Supplementary information for

Pathway Deciphering and Heterologous Production of Pyrethrins

Yi Li(李毅)[1#], Xiaochun Xiao(肖晓纯)[2#], Shun Wen(温顺)[2#], Yao Zhi(支瑶)[2], Fangfang Chen(陈芳芳)[2], Yan Yao(姚燕)[2], Xueting Fang(方雪婷)[2], Jing Yang(杨静)[2], Ranjun Li(李冉君)[2], Denghong Yan(闫登鸿)[2], Kexin Liu(刘科新)[3], Qijun Jiang(蒋其军)[3], Kun Shen(沈坤)[2], Anwei Hou(侯安伟)[4], Zixin Deng(邓子新)[1,2], Li Lu(鲁丽)[*2], Tiangang Liu(刘天罡)[*1]

[#]These authors contributed equally: Yi Li, Xiaochun Xiao, Shun Wen.
[*]Correspondence: Li Lu luliwhu@whu.edu.cn, Tiangang Liu liutiangang@sjtu.edu.cn

The PDF file includes:

Materials and Methods
Supplementary Figure 1 to 39
References

Other Supplementary Materials for this manuscript includes the following:

Supplementary Table 1 (Microsoft Excel format). Summary of sequencing data of the *Tanacetum cinerariifolium* genome.
Supplementary Table 2 (Microsoft Excel format). Statistics of Hi-C assisted genome assembly.
Supplementary Table 3 (Microsoft Excel format). Statistics of chromosome length of the *T. cinerariifolium* genome.
Supplementary Table 4 (Microsoft Excel format). Evaluation of genome assembly completeness using BUSCO.
Supplementary Table 5 (Microsoft Excel format). Statistics of gene function annotation.
Supplementary Table 6 (Microsoft Excel format). Summary of gene family clustering.
Supplementary Table 7 (Microsoft Excel format). Summary of gene family expansions and contractions.
Supplementary Table 8 (Microsoft Excel format). Expression values of *TPS* candidate genes from the transcriptome of *T. cinerariifolium*.
Supplementary Table 9 (Microsoft Excel format). Expression values of 25 candidate *CYP450* genes from the transcriptome of *T. cinerariifolium*.
Supplementary Table 10 (Microsoft Excel format). Expression values of *UGT* genes from the transcriptome of *T. cinerariifolium*.
Supplementary Table 11 (Microsoft Excel format). Sequences of UGTs used for phylogenetic analysis in this study.
Supplementary Table 12 (Microsoft Excel format). Expression values of *SCPL* genes from the transcriptome of *T. cinerariifolium*.
Supplementary Table 13 (Microsoft Excel format). Yeast strains used in this study.
Supplementary Table 14 (Microsoft Excel format). Plasmids used in this study.

Supplementary Table 15 (Microsoft Excel format). Primer sequences used for plasmid and strain construction.
Supplementary Table 16 (Microsoft Excel format). Amino acid sequences of genes cloned and described in this study.
Supplementary Table 17 (Microsoft Excel format). *Ks* values and duplication time of genes involved in pyrethrins biosynthesis in *T. cinerariifolium*.

Materials and Methods

Plant materials

Tanacetum cinerariifolium plant samples were collected from Yuxi City (24° 20′ 49″ N, 102° 31′ 37″ E), Yunnan Province, China, one of the three largest pyrethrum production bases in the world. Fresh leaves, stems, roots and flower tissues were harvested and immediately frozen in liquid nitrogen, and then stored at −80 °C in the laboratory prior to genome sequencing, transcriptome sequencing, and metabolite analysis. The *Nicotiana benthamiana* plants were cultivated in an artificial climate chamber at a temperature of 23 ℃, with a 16-h light and 8-h dark cycle for maintenance. Before the *Agrobacterium* infection, the *N. benthamiana* plants grew for about 4 weeks, with regular watering during this period.

Chemicals

All solvents used for extractions, chemical synthesis and preparative HPLC were of HPLC grade. All solvents were purchased from Fisher Scientific. Chrysanthemol (**3**, CAS: 5617-92-5), chrysanthemic acid (**4**, CAS: 4638-92-0), α-linolenic acid (α-LeA, **10**, CAS: 463-40-1), *cis*-OPDA (**13**, CAS: 85551-10-6), and *cis*-jasmone (**17**, CAS: 488-10-8) standards were purchased from Sigma-Aldrich. Synthesis of compounds that are not commercially available is described in synthesis of compounds section.

Molecular biology kits

All gene and fragment amplifications were performed using Phanta® Max Super-Fidelity DNA Polymerase (Vazyme, China), while colony PCR reactions were performed using Taq Master Mix (Vazyme, China). All primers were purchased from Sigma Aldrich. PCR product purifications were performed using the FastPure Gel DNA Extraction Mini Kit (Vazyme, China). Plasmid purifications were performed using FastPure Plasmid Mini Kit (Vazyme, China). cDNA was prepared using Superscript™ IV VILO™ Master Mix and TURBOTM DNase (Thermo Fisher). All restriction enzymes were purchased from NEB (New England Biolabs).

DNA extraction, library preparation, and sequencing

For genome, high-quality genomic DNA was isolated from the fresh leaves of *T. cinerariifolium* using a modified cetyltrimethylammonium bromide method. For PacBio HiFi sequencing, libraries were prepared using Pacific Biosciences SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, USA) following the manufacturer's protocol. A SMRT Cell sequencing library was constructed and then

sequenced on a PacBio Revio system (Pacific Biosciences, USA). There was ~ 393 G raw PacBio data in total with an N50 length of ~21.7 Kb. For Illumina sequencing, qualified DNA samples were randomly interrupted to prepare a 150-bp inserted fragment library. Then, the Agilent 2100 Bioanalyzer system was used to verify the insertion size of the library, and qPCR was performed to ensure the effective quantitative concentration of the library. A total of ~503 Gb raw data was generated. The library was sequenced on the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA, USA). Finally, a Hi-C library was prepared from cross-linked chromatin in *T. cinerariifolium* cells according to the standard Hi-C protocol. The library was sequenced using the 150-bp paired-end mode on the Illumina NovaSeq 6000 platform. In total, we generated approximate 696.7 Gb (~97×) of Hi-C data.

RNA purification and sequencing

Total RNA was extracted from three replicate samples of *T. cinerariifolium* roots, stems, leaves, and different stages of flowers (S2, S4, and S6) using the TRIzol reagent (Invitrogen, CA, USA) according to the manufacturer's instructions. For long-read isoform sequencing, all libraries were subsequently sequenced on the PacBio Sequel platform with P6C4 polymerase enzyme and chemistry. Illumina short-read RNA paired-end libraries were then constructed using NEBNext® Ultra™ II RNA Library Prep Kit for Illumina® (New England Biolabs, MA, USA) and sequenced on the Illumina HiSeq 6000 platform.

Genome assembly

Hifiasm (v0.16.1) was used for self-correction and assembly with default parameters[1]. To improve the accuracy of the assembled contigs, the highly accurate short reads were then used to correct the assemblies with Pilon (v1.23) for two rounds[2]. Given the high heterozygosity presented in some genomic regions, we further used purge_dups (v1.2.5) to improve the resulting assembly by removing duplications and reassigning allelic contigs. The scaffolds were then anchored into chromosome-level scaffolds from Hi-C data using Juicer (v1.6)[3] and 3D-DNA (v180922)[4]. The quality of assembly following each round of polish was assessed using BUSCO (v5.4.7)[5] and QUAST (v5.0.2)[6]. The HiFi reads, Iso-Seq data, Illumina short reads, assembled transcripts, and RNA-Seq reads were mapped to the genome assembly with BWA (v0.7.13)[7], minimap2 (v2.17)[8], and HISAT2 (v2.1.0)[9], and the overall alignment rates were calculated, respectively.

Genome annotation

For gene prediction, we first filtered the PacBio Iso-Seq raw reads using SMRT Analysis to obtain circular consensus sequences (CCS) and non-CCS subreads, with subsequent analysis using the Iso-Seq3.1.2 pipeline[10]. After quality evaluation, CCS reads were corrected to remove primers and SMRT adapters using Lima (v1.9, https://lima.how/) to generate full-length (FL) reads. Full-length, non-chimeric (FLNC) reads were then collapsed, according to default parameters in Iso-Seq3 Cluster, to high-quality transcripts. Open reading frames were identified using

TransDecoder software (v5.5.0, https://github.com/TransDecoder/TransDecoder) to obtain putative genes, coding sequences (CDS) and protein sequences.

For genome repeat sequence annotation, we employed both homology-based and *de novo* approaches to search for transposable elements (TEs). De novo identification of repeat elements was performed using RepeatModeler (v2.0.3)[11]. Long terminal repeat retrotransposons (LTR-RTs) were predicted by structural characteristics using LTR_Finder (v1.07)[12] and LTRharvest[13]. LTR_retriever (v2.9.0)[14] was used to eliminate the redundancy of LTR sequence prediction results. All TE sequences generated by the different methods were combined with RepBase-20181026[15] and used as a library for repeat identification in RepeatMasker (v4.1.2)[16]. In addition, tandem repeats were identified with Tandem Repeat Finder (v4.09) [17] using default parameters.

Gene function annotation

Transcriptome prediction, homology prediction and *de novo* prediction were combined to annotate the gene structure. For homolog prediction, miniprot was used to align the homologous protein sequences from related species to the assembly and then report the corresponding gene structures. *De novo* prediction based on the genome sequence after masking the repeat sequences was analyzed using GeneMark-ET (v4.71)[18] and Augustus (v3.3.3)[19]. RNA-seq reads from seven tissues were aligned to the genome using HISAT2[9] and cDNAs were identified by using StringTie (v2.0.6)[20] and TransDecoder (v5.5.0, https://github.com/TransDecoder/TransDecoder). The gene set predicted by different methods was integrated to obtain the final set with Evidence Modeler (v2.0.0)[21].

