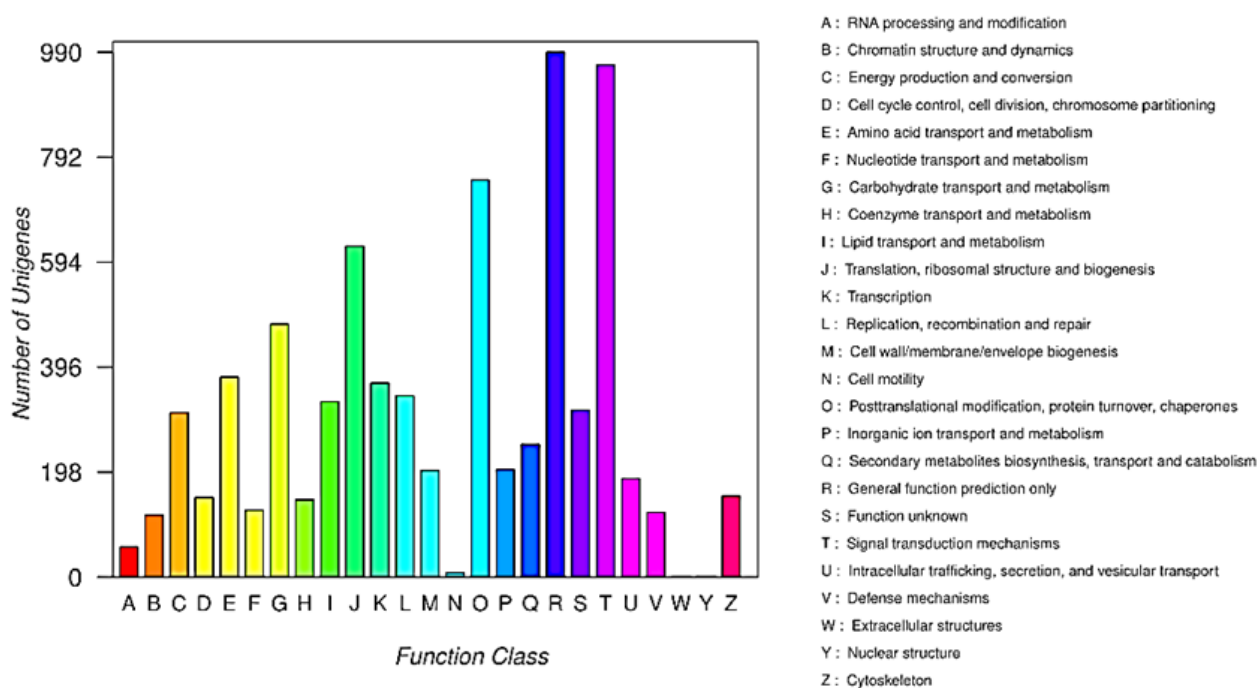
Primer	Primer Sequence
XsGAO1F	ATGGAAGTCTCCCTCACCCTC
XsGAO1R	TTAAAACTTGGTACCAATATCAACCC
XsGAO2F	ATGGAAGTCCATTTTCCCAC
XsGAO2R	TCCAAATATCACTATCCTTCG
M13F	CAGGAAACAGCTATGAC
M13R	GTAAAACGACGGCCAGT

Table A2. Primers used in RT-PCR, qRT-PCR, and expression in yeast.

Primer	Primer Sequence
QXsGAO1F	CTAATAAGGTGTCCGAGAG
QXsGAO1R	GGCAGGTCTGAATATCT
QXsGAO2F	TCTCAACCATAGTAATCTCA
QXsGAO2R	CGATGTCTGTGTAATTGTAT
ActinF	TACTACAACGGCAGAACGGGAAA
ActinR	TCATAGACGGCTGGAACAAAACC
GFP-XsGAO1F	acgggggactcttgaccatggATGGAAGTCTCCCTCACCCTC
GFP-XsGAO1R	gcccttgctcaccatactagtAAAACTTGGTACCAATATCAACCCA
GFP-XsGAO2F	acgggggactcttgaccatggATGGAAGTCCATTTTCCCACC
GFP-XsGAO2R	gcccttgctcaccatactagtCATTGTGTTGTAAGGTGTTGGGA
LsGASF	ATGGCAGCAGTTGACACTAATG
LsGASR	TTACATGGATACAGAACCAAC
LsGAOF	ATGGAGCTTTCAATAACCACC
LsGAOR	CTAAAACTCGGTACGAGTAACAAC
pYeDP60-XsGAO1F	acacactaaattaccgatccATGGAAGTCTCCCTCACCCTC
pYeDP60-XsGAO1R	gggagatccccgcggaattcTTAAAACTTGGTACCAATATCAACCC
pYeDP60-XsGAO2F	acacactaaattaccgatccATGGAAGTCCATTTTCCCACC
pYeDP60-XsGAO2R	gggagatccccgcggaattcTCACATGTGTTGTAAGGTGTTGG
pYeDP60-LsGAOF	acacactaaattaccgatccATGGAGCTTTCAATAACCACCTCC
pYeDP60-LsGAOR	gggagatccccgcggaattcCTAAAACTCGGTACGAGTAACAACCTC

Table A3. Reference sequences used for phylogeny construction.