Ribosomal RNAs (rRNAs) were predicted using RNAmmer (v1.2)[22], transfer RNAs (tRNAs) were predicted using tRNAscan-SE (v2.0.12)[23], and other non-coding RNAs, including microRNAs (miRNAs) and small nuclear RNAs (snRNAs), were identified using Infernal (v1.1.5)[24] against Rfam database (v14.10)[25]. Gene functional annotation was performed using the Pfam (http://pfam.sanger.ac.uk/), NR (https://ftp.ncbi.nlm.nih.gov/blast/db/), SwissProt (https://www.expasy.org/resources/uniprotkb-swiss-prot), TrEMBL[26], and KEGG (https://www.genome.jp/kegg/) databases. An E-value threshold of 1e-5 was applied, and annotations with the lowest E-values were selected for further analysis. Furthermore, InterProScan (v5.66-98.0)[27] was employed to identify motifs and domains by interrogating the InterPro databases (https://www.ebi.ac.uk/interpro/), subsequently providing Gene Ontology (GO, https://www.geneontology.org/) term annotations.

Evolutionary analysis of gene families

To investigate the evolutionary dynamics of gene families in *T. cinerariifolium*, we performed genome-wide gene family identification using OrthoFinder[28] for pyrethrum and 11 other plant species: *Artemisia annua*[29], *Arabidopsis thaliana*[30], *Amborella trichopoda*[31], *Chrysanthemum lavandulifolium*[32], *Chrysanthemum nankingense*[33], *Helianthus annuus*[34], *Mentha suaveolens*[35], *Nicotiana benthamiana*[36], *Taraxacum*

mongolicum[37], *Jasminum sambac*[38], and *Vitis vinifera*[39]. Genome and protein sequences of these species were retrieved from public databases. OrthoFinder was run with default parameters, resulting in the identification of 26,788 gene families in the pyrethrum genome, among which 8,527 gene families (containing 17,454 genes) were pyrethrum-specific. A maximum-likelihood species tree was constructed with RAxML[40] (model: PROTGAMMAWAG) based on 215 single-copy orthologous genes shared by all 12 species.

Gene family expansion and contraction across the 12 species were analyzed with CAFÉ[41] under default settings. The birth–death model was used to infer ancestral gene family sizes at each node, allowing the detection of changes relative to ancestors. In pyrethrum, 1,173 gene families were significantly expanded and 1,848 were contracted (**Supplementary Table 7**). KEGG enrichment analyses were performed for pyrethrum-specific and expanded gene families (**Supplementary Figs. 2, 3**). Finally, divergence times among the selected species were estimated using the mcmctree program in PAML[42].

Whole-genome duplication analysis

The synonymous substitution rate (*Ks*) method is commonly employed to identify whole-genome duplication (WGD) events. By plotting intra- and inter-species *Ks* distributions, we can infer whether a WGD occurred in the studied species and determine whether it took place before or after the speciation of two lineages. Specifically, protein sequences from different species were aligned using BLAST, and syntenic blocks were detected with MCScanX[43]. The paralogous and orthologous gene pairs within these blocks were then used to calculate the synonymous substitution rate (*Ks*), nonsynonymous substitution rate (*Ka*), and the *Ka*/*Ks* ratio with the yn00 module in PAML. Finally, density plots of *Ks* values were generated to visualize the distribution.

Microcolinearity analysis of genes

Microsynteny analysis was performed using the JCVI toolkit (based on the MCScan algorithm)[44]. Whole-genome alignments were conducted using gene annotations (BED format) and protein sequences to identify syntenic blocks. High-confidence anchor gene pairs were used to reconstruct syntenic relationships between species. Local gene neighborhoods (10 flanking genes) around target genes were extracted to examine microsynteny conservation. Only genes located on chromosomes were retained, and syntenic anchors were filtered based on collinearity score and gene density to ensure robustness.

Transcriptomic and Co-expression analysis

The expression levels of each transcript were quantified using HISAT2[9]. The FPKM (fragments per million mapped reads) value for each gene was calculated based on the gene length and the number of reads mapped to that gene. To construct the gene co-expression network, we used the gene expression data of the identified pathway genes from the *T. cinerariifolium* transcriptome samples. And then, we used the R software

package "WGCNA"[45] to establish a reliable co-expression network. In addition, we used genes that have been functionally characterized, such as *TcCDS, TcGLIP*, *TcJMH*, and the contents of acid moieties and pyrethrins in *T. cinerariifolium* tissues as the targets. We used Pearson correlation analysis to screen the candidate genes, with the screening criteria: correlation coefficient ≥ 0.7, *p*-value ≤ 0.05.

Metabolite analysis of *Tanacetum cinerariifolium*

In order to investigate the contents of each intermediate in the pyrethrins biosynthetic pathway, we conducted metabolite analysis on the *T. cinerariifolium* plant samples used for RNA analysis. The plant samples such as leaves, stems, roots and flowers, were rapidly ground into powder using liquid nitrogen. 100 mg of the tissue powder was weighed, then 1 mL of ethyl acetate (EA) was added for extraction, and the mixture was sonicated for 15 min at room temperature. The extracted homogenate was centrifuged at 13,000 × g for 10 min. Subsequently, the supernatant was placed in a nitrogen concentrator with a nitrogen flow rate of 3 mL/min. The extraction solvent EA was blown away under a nitrogen atmosphere, and the sample was concentrated to 200 μL. After another centrifugation, the supernatant was taken for subsequent GC-MS and LC-MS analysis.

Identification candidate genes

For the candidate genes in the pyrethrum genome that are involved in the biosynthesis of pyrethrins, we identified them using Hidden Markov Model[46] search based on the corresponding Pfam IDs from the Pfam database, such as LOX (lipoxygenase, PF00305), AOC (allene oxide cyclase, PF06351), ACX (acyl-CoA oxidase, PF01756), KAT (3-ketoacyl coenzyme a thiolase, PF02803 and PF00108), ACS (acyl-CoA synthetase, PF00501), SCPL (serine carboxypeptidase-like acyltransferase, PF00450), CYP450 (PF00067), and UGT (PF00201). In addition, we also conducted a blast using the homologous genes reported in the literature as references, such as *AtACX* and *AtMFP* from *Arabidopsis thaliana*, and these protein sequences were downloaded from the online database TAIR (https://www.arabidopsis.org/index.jsp). The protein sequences of all identified genes (such as *TcTPS*, *TcCYP450*, and *TcUGT*) were aligned using MAFFT (v7.475)[47]. Finally, phylogenetic trees of these genes were constructed using the IQ-TREE[48].

Candidate gene cloning

The full-length sequences of the genes of interest were amplified from cDNA using the primers listed in **Supplementary Table 15** and Phanta® Max Super-Fidelity DNA Polymerase (Vazyme) or PrimeSTAR® GXL DNA Polymerase (Takara) according to the manufacturer's instructions. For transient expression in *N. benthamiana*, the PCR products were purified from agarose gel and ligated into the modified pEAQ vector (digested with *Xho*I and *Sal*I purchased from Vazyme) using the Hieff Clone® Plus Multi One Step Cloning Kit (Yeasen). Gibson reactions were transformed into competent *E. coli* DH5α cells. Recombinant colonies were selected on LB agar plates supplemented with kanamycin (50 μg/mL). Positive clones were identified by colony

PCR. Plasmids were *iso*lated from positive colonies grown overnight. Identities of the inserted sequences were confirmed by Sanger sequencing. The constructs were then used to transform Agrobacterium tumefaciens strain GV3101. Recombinant colonies were selected on LB agar containing rifampicin (50 μg/mL), gentamicin (25 μg/mL) and kanamycin (50 μg/mL). Single colonies were grown in 10 mL of LB (50 μg/mL rifampicin, 25 μg/mL gentamicin, and 50 μg/mL kanamycin) at 28 °C for 1 day; glycerol stocks (20% glycerol) were prepared and stored at −80 °C for future use.

For heterologous expression in *E. coli*, coding sequence (CDS) was ligated into the pET28a vector using the Hieff Clone® Plus Multi One Step Cloning Kit (Yeasen) with the primers listed in **Supplementary Table 15**. Gibson reactions were transformed into competent *E. coli* DH5α cells. The constructed plasmids were transformed into Rosetta *E. coli*. Recombinant colonies were selected on LB agar containing chloramphenicol (34 μg/mL) and kanamycin (50 μg/mL). Single colonies were grown in 10 mL of LB (34 μg/mL chloramphenicol and 50 μg/mL kanamycin) at 37 °C for 12 h; glycerol stocks (20% glycerol) were prepared and stored at −80 °C for future use. Genes codon-optimized for *S. cerevisiae* were synthesized by GenScript Co., Ltd. (China).