Name	Species	Gene Bank Accession No.
CcVsGAO1		AIA09035.1
CcVsGAO2	<i>Cynara cardunculus</i> var.	XP_024977750.1
CcVsGAO3	scolymus	AIA09037.1
CcVs-GAOLike		XP_024977969.1
TpGAO	<i>Tanacetum parthenium</i>	AHN62855.1
LsGAO1		XP_023734551.1
LsGAO2	<i>Lactuca sativa</i>	ADF32078.1
LsGAO3		AIX97103.1
BsGAO	<i>Barnadesia spinosa</i>	ADF43083.1
HaGAO	<i>Helianthus annuus</i>	ADF43082.1
ScGAO	<i>Saussurea costus</i>	ADF43081.1
CiGAO	<i>Cichorium intybus</i>	ADF43080.1
TcGAO	<i>Tanacetum cinerariifolium</i>	AGO03789.1
CeGAO1		AZI95573.1
CeGAO2	<i>Cichorium endivia</i>	AZI95575.1
CeGAO3		AZI95574.1
OsJGGAO	<i>Oryza sativa</i> Japonica Group	XP_015624875.1

**Figure A1.** COG classification of *X. sibiricum* unigenes.

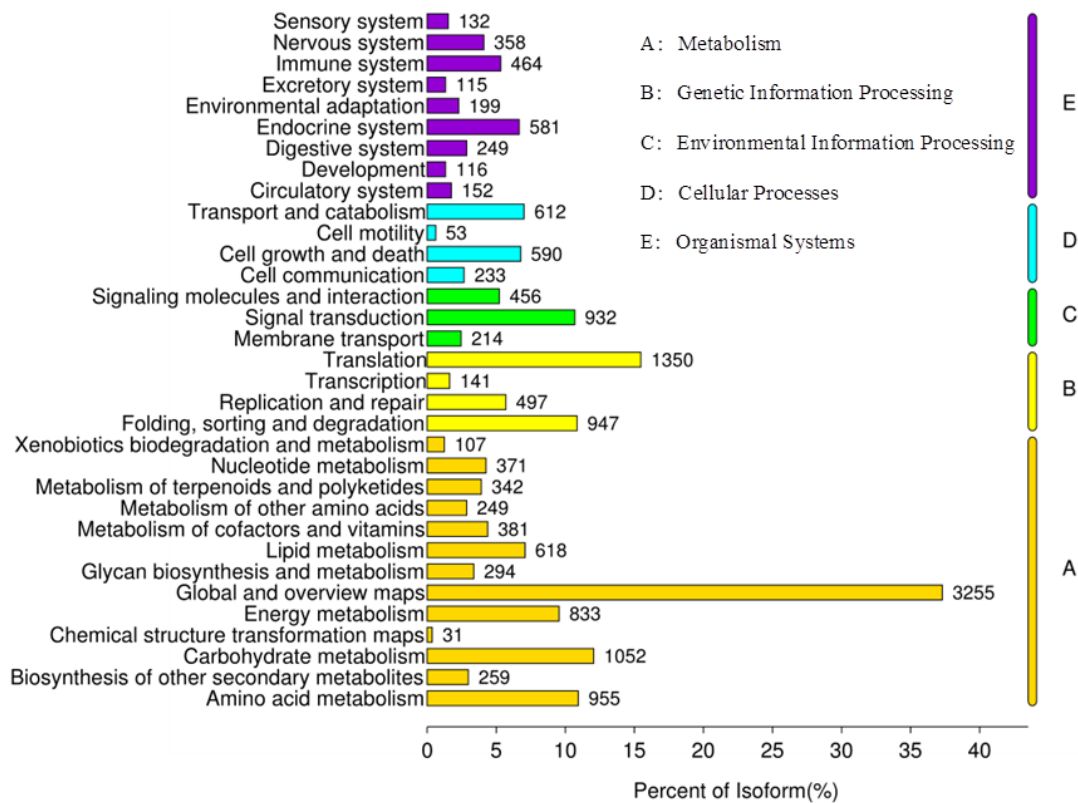


Figure A2. Pathway analysis of associated *X. sibiricum* unigenes in KEGG.