Agrobacterium-mediated transient expression in *N. benthamiana*

Plasmids containing genes of interest were transformed into *Agrobacterium tumefaciens* GV3101 ($Gent^R$) using the freeze-thaw method, plated onto selective LB agar plates (50 μg/mL of kanamycin) and grown for 2 days at 30 °C. Positive transformants were verified through colony PCR and these were then inoculated into 4 mL of liquid LB cultures (50 μg/mL of kanamycin, 50 μg/mL of rifampicin and 25 μg/mL of gentamicin), which were shaken for 24 h at 30 °C. This mixture was centrifuged at 4,000 × g for 10 min to pellet the cells, after which the supernatant was removed. The cell pellet for each strain was resuspended in 4 mL of *Agrobacterium* induction buffer (10 mmol/L MES, 10 mmol/L $MgCl_2$, 100 μmol/L acetosyringone, pH 5.6) and incubated at room temperature for 1–2 h. The concentrations of cell resuspensions were measured by taking their optical density OD_{600}. For individually tested strains, *Agrobacterium* suspensions were diluted to OD_{600} = 0.4–0.6. When strains were used in combination, the *Agrobacterium* suspensions for each strain were mixed in equal concentration. Prepared *Agrobacterium* suspensions were infiltrated into the abaxial side of *N. benthamiana* leaves of 4–5-weeks-old plants with a needleless 1mL syringe. For a typical experiment, three leaves from three different plants were used for each strain mixture to minimize any batch effects or biological variation among plants. Following infiltration, plants were grown as usual for 3–4 days, after which leaves were ex*cis*ed for subsequent metabolite extraction. For substrate co-infiltration experiments, plants were grown for 3 days after *Agrobacterium* infiltration, after which substrate (100 ppm in water with 1% DMSO) was infiltrated into the infected portion of the leaf using a needleless syringe. Leaves were harvested 1 day later, snap-frozen in liquid nitrogen and stored at −80 °C for later processing.

Luciferase complementation assay (LCA)

The CDSs of the *TcCDS* and *TcTPS1* were cloned into the C and N terminus of luciferase (LUC) in the pCAMBIA-LUC vector to generate *TcCDS*-cLUC, and *TcTPS1*-nLUC constructs, respectively. The *Agrobacterium* strain GV3101 carried with nLUC and cLUC recombinant plasmids was co-transfected into *N. benthamiana* leaves by mixing equal volumes. Plants were then incubated at 22 °C for 2 days before the LUC activity was measured.

Expression and enzymatic reaction of recombinant enzymes in *E. coli*

Chemically competent Rosetta, *E. coli* cells were transformed with the pET28a vector carrying a coding sequence of interest and selected on LB agar with chloramphenicol (34 μg/mL) and kanamycin (50 μg/mL). Colonies for each construct were picked, inoculated in 5 mL of LB medium with chloramphenicol (34 μg/mL) and kanamycin (50 μg/mL) and incubated overnight at 37 °C and 220 rpm. 1 mL of the culture was transferred to 100 mL of fresh LB medium and grown at 37 °C and 220 rpm until OD_{600} reached ~0.6, when the temperature was dropped to 18 °C and 0.1 mmol/L IPTG was added. After overnight incubation at 18 °C and 220 rpm, cells were harvested via centrifugation at 3,000 × g for 10 min at 4 °C. The pellet was resuspended in 8 mL of lysis buffer containing 50 mmol/L sodium phosphate (pH 8.0). After the cells were disrupted via sonication, the crude enzyme preparation was either used immediately for enzymatic assays or stored at −80 °C for later use. For enzymatic reactions, 1 mmol/L $MgCl_2$, 1 mmol/L substrate and co-factors were added to 1 mL crude preparation described above. Incubation was carried out at 30 °C for 24 h and then stopped by adding 1 mL EA.

As for TcCDS, TcTPS1, TcUGT6 and TcGLIP, the crude lysates were centrifuged at 10,000 × g for 15 min. Ni^{2+}-NTA agarose resin was first equilibrated with 20 mmol/L imidazole buffer. After adding supernatant, the target protein was washed with 30 mmol/L imidazole buffer and then eluted with 250 mmol/L imidazole buffer. The purity of the eluted protein was estimated by SDS–PAGE, and the protein sample was stored at -80 °C. For enzymatic reactions, reaction mixtures consisted of 100 μg TcCDS, TcTPS1, TcUGT6 or TcGLIP, 150 μmol/L substrate, 1 mmol/L $MgCl_2$, 2 mmol/L DTT, 12.5% glycerol, 0.1% Tween-20, 1 mmol/L ascorbic acid in a total volume of 100 μL Tris-HCl (pH 7.5, 50 mmol/L). Reactions were incubated at 30 °C for 12 h. Empty vector served as the negative control. After incubation, the reactions were quenched by addition of 1 volume of EA, filtered through 0.22-μm PTFE filters and analyzed by untargeted GC-MS. Enzymatic kinetic parameters were calculated using GraphPad Prism software (version 9.0).

Heterologous expression in *S. cerevisiae*

The *S. cerevisiae* strain *JCR27*[49], which was engineered to boost terpenoid production, was used as the parent strain for the production of *trans*-Chrysanthemic acid. The *S. cerevisiae* strain *30000B* was as the parent strain for the production of 3,7-ddh-JA. DNA transformation of *S. cerevisiae* was performed using the LiAc/SS carrier DNA/PEG method[50]. Mutant strains were confirmed by colony-PCR. Growth

of yeast cultures and heterologous protein production was performed much as previously described[51]. Freshly streaked colonies from each strain of interest were inoculated in 4 ml synthetic drop-out medium and were then grown at 30 °C with 220 rpm shaking, used as seed cultures. After 2 days of growth, the seed cultures were transferred to 250-mL shake flasks, each containing 50 mL YPDG medium (10 g/L yeast extract, 20 g/L peptone, 1 g/L glucose and 1% galactose), and adjusted to an initial optical density at 600 nm (OD_{600}) of 0.1. The yeast cultures were incubated at for 3-5 days before extraction.

For whole-cell biotransformation, the *S. cerevisiae* strain *30000B* containing corresponding plasmids was precultured in 5 mL of SD medium for 24 h at 30 °C and 220 rpm. The seed culture was scaled up in a 250 mL baffled culture flask containing 50 mL of SD liquid medium with a starter OD 600 incubation, when the OD 600 of 0.1. After 48 h of reached ~5.0, 50 OD 600 units of cell pellets were collected through centrifugation at 3,000 × g for 5 min and then resuspended in 1 mL of 50 mM potassium phosphate buffer, pH 7.4. Resting cell assays were initiated by adding a substrate stock solution. After incubation at 30 °C and 220 rpm, all samples were extracted with 1 mL of ethyl acetate on a vortex mixer for 5 min at full speed, and the organic phase was subjected to subsequent GC-MS analysis.

Subcellular localization of TcCDS and TcTPS1 using GFP

The *TcCDS* and *TcTPS1* genes without the stop codon were cloned into the vectors pL1300-cYFP and pL1300-nYFP using the *Sal*I restriction site, respectively and then transformed into GV3101 competent cells. Transgenic *N. benthamiana* mesophyll protoplasts were *iso*lated according to Sheen[52]. Images were taken using a confocal laser-scanning microscope (Leica, Germany) with a fluorescence band and filters of 620–750 nm for chlorophyll imaging and 500–530 nm for YFP.

LC-MS analysis

The relative content of *cis*-OPDA, *iso*-OPDA, 4,5-ddh-JA, and 3,7-ddh-JA in samples were measured by LC–MS as described below. All the supernatant of samples were filtered through a Fisher polytetrafluoroethylene (PTFE) syringe filter (0.22 μm) before being injected. For LC-MS analysis of these compounds, samples were analyzed using an Agilent liquid chromatograph (1290 Infinity Liquid Chromatography, 1290 Infinity LC, Agilent) coupled with a triple quadrupole mass spectrometer (6475 Triple Quadrupole Mass Spectrometer, 6475/TQ-MS, Agilent). The liquid chromatographic system was equipped with a C18 column (250 mm × 4.6 mm, 5 μm, Agilent), and the injection volume for each sample was 3 μL.

The eluent for the liquid phase was set as follows: ultra-pure water containing 0.1% (v/v) formic acid (A); acetonitrile with 0.1% (v/v) formic acid (B). The gradient of the mobile phase is set as: 0 min, 10% B; 0–15 min, 10%–60% B; 15–25 min, 60%–95% B; 25–32 min, 95% B; 32–32.1 min, 95%–10% B; 32.1–40 min, 10% B; The column temperature was 40 °C, and the column flow rate was 0.4 mL/min. The mass spectrometry was performed using TQ-MS. The conditions of the mass spectrometer were set as follows: full-scan mode, scan range: 50–500 *m*/*z*; scan time: 300 ms;

Fragmentor: 120 V; CAV: 8 V; Ion mode: positive and negative; Capillary voltage: 4000 V; Nozzle voltage: 1500 V; Gas temperature: 300 °C; Sheath gas temperature: 250 °C.

To detect glycosylated products and chrysanthemyl-CoA compounds in different tissues of pyrethrum, tissues were ground into powder under liquid nitrogen and extracted with 80% methanol/water. The mixture was then subjected to ice-bath sonication for 30 min, followed by centrifugation at 10,000 × g for 10 min. The supernatant was filtered through a PTFE membrane, and the resulting extract was analyzed by high-resolution LC-MS/MS.

The high-resolution mass spectrometry analysis was carried out using an UPLC system (Thermo Scientific, USA) coupled to a Q Exactive Plus mass spectrometer (Thermo scientific, USA) equipped with an ACQUITY UPLC BEH C18 column (100 mm × 2.1 mm, 1.7 μm, Waters) maintained at 45 °C. The mobile phase consisted of water containing 0.1% formic acid as phase A and acetonitrile containing 0.1% formic acid as phase B, with a linear gradient from 95% to 0% A. The flow rate was set to 0.4 mL/min, and the injection volume was 1 μL. Mass spectrometry analysis utilized a heated electrospray ionization (HESI) source in negative ionization modes. Full-scan spectra (100–1500 m/z) were acquired in data-dependent acquisition mode with one full scan (resolution: 70,000; automatic gain control (AGC)).