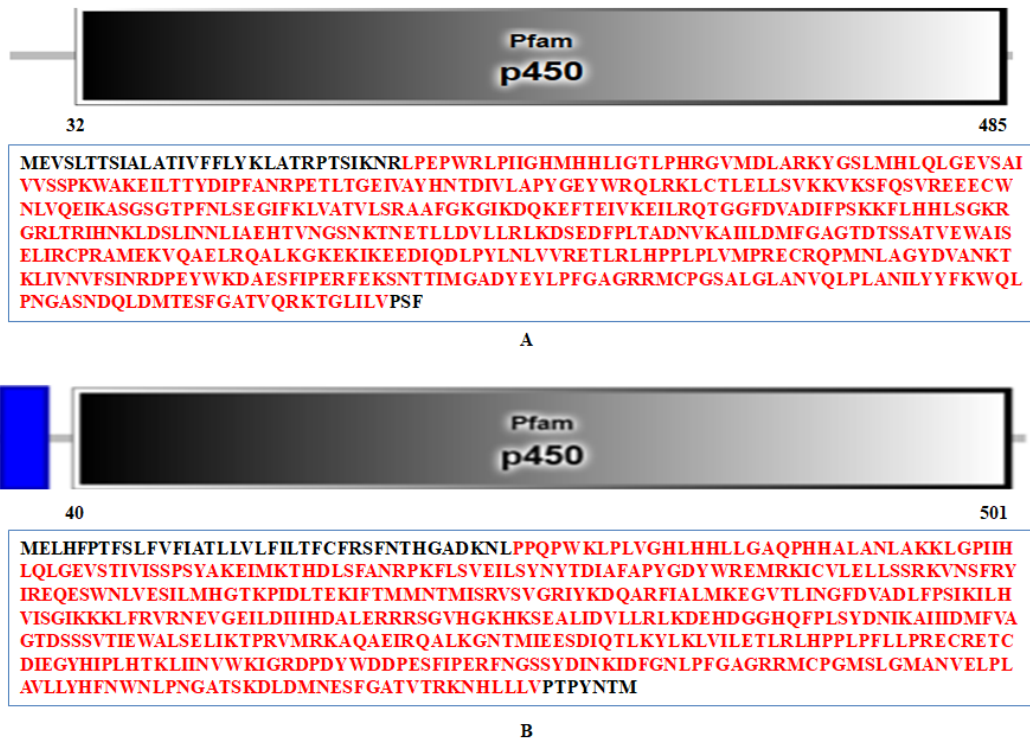


Figure A3. Domain prediction and amino acid sequences of XsGAO1 (A) and XsGAO2 (B). (Function domains shown in red).

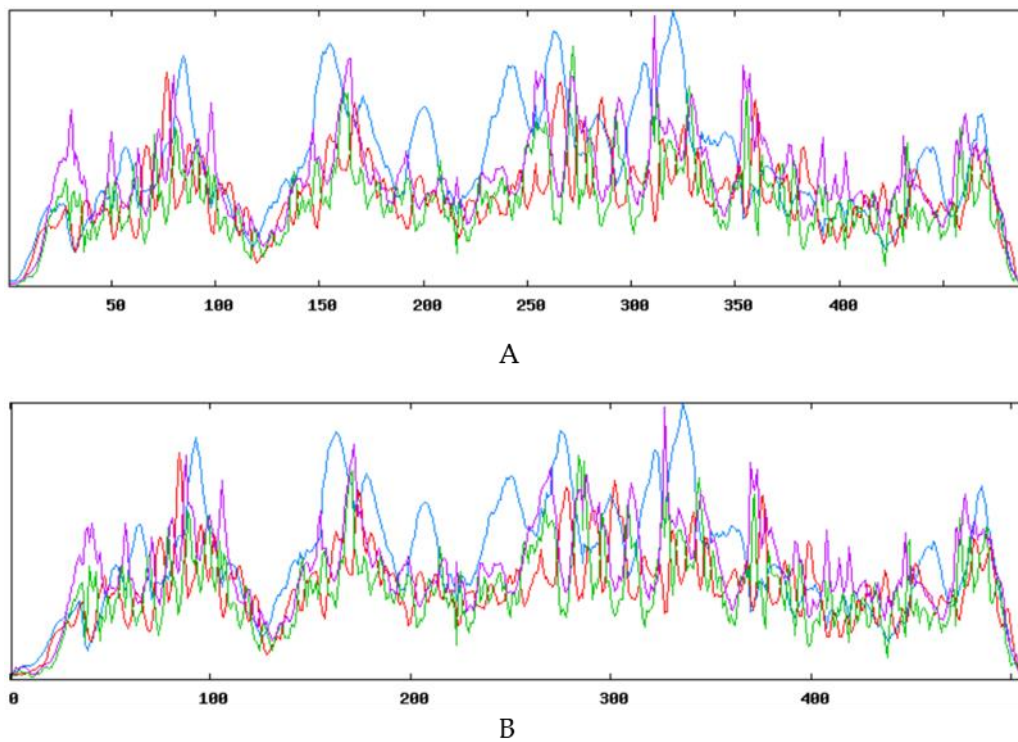


Figure A4. Secondary structures of (A) XsGAO1 and (B) XsGAO2. α -helixes are indicated in blue, β -sheets are indicated in green, and random coils are indicated in red.