GC-MS analysis

For compounds such as *cis*-jasmone, jasmolone, and jasmolin I, we conducted monitoring and analysis using GC-MS. Using the Agilent gas chromatography (Gas chromatography, 8890/GC, Agilent) coupled with a triple quadrupole mass spectrometer (7000E Triple Quadrupole Mass Spectrometer, 7000E/TQ-MS, Agilent), and in combination with the DB-5Q column (30 m × 0.25 μm × 0.25 mm). The analysis parameters for the GC-MS were as follows: each sample injection volume is 1 μL; The column temperature was programmed as follows: 50 °C hold for 1 min, then increased at a rate of 15 °C/min to 110 °C, followed by a 6 °C/min increase to 200 °C, hold for 1 min; and then 10 °C/min to 280 °C. Total time for the program was 30.33 min. The injection port temperature was 240 °C; split ratio: 15:1; MSD transmission line temperature: 280 °C; ion source temperature: 280 °C; quadrupole temperature: 150 °C; peak width set to 0.8 seconds; scanning mode: MS1 scan; mass range: 45–500 m/z, scan time: 200 ms.

For precise mass confirmation of the compounds, synthesized standards and reaction products were further analyzed using a Thermo Scientific™ GC 240 coupled with Orbitrap™ Exploris™ mass spectrometer. The GC temperature program was identical to that of the Agilent 8890 GC system. MS parameters were as follows: scan range: 45–500 m/z; Orbitrap resolution: 60,000; AGC target: standard; Ion source temp: 250 °C; Ionization mode: EI; Mass tolerance: 5 ppm.

Protein crystal modeling

The three-dimensional structure of the receptor protein was predicted usingAlphaFold3[53]. Molecular docking was performed using the AutoDock Vina 1.2.3

computational engine[54]. Upon completion, all generated poses were ranked by the system according to the calculated binding free energy (ΔG, in kcal/mol) by AutoDock Vina. The top-ranked pose (i.e., the one with the lowest binding free energy) was selected for subsequent analysis.

The predicted binding mode of the protein-ligand complex was visualized and analyzed using PyMOL 3.0.0 (Schrödinger, LLC, https://pymol.org/pymol.html). In the rendered images, the receptor protein was represented as a cartoon. Key amino acid residues within the predicted binding pocket that interact with the ligand were displayed as sticks. The ligand molecule was also presented in stick representation. Critical non-covalent interactions, such as hydrogen bonds, were detected and annotated using the analysis tools built into PyMOL.

Chemical synthesis

10-OH chrysanthemol was synthesized according to the reported method[55, 56]. Subsequently, the main product was obtained through silica gel chromatography. The structure of the product confirmed as 10-OH chrysanthemol (**Supplementary Fig. 10e-10g**).

As for the synthesis of 4,5-ddh-JA, following the reported method[57], we used MeJA as the precursor and carried out the reaction under the catalysis of Palladium (II) acetate, diethylallyl phosphate, and Na_2CO_3 at 80 °C for 24 h. Then, 4,5-ddh-MeJA was obtained through silica gel column. Under the catalysis of sodium methoxide, the reaction was carried out at 60 °C for 24 h and then purified by column chromatography to obtain 3,7-ddh-MeJA (**Supplementary Fig. 17a, 17b**). GC-MS analysis of 4,5-ddh-MeJA showed prominent ions at *m*/*z* 222.1252 (M^+), 193.0864 (M^+ - C2H5), 154.0625 (M^+ - C5H8); 3,7-ddh-MeJA at *m*/*z* 222.1252 (M^+), 193.0864 (M^+ - C2H5), 149.0966 (M^+ - C3H5O2) (**Supplementary Fig. 17c**). Finally, the methyl groups were hydrolyzed using porcine liver esterase (PLE, (Sigma Chemical; Lot No. E2884-1KU))[58], resulting in the formation of 4,5-ddh-JA and 3,7-ddh-JA (**Supplementary Fig. 17d**).

Regarding the synthesis of *cis*-OPDA, we followed the reported method and made some modifications[59]. In brief: Using the GmLOX, AtAOC, and AtAOS proteins expressed in Rosetta (DE3) *E. coli*, mix them in equal volumes. Add 25 mg of α-LeA (sigma) in 200 μL DMSO at 25 °C for 24 h. After the reaction, adjust the pH of the reaction mixture to 5 by adding 1 mol/L hydrochloric acid. Extract the reaction mixture with ethyl acetate 4 times. Subsequently, all the ethyl acetate extracts were combined and the solvent was removed through vacuum and low-temperature evaporation. Add 200 μL ethyl acetate to dissolve the extracted substance again. Purify the extract using a silica gel plate: SiO_2 gel plate, 0.5 mm × 20 cm × 20 cm; petroleum ether: ethyl acetate = 25:25 solvent). The synthesis method of isotopically labeled *cis*-OPDA is the same (**Supplementary Fig. 15**).

After obtaining *cis*-OPDA, it was dissolved in 10% methanol hydrochloric acid, and the reaction was carried out at 60 °C for 1 h. And then, following the strategy for synthesizing 3,7-ddh-JA, *iso*-OPDAMe was obtained from the reaction (**Supplementary Figs. 15, 17**). GC-MS analysis of the methyl ester of *cis*-OPDA

showed prominent ions at *m*/*z* 306.2195 (M^+), 275.2012 (M^+ - CH3O), 238.1570 (M^+ - C5H8), 149.0967 (M^+ - C9H17O2); *iso*-OPDAMe at *m*/*z* 306.2195 (M^+), 275.2012 (M^+ - CH3O), 177.1280 (M^+ - C7H13O2), 149.0967 (M^+ - C9H17O2) (**Supplementary Fig. 17**). Finally, the methyl group was removed through the PLE hydrolysis, resulting in *iso*-OPDA (**Supplementary Fig. 15**).

Preparative synthesis of chrysanthemoyl-glucoside

Chrysanthemoyl-glucoside (**8**) was synthesized enzymatically using purified recombinant TcUGT6. In brief, chrysanthemic acid (**4**) (2 mmol/L) was incubated with uridine 5′-diphosphoglucose disodium salt (4 mmol/L; CAS: 28053-08-9) and purified TcUGT6 (2 µg) in 50 mmol/L Tris–HCl pH 7.5 containing 1 mmol/L $MgCl_2$ and 2 mmol/L DTT in a total volume of 1 mL at 30 °C for 12 h; UDP-glucose was supplied in excess to drive the reaction to completion. The reaction was quenched by adding 1 volume methanol, and nucleotides and salts were removed by C18 solid-phase extraction. The residue was further purified by semi-preparative reversed-phase HPLC (Eclipse XDB-C18, 5 µm, 9.4 mm × 250 mm; mobile phase A, water containing 0.1% formic acid; mobile phase B, acetonitrile; gradient, 60%–70% B over 20 min; flow rate, 3 mL/min; detection at 210 nm). The target fraction was collected, lyophilized, and re-dissolved in methanol. The identity of **8** was confirmed by LC-MS/MS ($[M+HCOOH–H]^-$ ion; *m*/*z* = 375.1), and its purity was estimated to be > 95% by analytical HPLC-UV at 210 nm. The compound was aliquoted and stored at −80 °C until use as the acyl donor in TcGLIP assays.

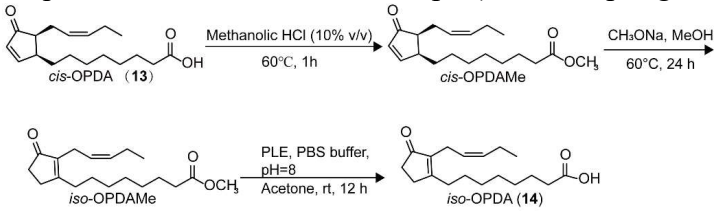
TcGLIP acyltransferase assay with defined donors.

Standard reactions (200 µL) contained 50 mmol/L Tris–HCl (pH 8.0), 2 mmol/L jasmolone (**18**), 4 mmol/L donor (purified Chrysanthemoyl-glucoside (**8**), or equimolar chrysanthemoyl-CoA (**7**), or chrysanthemic acid (**4**)), 2 µg purified TcGLIP; reactions lacking enzyme or lacking **18** received buffer instead of the omitted component. Donor stock solutions of purified chrysanthemoyl-glucoside (**8**) were prepared in DMSO so that the final organic content did not exceed 5 % (v/v); the same solvent proportion was maintained across all conditions, including controls. Reactions were incubated at 25 °C for 12 h and 24 h, and terminated by addition of 10 µL acetic acid. The reaction mixture was then extracted with an equal volume of n-hexane, and the organic phase was analyzed by GC-MS under the same conditions described above. Jasmolin I (**21**) was identified by comparison of its retention time and mass spectrum with those of an authentic standard, and the characteristic ion at *m*/*z* = 123.1 was monitored. Each condition was performed with three biological replicates.

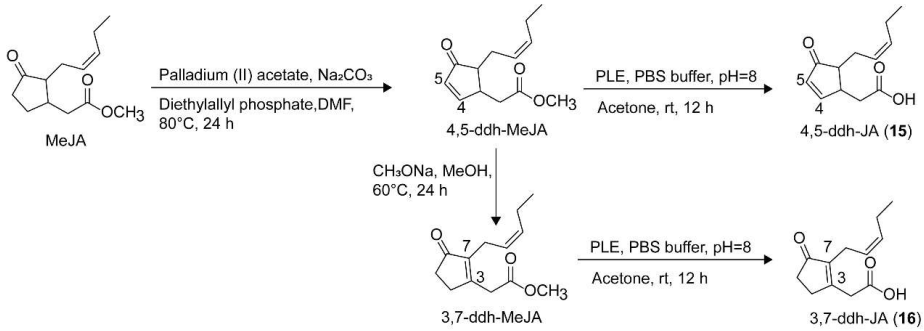
NMR analysis

NMR spectra were recorded on a Bruker 400 MHz NMR spectrometer (Avance III) or Bruker 600 MHz Avance III spectrometers (Bruker), at ambient temperature with $CDCl_3$ as the solvent unless otherwise stated. Chemical shifts are reported in parts per million relative to $CDCl_3$ (1H, δ 7.26; 13C, δ 77.1). Data for 1H NMR are reported as

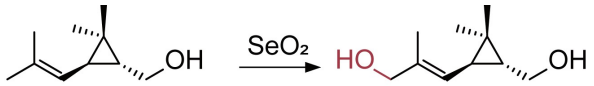
follows: chemical shift, integration, multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, quint = quintet, sixt = sixtet, m = multiplet), and coupling constants.