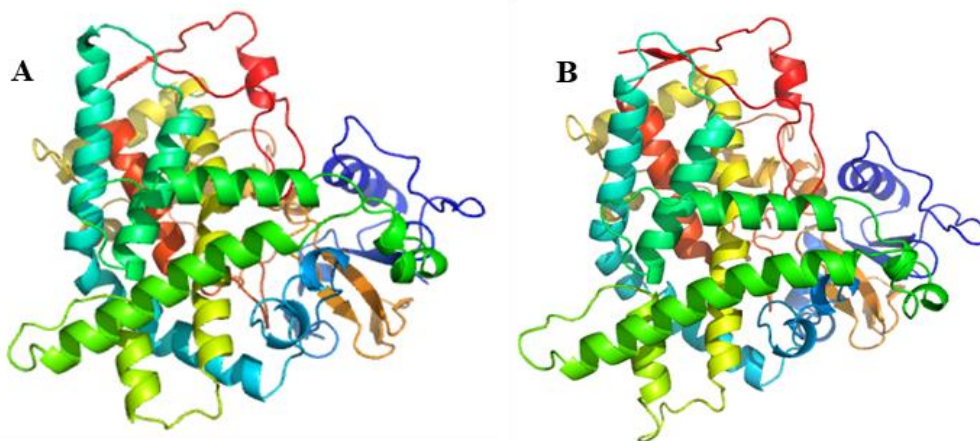


Figure A5. Predicted 3D structures of (A) XsGAO1 and (B) XsGAO2.

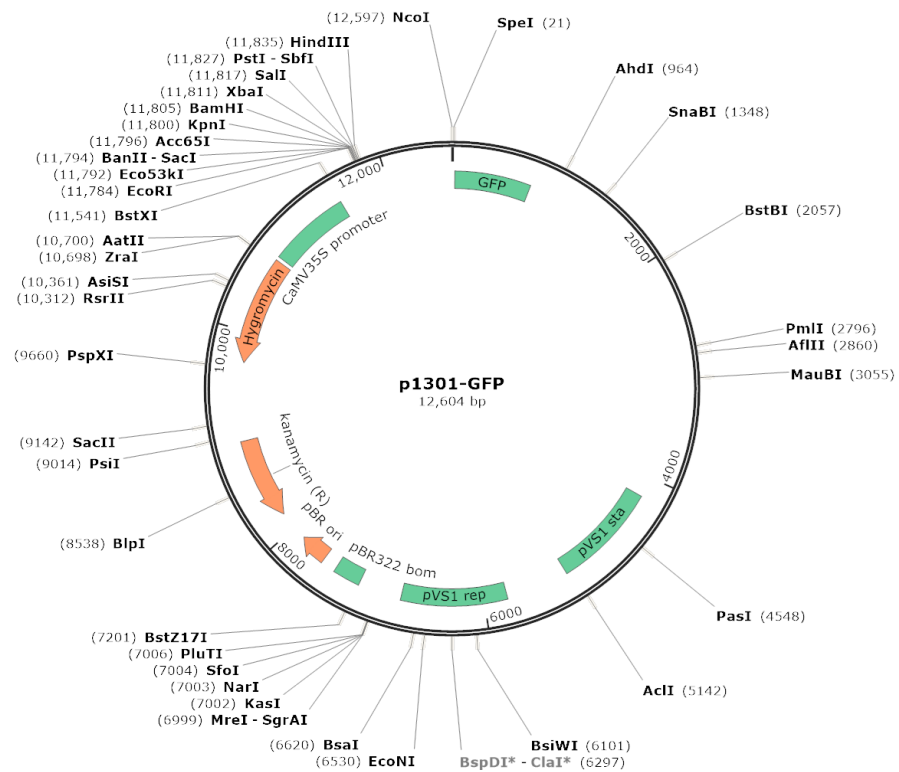


Figure A6. Map of the vector used for subcellular localization of XsGAO1 and XsGAO2 created with SnapGene. BspDI* and ClaI* are isoschizomers.

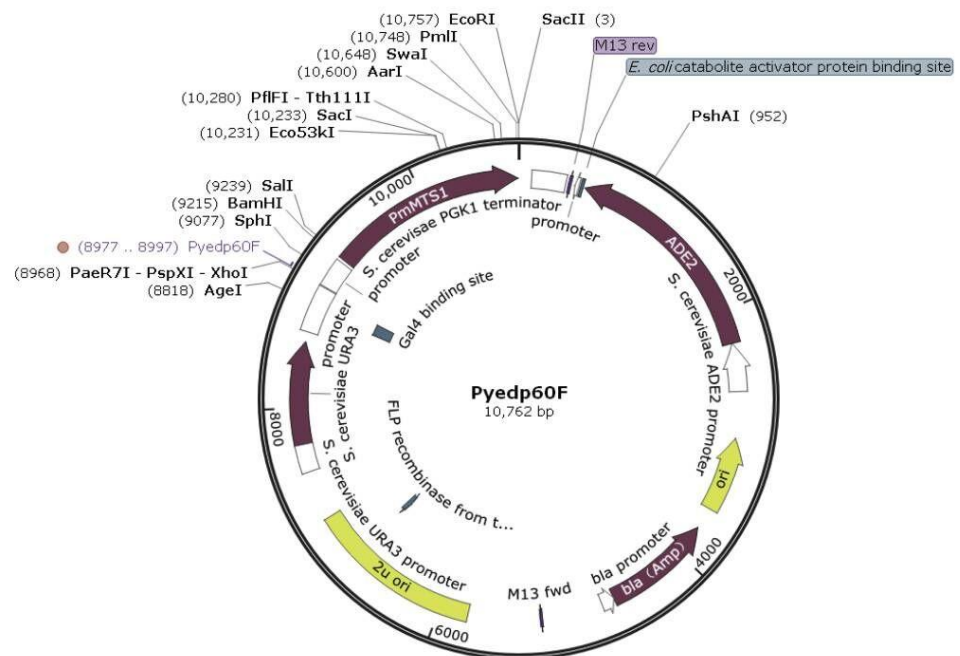


Figure A7. Map of the vector used in the yeast expression experiment created with SnapGene. Different protein showed with distinguished color.

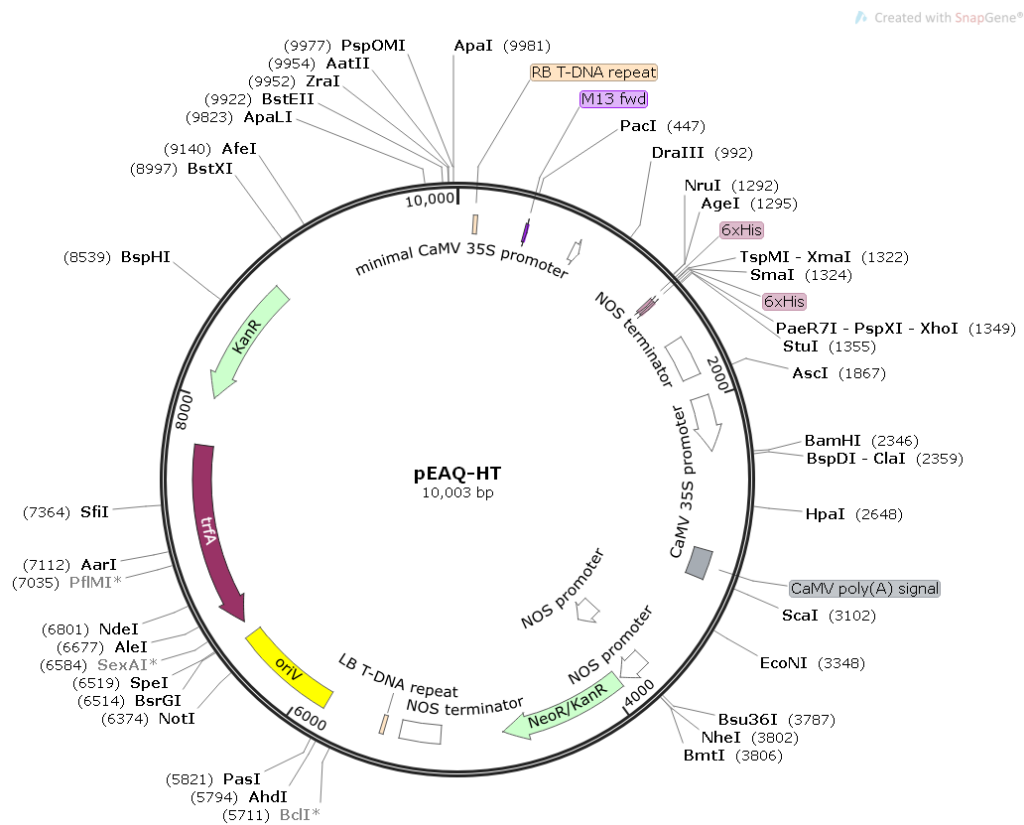


Figure A8. Map of the vector used for function validation of XsGAO1 and XsGAO2 in tobacco created with SnapGene. BclI* is methylation sensitive restriction endonuclease. Different protein showed with distinguished color.