Data of **14 (*iso*-OPDA):** ^{1}**H NMR** (600 MHz, CDCl3): δ 5.41-5.30 (m, 1H), 5.26-5.16 (m, 1H), 2.93 (d, *J* = 7.2 Hz, 2H), 2.51-2.47 (m, 2H), 2.42 (t, *J* = 8.4, 2H), 2.38-2.35 (m, 4H), 2.15 (quint, *J* = 7.2 Hz, 2H), 1.65-1.62 (m, 2H), 1.54-1.49 (m, 2H), 1.36-1.33 (m, 6H), 0.99 (t, *J* = 7.8 Hz, 3H). ^{13}C NMR (150 MHz, CDCl3): δ 209.65, 177.94, 174.50, 139.36, 132.39, 125.52, 34.34, 33.75, 31.42, 29.67, 29.31, 29.18, 29.04, 27.50, 24.75, 21.39, 20.74, 14.30.

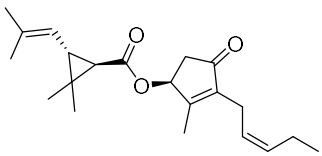


Data of **16 (3,7-ddh-JA)**: ^{1}**H NMR** (600 MHz, CDCl3): δ 5.47-5.37 (m, 1H), 5.25-5.13 (m, 1H), 3.52 (s, 2H), 3.40-2.95 (m, 2H), 2.69-2.61 (m, 2H), 2.48-2.41 (m, 2H), 2.18-2.09 (m, 2H), 0.99 (t, *J* = 7.8 Hz, 3H). ^{13}C NMR (150 MHz, CDCl3): δ 208.8, 173.8, 163.5, 142.2, 133.3, 124.2, 36.3, 34.3, 29.9, 21.3, 20.6, 14.1.



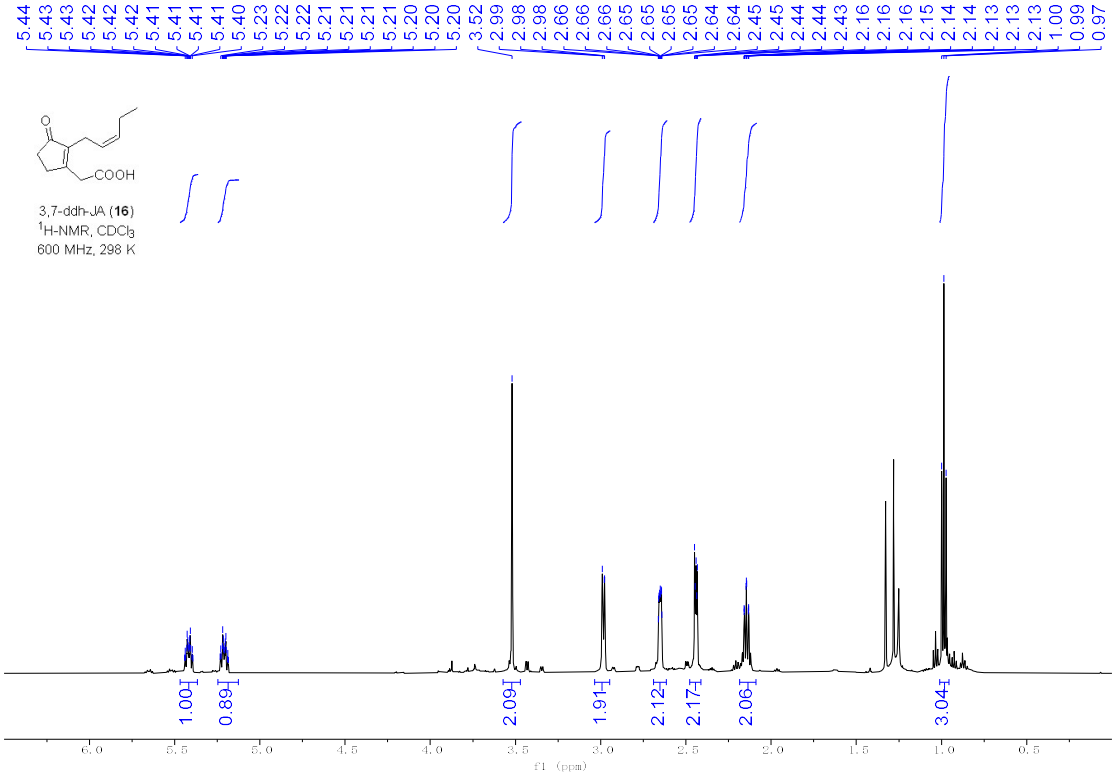
Chrysanthemol (3) 10-OH chrysanthemol (5)

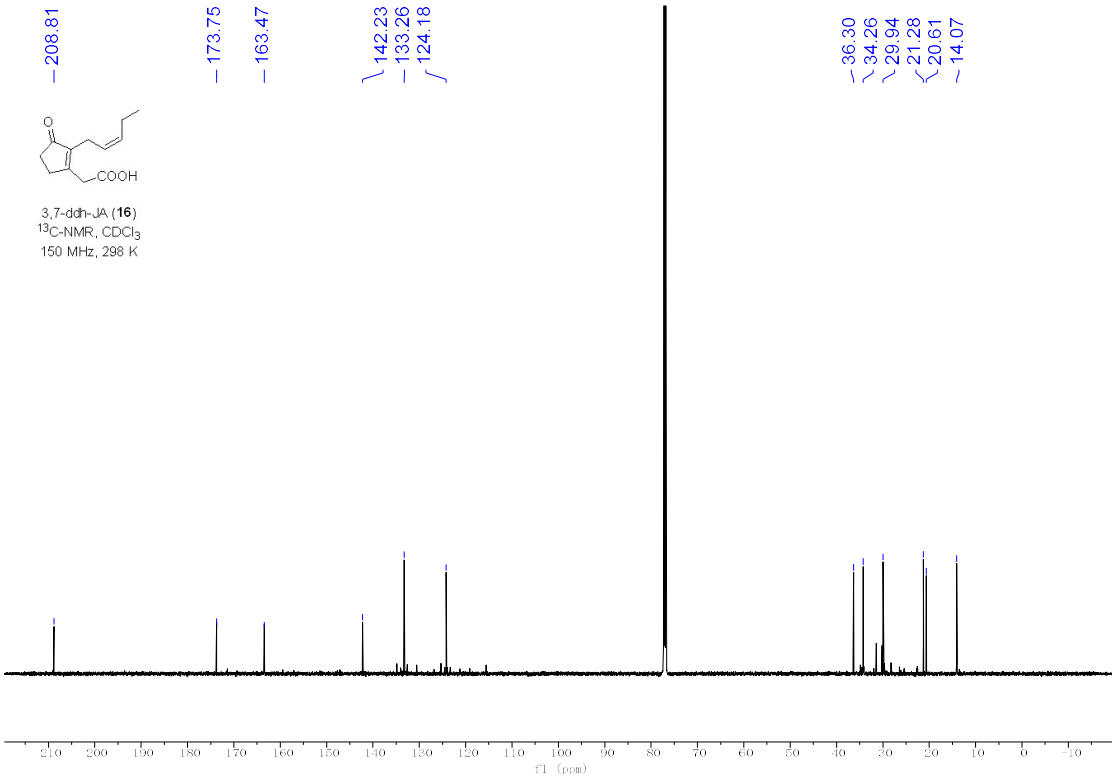
Data of **5 (10-OH Chrysanthemol)**: ^{1}**H NMR** (600 MHz, $CDCl_3$) δ 5.16 (dd, *J* = 8.6, 1.4 Hz, 1H), 4.01 (d, *J* = 1.2 Hz, 2H), 3.77 (dd, *J* = 11.5, 6.6 Hz, 1H), 3.60 (dd, *J* = 11.5, 8.4 Hz, 1H), 1.74 (d, *J* = 1.6 Hz, 3H), 1.27 – 1.24 (m, 1H), 1.17 (s, 3H), 1.09 (s, 3H), 0.93 (ddd, *J* = 8.4, 6.6, 5.1 Hz, 1H). ^{13}C NMR (151 MHz, $CDCl_3$) δ 135.9, 125.5, 68.9, 63.5, 35.5, 28.3, 22.9, 22.7, 21.4, 14.2.