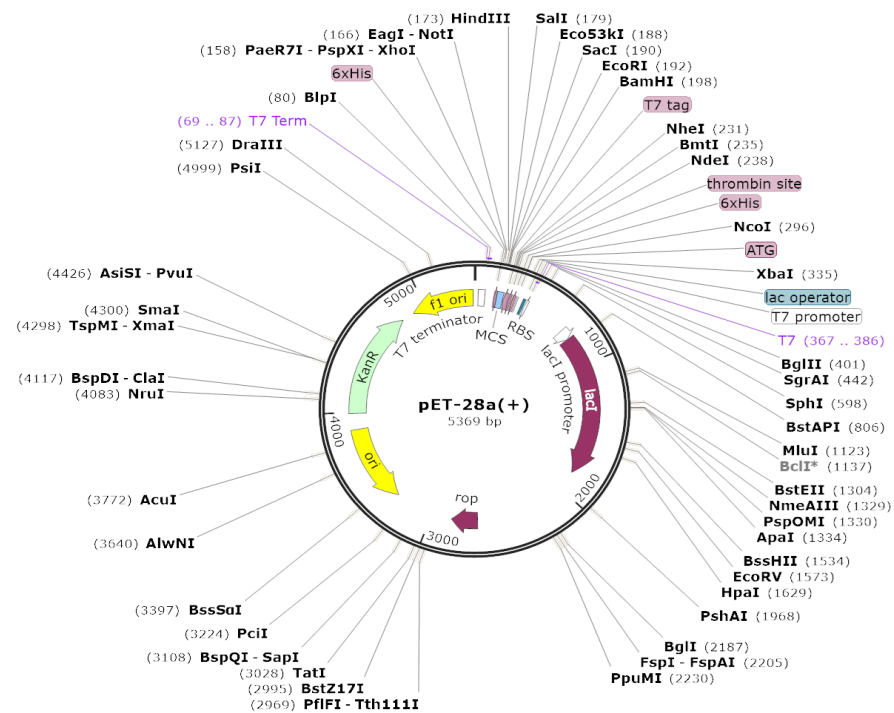


Figure A9. Map of the vector used for prokaryotic expression created with SnapGene. Different protein showed with distinguished color.

References

1. Guo, L.M.; Lv, J.L.; Zhang, L.B. Research progress on anti-inflammatory mechanism of natural sesquiterpenoids. *China J. Chin. Mater. Med.* **2018**, *43*, 3989–3999.
2. Xu, D.; Liu, D.; Zhang, C.H.; Zhang, R.F.; Wang, Q.L.; Zou, K.; Wang, H. Antitumor constituents from roots of *Ligularia veitchiana*. *J. Cent. China Norm. Univ. (Nat. Sci.)* **2021**, *55*, 72–75.
3. Yan, Z.P.; Xu, T.T.; An, Z.T.; Hu, Y.; Chen, W.Z.; Ma, J.X.; Shao, C.L.; Zhu, F.S. Costunolide induces mitochondria-mediated apoptosis in human gastric adenocarcinoma BGC-823 cells. *BMC Complementary Altern. Med.* **2019**, *19*, 151–162. [[CrossRef](#)] [[PubMed](#)]
4. Qiao, W.; Wang, X.; Li, X.; Li, J.; Li, X. Advances in broad-spectrum antitumor effect of four sesquiterpenoids. *Chin. J. Biochem. Pharm.* **2016**, *36*, 5–8.
5. Jiang, Y.Q.; Dong, Y.J.; Zhou, F.J.; Chen, J.P.; Zhou, Y.T.; Tian, C.W.; Chen, C.Q. Research progress on artemisinin and its derivatives. *Chin. Tradit. Herb. Drugs* **2022**, *53*, 599–609.
6. Ramírez-Erosa, I.; Huang, Y.G.; Hickie, R.A.; Sutherland, R.G.; Barl, B. Xanthatin and xanthinosin from the burs of *Xanthium strumarium* L. as potential anticancer agents. *Can. J. Physiol. Pharmacol.* **2007**, *85*, 1160–1172. [[CrossRef](#)]
7. Zhang, L.; Ruan, J.S.; Yan, L.G.; Li, W.D.; Wu, Y.; Tao, L.; Zhang, F.; Zheng, S.Z.; Wang, A.Y.; Lu, Y. Xanthatin Induces Cell Cycle Arrest at G2/M Checkpoint and Apoptosis via Disrupting NF- κ B Pathway in A549 Non-Small-Cell Lung Cancer Cells. *Molecules* **2012**, *17*, 3736–3750. [[CrossRef](#)]
8. Aranjan, J.M.; Manuel, A.; Rao, C.M.; Udupa, N.; Rao, J.V.; Joy, A.M.; Gandhi, P.; Radhakrishnan, E.K. Preliminary evaluation of in vitro cytotoxicity and in vivo antitumor activity of *Xanthium strumarium* in transplantable tumors in mice. *Am. J. Chin. Med.* **2013**, *41*, 145–162. [[CrossRef](#)]
9. Wang, L.J.; Fang, X.; Yang, C.Q.; Li, J.X.; Chen, X.Y. Biosynthesis and regulation of secondary terpenoid metabolism in plants. *Scientia Sinica Vitae* **2013**, *43*, 1030–1046.
10. Gou, J.B.; Hao, F.H.; Huang, C.Y.; Kwon, M.; Chen, F.F.; Li, C.F.; Liu, C.Y.; Ro, D.; Tang, H.R.; Zhang, Y.S. Discovery of a non-stereoselective cytochrome P450 catalyzing either 8 α - or 8 β -hydroxylation of germacrene A acid from the Chinese medicinal plant, *Inula hupehensis*. *Plant J.* **2018**, *93*, 92–106. [[CrossRef](#)]
11. Liu, Q.; Kashkooli, A.B.; Manzano, D.; Pateraki, I.; Richard, L.; Kolkman, P.; Lucas, L.F.; Guallar, V.; Vos, R.C.; Franssen, M.C.R.; et al. Kauniolide synthase is a P450 with unusual hydroxylation and cyclization-elimination activity. *Nat. Commun.* **2018**, *9*, 4657. [[CrossRef](#)] [[PubMed](#)]
12. Liao, P.; Hemmerlin, A.; Bach, T.J.; Chye, M.L. The potential of the mevalonate pathway for enhanced isoprenoid production. *Biotechnol. Adv.* **2016**, *34*, 697–713. [[CrossRef](#)] [[PubMed](#)]
13. Zong, Z.; Cheng, L.; Chen, Z.J.; Wang, L.; Wang, C.; Qi, Y.G.; Liu, A.J. Research progress on biosynthesis of terpenoids for food. *China Brew.* **2018**, *37*, 22–27.
14. Shi, M.; Wang, Y.; Zhou, W.; Hua, Q.; Kai, G.Y. Research progress in terms of the biosynthesis and regulation of terpenoids from medicinal plants. *Sci. Sin. Vitae* **2018**, *48*, 352–364.
15. Xiao, L.; Chen, R.B.; Wu, Y.; Zhang, L. Research progress on effect of AP2/ERF transcription factors in regulating secondary metabolite biosynthesis. *China J. Chin. Mater. Med.* **2020**, *45*, 5412–5420.
16. Little, D.B.; Croteau, R.B. Alteration of product formation by directed mutagenesis and truncation of the multiple-product sesquiterpene synthases δ -selinene synthase and γ -humulene synthase. *Arch. Biochem. Biophys.* **2002**, *402*, 120–135. [[CrossRef](#)]
17. Yoshikuni, Y.; Martin, V.J.; Ferrin, T.E.; Keasling, J.D. Engineering cotton (+)- δ -cadinene synthase to an altered function: Germacrene D-4-ol synthase. *Chem. Biol.* **2006**, *13*, 91–98. [[CrossRef](#)]
18. Frey, M.; Schmauder, K.; Pateraki, I.; Spring, O. Biosynthesis of Eupatolide-A metabolic route for sesquiterpene lactone formation involving the P450 enzyme CYP71DD6. *ACS Chem. Biol.* **2018**, *13*, 1536–1543. [[CrossRef](#)]
19. Wallaart, T.E.; Bouwmeester, H.J.; Hille, J.; Poppinga, L.; Maijers, N.C. Amorpha-4,11-diene synthase: Cloning and functional expression of a key enzyme in the biosynthetic pathway of the novel antimalarial drug artemisinin. *Planta* **2001**, *212*, 460–465. [[CrossRef](#)]
20. Kraker, J.M.; Franssen, M.C.; Joerink, M.; Groot, A.D.; Bouwmeester, H.J. Biosynthesis of costunolide, dihydrocostunolide, and leucodin. Demonstration of cytochrome P450-catalyzed formation of the lactone ring present in sesquiterpene lactones of chicory. *Plant Physiol.* **2002**, *129*, 257–268. [[CrossRef](#)]
21. Liu, Q.; Manzano, D.; Tanic, N.; Pesic, M.; Bankovic, J.; Pateraki, I.; Ricard, L.; Ferrer, A.; Vos, R.D.; Krol, S.V.D.; et al. Elucidation and in planta reconstitution of the parthenolide biosynthetic pathway. *Metab. Eng.* **2014**, *23*, 145–153. [[CrossRef](#)] [[PubMed](#)]
22. Li, Y.J.; Chen, F.F.; Li, Z.Q.; Zhang, Y.S. Identification and functional characterization of sesquiterpene synthases from *Xanthium strumarium*. *Plant Cell Physiol.* **2016**, *57*, 630–641. [[CrossRef](#)] [[PubMed](#)]
23. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Lin, F.; Raychowdhury, R.; Zeng, Q.D.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)] [[PubMed](#)]
24. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
25. Nguyen, D.T.; Gopfert, J.C.; Ikezawa, N.; MacNevin, G.; Kathiresan, M.; Gonrad, J.; Spring, O.; Ro, D.K. Biochemical conservation and evolution of germacrene a oxidase in Asteraceae. *J. Biol. Chem.* **2010**, *285*, 16588–16598. [[CrossRef](#)]