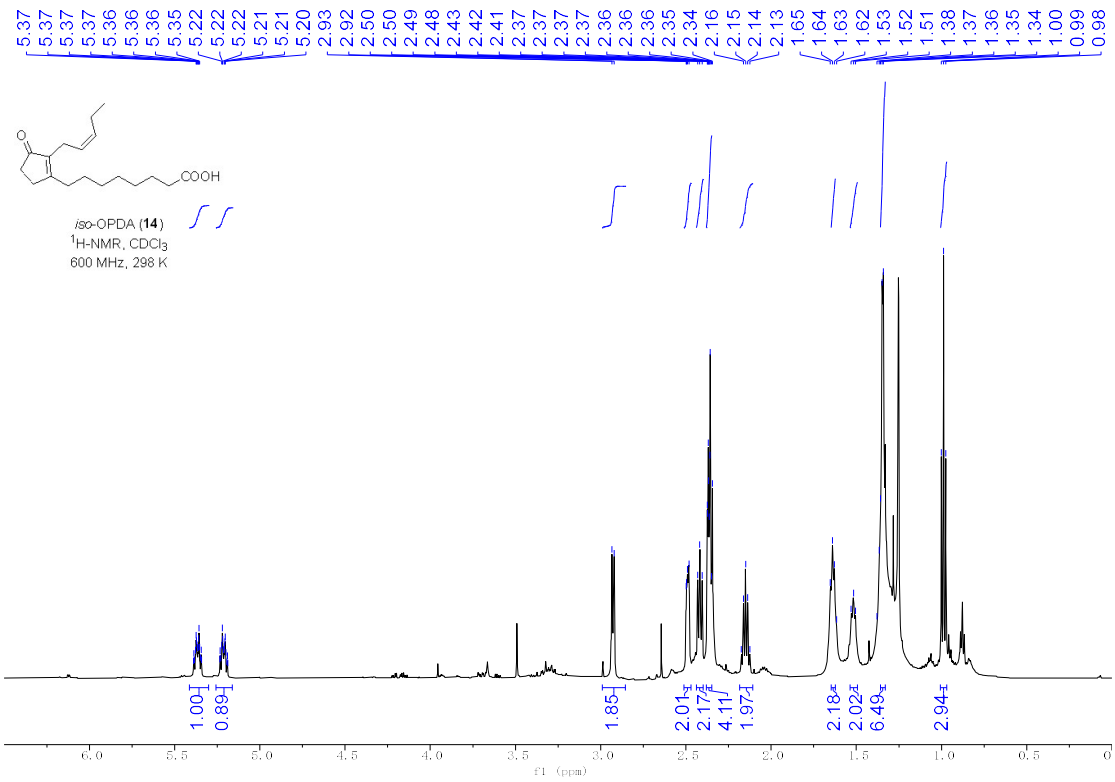


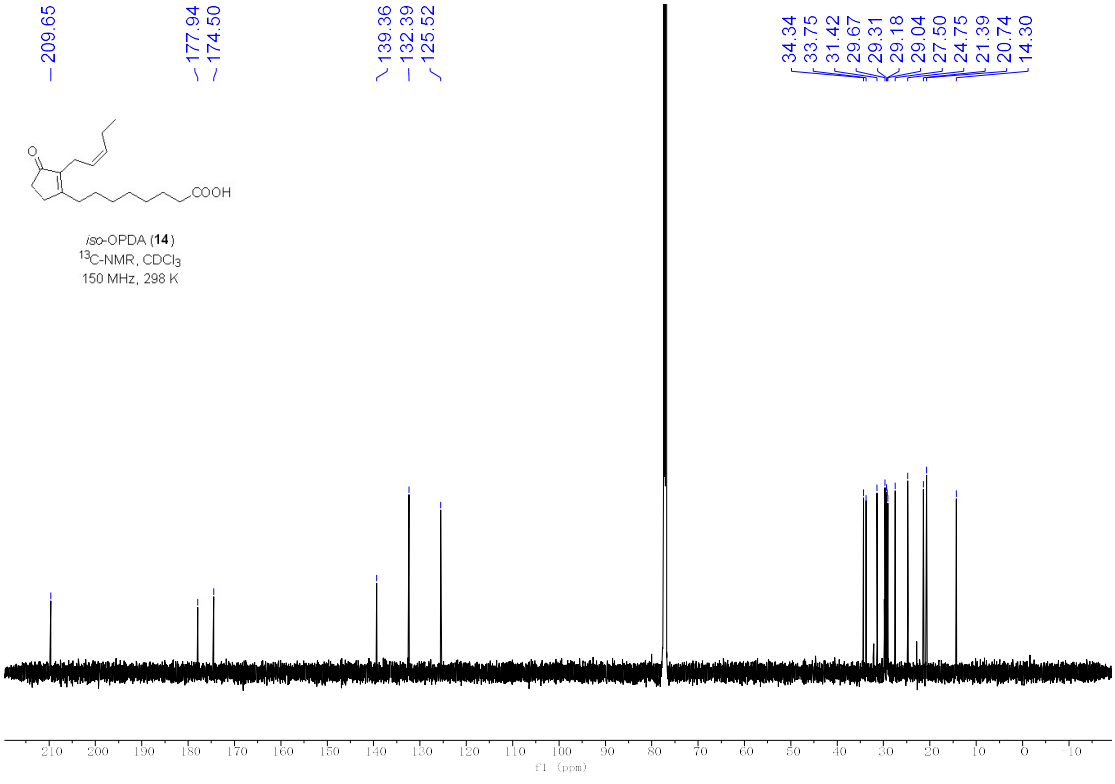
Data of **21 (jasmolin I):**^{1}H NMR (600 MHz, CDCl3): δ 5.68-5.63 (m, 1H), 5.47-5.37 (m, 1H), 5.28-5.19 (m, 1H), 4.92-4.89 (m, 1H), 2.98 (d, *J* = 7.2 Hz, 2H), 2.87 (dd, *J* = 6.0, 18.6 Hz, 1H), 2.22 (dd, *J* = 2.4, 18.6 Hz, 1H), 2.18-2.11 (m, 2H), 2.11-2.06 (m, 1H), 2.04 (s, 3H), 1.72 (s, 3H), 1.71 (s, 3H), 1.41 (d, *J* = 5.4 Hz, 1H), 1.26 (s, 3H), 1.14 (s, 3H), 0.99 (t, *J* = 7.8 Hz, 3H). ^{13}C NMR (150 MHz, CDCl3): δ 204.17, 172.49, 164.99, 142.91, 136.05, 133.32, 124.06, 120.96, 73.14, 42.23, 34.71, 33.09, 29.27, 25.72, 22.26, 21.36, 20.75, 20.56, 18.65, 14.24, 14.20.

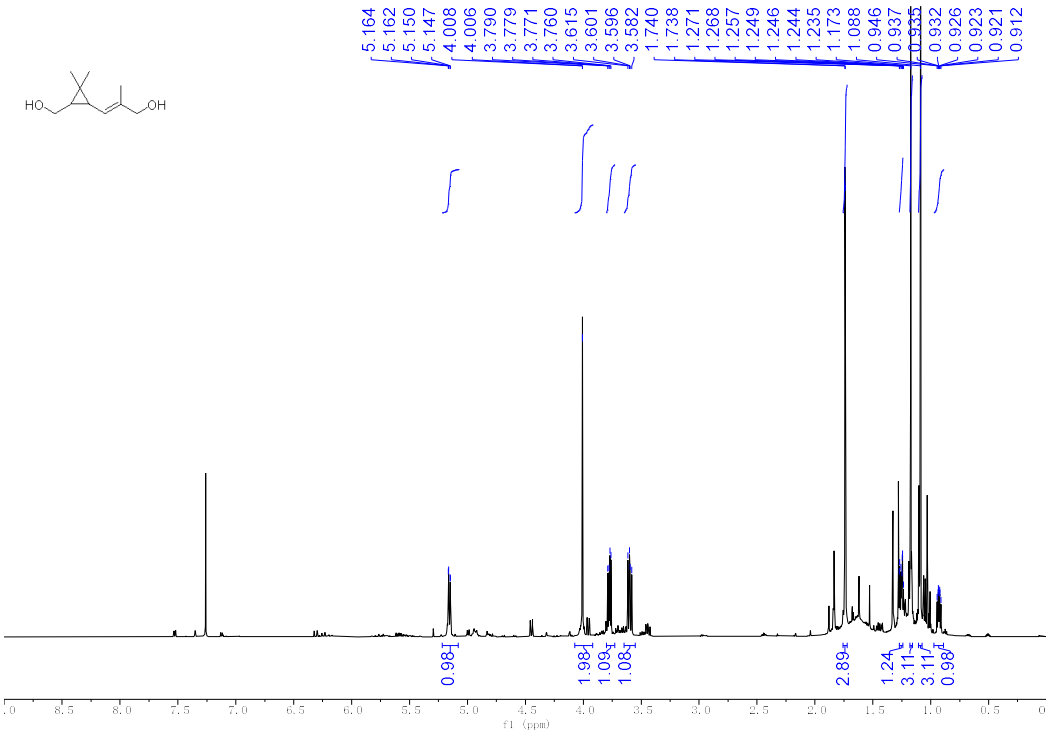
NMR spectrum

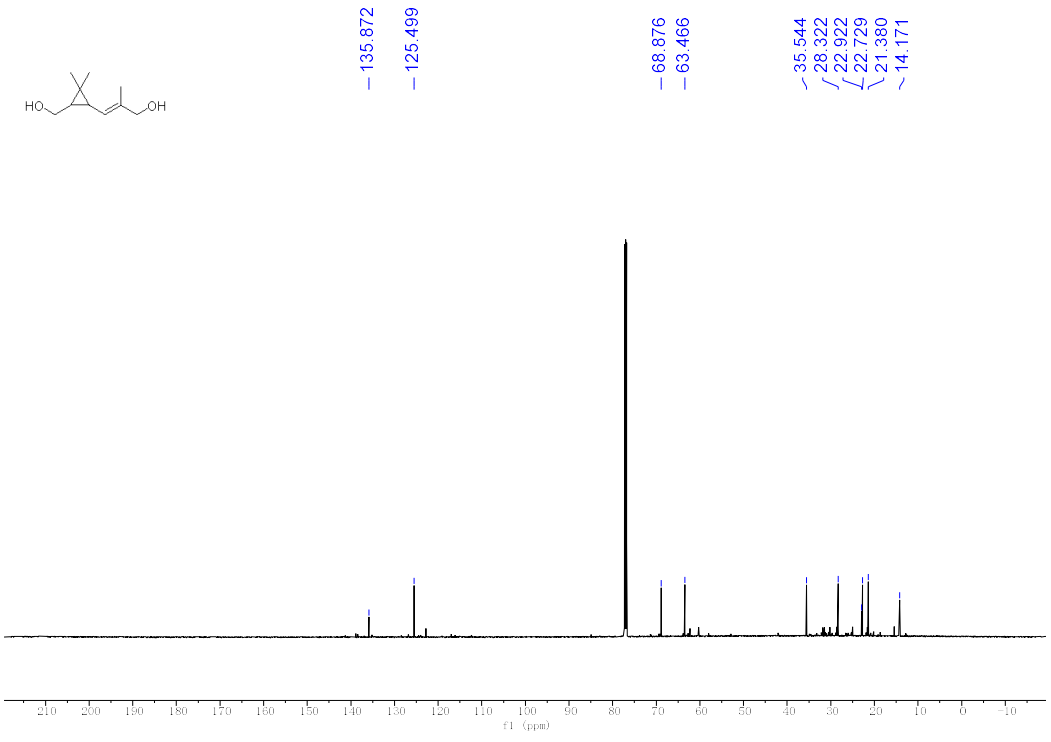


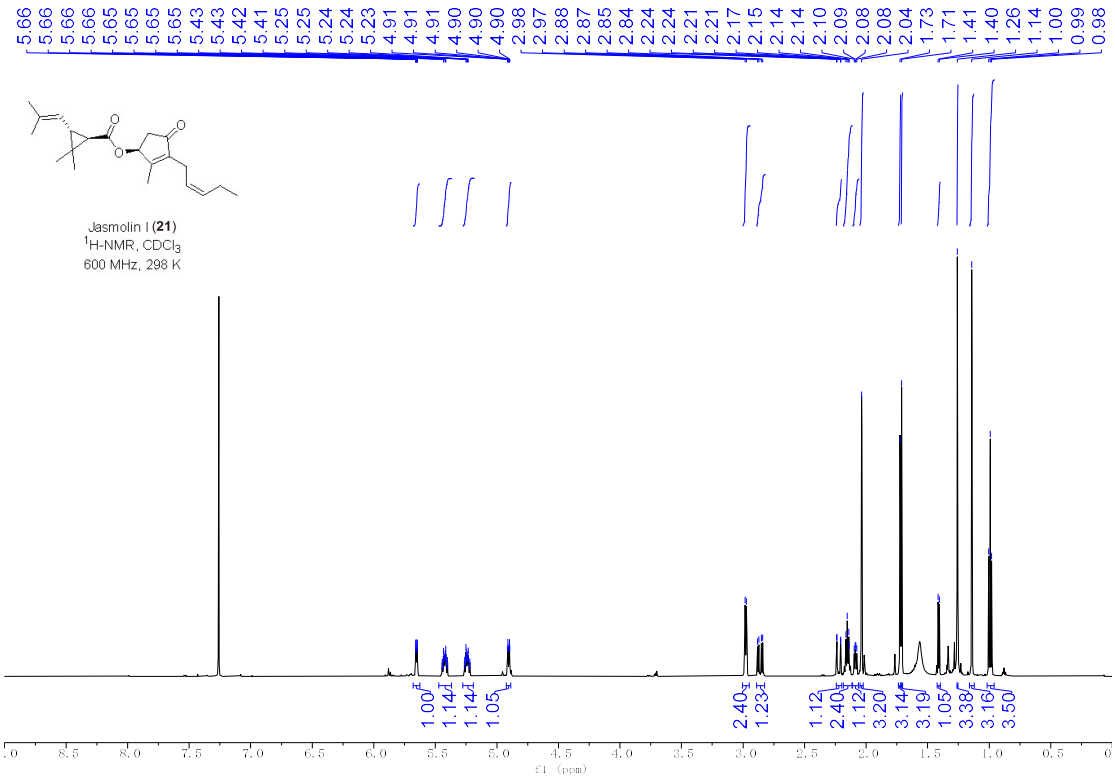


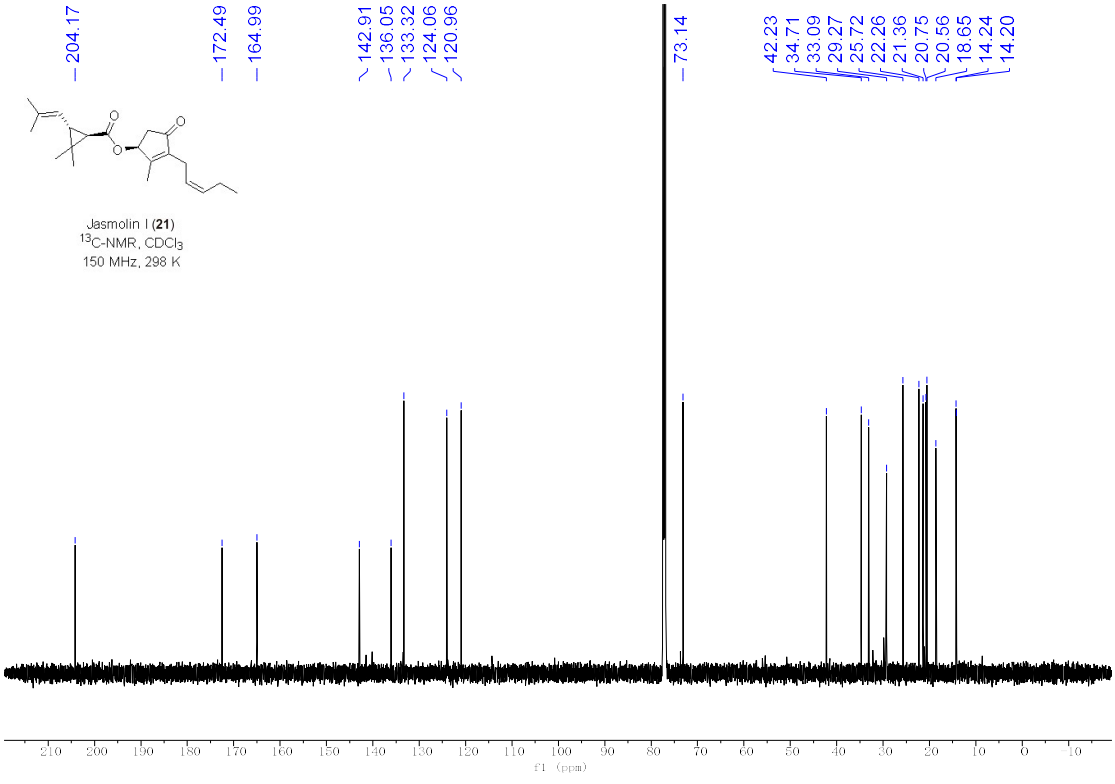

5.37 5.37 5.37 5.37 5.36 5.36 5.36 5.35 5.22 5.22 5.22 5.21 5.21 5.20 2.93 2.92 2.50 2.50 2.49 2.48 2.43 2.42 2.41 2.37 2.37 2.37 2.37 2.36 2.36 2.36 2.35 2.34 2.16 2.15 2.14 2.13 1.65 1.64 1.63 1.62 1.53 1.52 1.51 1.38 1.37 1.36 1.35 1.34 1.00 0.99 0.98
COOH
iso-OPDA (14)
1H-NMR, CDCl3
600 MHz, 298 K
1.00 0.89 1.85 2.01 2.17 4.11 1.97 2.18 2.02 6.49 2.94
f1 (ppm)


209.65 177.94 174.50 139.36 132.39 125.52 34.34 33.75 31.42 29.67 29.31 29.18 29.04 27.50 24.75 21.39 20.74 14.30
COOH
iso-OPDA (14)
13C-NMR, CDCl3
150 MHz, 298 K
f1 (ppm)

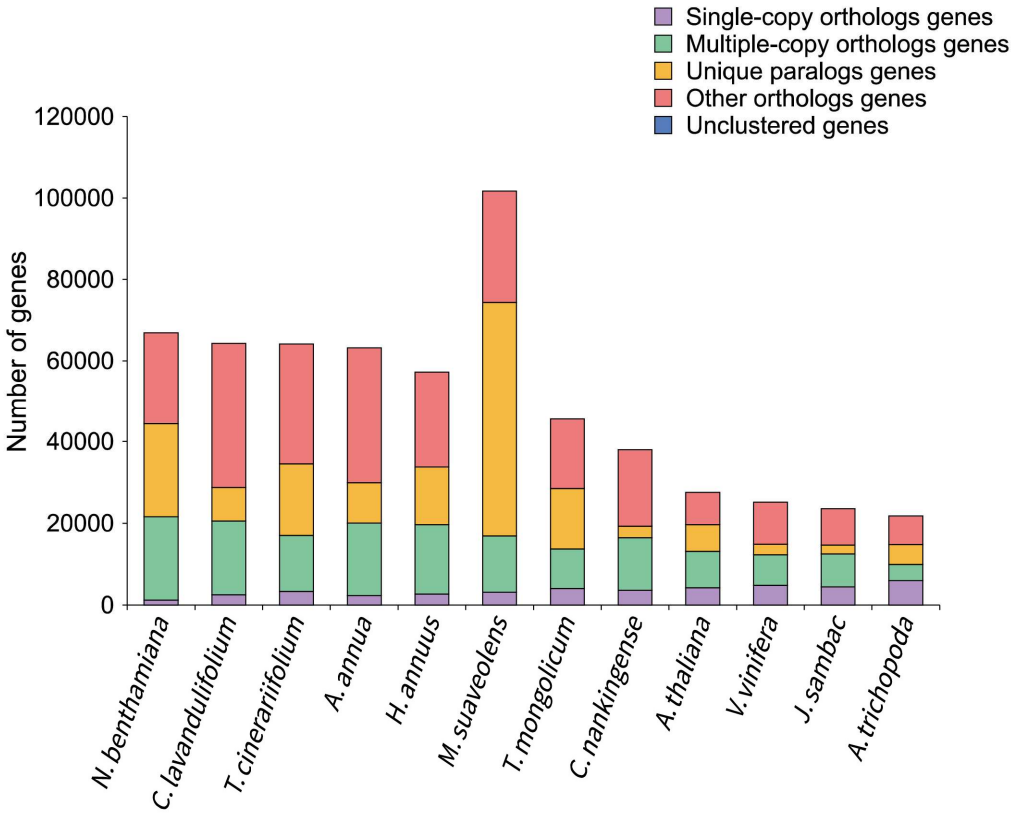

HO
OH
5.164
5.162
5.150
5.147
4.008
4.006
3.790
3.779
3.771
3.760
3.615
3.601
3.596
3.582
1.740
1.738
1.271
1.268
1.257
1.249
1.246
1.244
1.235
1.173
1.088
0.946
0.937
0.935
0.932
0.926
0.923
0.921
0.912
0.98
1.98
1.09
1.08
2.89
1.24
3.11
3.11
0.98
f1 (ppm)