26. Hui, F.; Liu, X.Y.; Li, Z.Y.; Liu, F.S.; Yang, S.H. Application of transcriptome sequencing in study of medicinal plants. *Chin. Tradit. Herb. Drugs* **2019**, *50*, 6149–6155.
27. Li, Y.J. Cloning and Functional Analysis of Genes Involved in the Biosynthesis Sesquiterpene Lactones in *Xanthium strumarium* L. Ph.D. Thesis, Chinese Academy of Sciences, Beijing, China, 2017.
28. Chen, F.F.; Hao, F.H.; Li, C.F.; Gou, J.B.; Lu, D.Y.; Gong, F.J.; Tang, H.R.; Zhang, Y.S. Identifying three ecological chemotypes of *Xanthium strumarium* glandular trichomes using a combined NMR and LC-MS method. *PLoS ONE* **2013**, *8*, 10. [[CrossRef](#)]
29. Nguyen, T.D.; Kwon, M.; Kim, S.U.; Fischer, C.; Ro, D.K. Catalytic plasticity of germacrene A oxidase underlies sesquiterpene lactone diversification. *Plant Physiol.* **2019**, *181*, 945–960. [[CrossRef](#)]
30. Hamberger, B.; Bak, S. Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philos. Trans. R. Soc. B* **2013**, *368*, 20120426. [[CrossRef](#)]
31. Kraker, J.W.D.; Franssen, M.C.R.; Groot, A.D.; Shibata, T.; Bouwmeester, H.J. Germacrenes from fresh costus roots. *Phytochemistry* **2011**, *58*, 481–487. [[CrossRef](#)]
32. Falara, V.; Akhtar, T.A.; Nguyen, T.T.; Spyropoulou, E.A.; Bleeker, P.M.; Schauvinhold, I.; Matsuba, Y.; Bonini, M.E.; Schillmiller, A.L.; Last, A.L.; et al. The tomato terpene synthase gene family. *Plant Physiology*. **2011**, *157*, 770–789. [[CrossRef](#)] [[PubMed](#)]
33. Kappers, I.F.; Aharoni, A.; Herpen, T.W.; Luckerhoff, L.L.; Dicke, M. Genetic engineering of terpenoid metabolism attracts bodyguards to Arabidopsis. *Science* **2005**, *309*, 2070–2072. [[CrossRef](#)] [[PubMed](#)]
34. Bennett, M.H.; Mansfield, J.W.; Lewis, M.J.; Beale, M.H. Cloning and expression of sesquiterpene synthase genes from lettuce (*Lactuca sativa* L.). *Phytochemistry* **2002**, *60*, 255–261. [[CrossRef](#)]
35. Liu, Q.; Majdi, M.; Cankar, K.; Goedbloed, M.; Charnikhova, T.; Verstappen, F.W.A.; Vos, R.C.H.D.; Beekwilder, J.; Krol, S.V.D.; Bouwmeester, H.J. Reconstitution of the costunolide biosynthetic pathway in yeast and *Nicotiana benthamiana*. *PLoS ONE* **2011**, *6*, e23255. [[CrossRef](#)]
36. Zheng, L.F.; Chen, F.F.; Li, Y.J.; Li, C.F. Identification of a unique germacrene A oxidase from *Xanthium strumarium*. *Phyton-Int. J. Exp. Bot.* **2022**, *91*, 989–998. [[CrossRef](#)]
37. Swinnen, G.; Jacobs, T.; Pauwels, L.; Goossens, A. CRISPR-Cas-Mediated Gene Knockout in Tomato. *Methods Mol. Biol.* **2020**, *2083*, 321–341.
38. Li, B.; Cui, G.; Shen, G.; Zhan, Z.; Huang, L.; Chen, J.; Qi, X. Targeted mutagenesis in the medicinal plant *Salvia miltiorrhiza*. *Sci. Rep.* **2017**, *7*, 43320. [[CrossRef](#)]
39. Shi, T.L.; Zhang, L.; Cheng, Q.Y.; Yu, J.S.; Liu, J.; Shen, Y.J. Xanthatin induces apoptosis by activating endoplasmic reticulum stress in hepatoma cells. *Eur. J. Pharmacol.* **2019**, *843*, 1–11. [[CrossRef](#)]
40. Wang, L.; Wang, J.; Li, F.; Liu, X.; Chen, B.; Tang, Y.X.; Wang, M.K. Cytotoxic sesquiterpene lactones from aerial parts of *Xanthium sibiricum*. *Planta Med.* **2013**, *79*, 661–665. [[CrossRef](#)]
41. Yao, S.B.; Wang, W.Z.; Li, Y.J.; Wang, Y.S.; Li, Y.J.; Gao, L.P.; Xia, T. The gene cloning and expression analysis of C4H in tea plant (*Camellia sinensis*). *J. Tea Sci.* **2015**, *1*, 35–44.
42. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; Mcgettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.; Wilm, A.; Lopez, R. Clustal W and clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [[CrossRef](#)] [[PubMed](#)]
43. Chen, L.N.; Wang, Q.Y.; Deng, J.; Ming, S.F.; Gong, L.; Yu, K.; Xu, J.Z.; Wu, N. Cloning and analysis of two HMGR genes (AIHMGR) in *Atractylodes lancea*. *Genom. Appl. Biol.* **2020**, *4*, 1732–1740.