HO
OH
135.872
125.499
68.876
63.466
35.544
28.322
22.922
22.729
21.380
14.171
f1 (ppm)

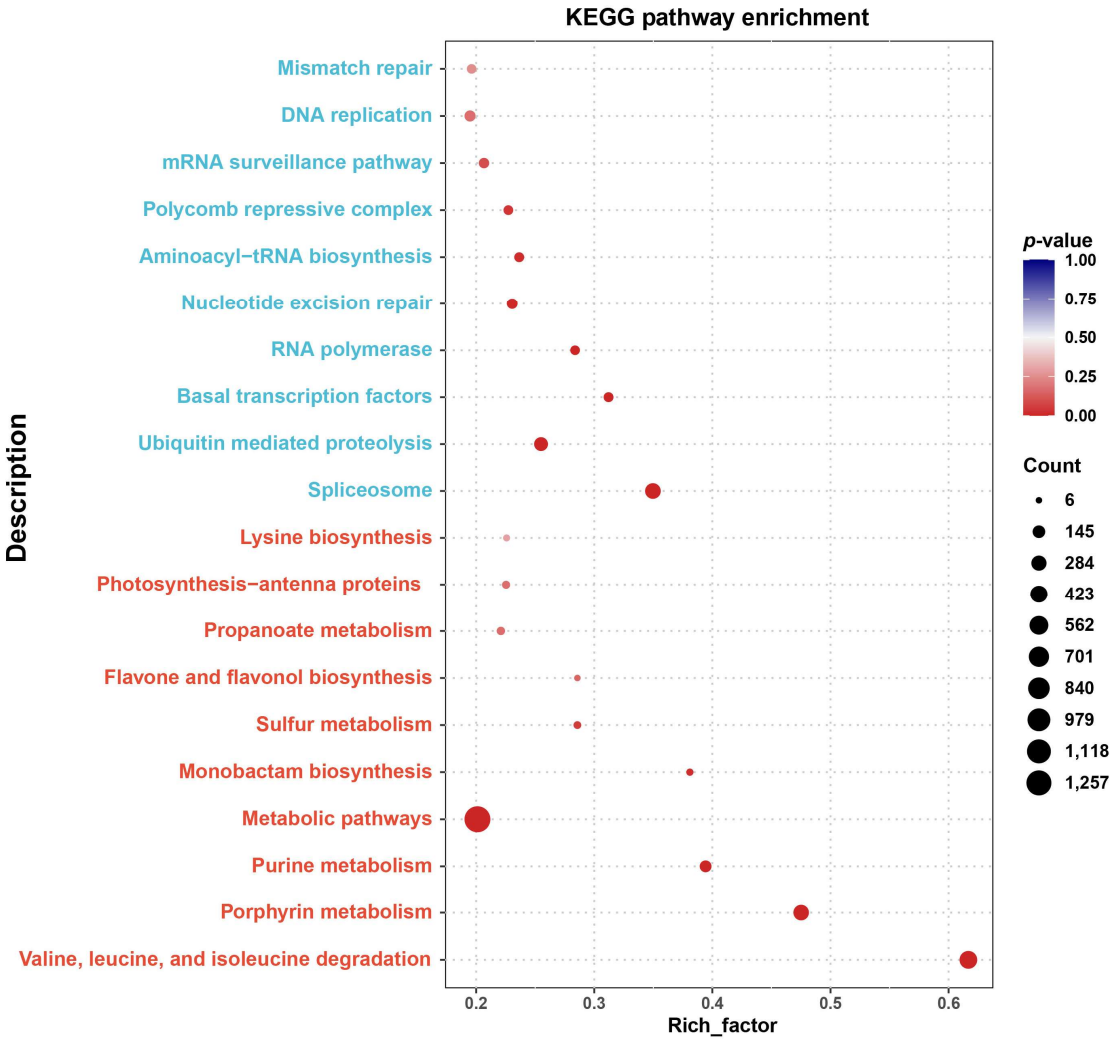

5.66 5.66 5.66 5.66 5.65 5.65 5.65 5.65 5.43 5.43 5.42 5.41 5.25 5.25 5.24 5.24 5.23 4.91 4.91 4.91 4.90 4.90 4.90 2.98 2.97 2.88 2.87 2.85 2.84 2.24 2.24 2.21 2.21 2.17 2.15 2.14 2.14 2.10 2.09 2.08 2.08 2.04 1.73 1.71 1.41 1.40 1.26 1.14 1.00 0.99 0.98
Jasmolin I (21)
1H-NMR, CDCl3
600 MHz, 298 K
1.00 1.14 1.14 1.05 2.40 1.23 1.12 2.40 1.12 3.20 3.14 3.19 1.05 3.38 3.16 3.50
f1 (ppm)


204.17 172.49 164.99 142.91 136.05 133.32 124.06 120.96 73.14 42.23 34.71 33.09 29.27 25.72 22.26 21.36 20.75 20.56 18.65 14.24 14.20
Jasmolin I (21)
13C-NMR, CDCl3
150 MHz, 298 K
f1 (ppm)

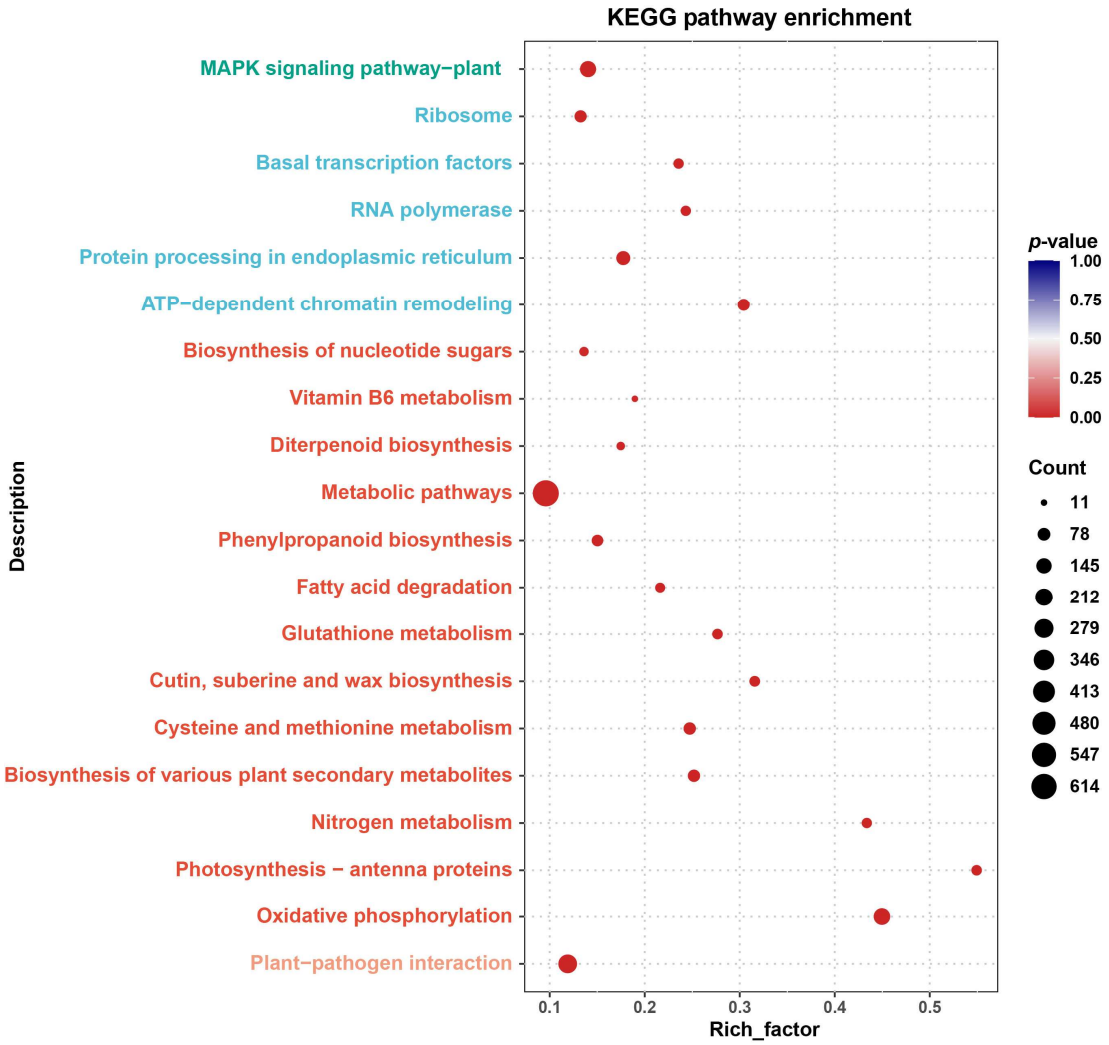
Supplementary Figures



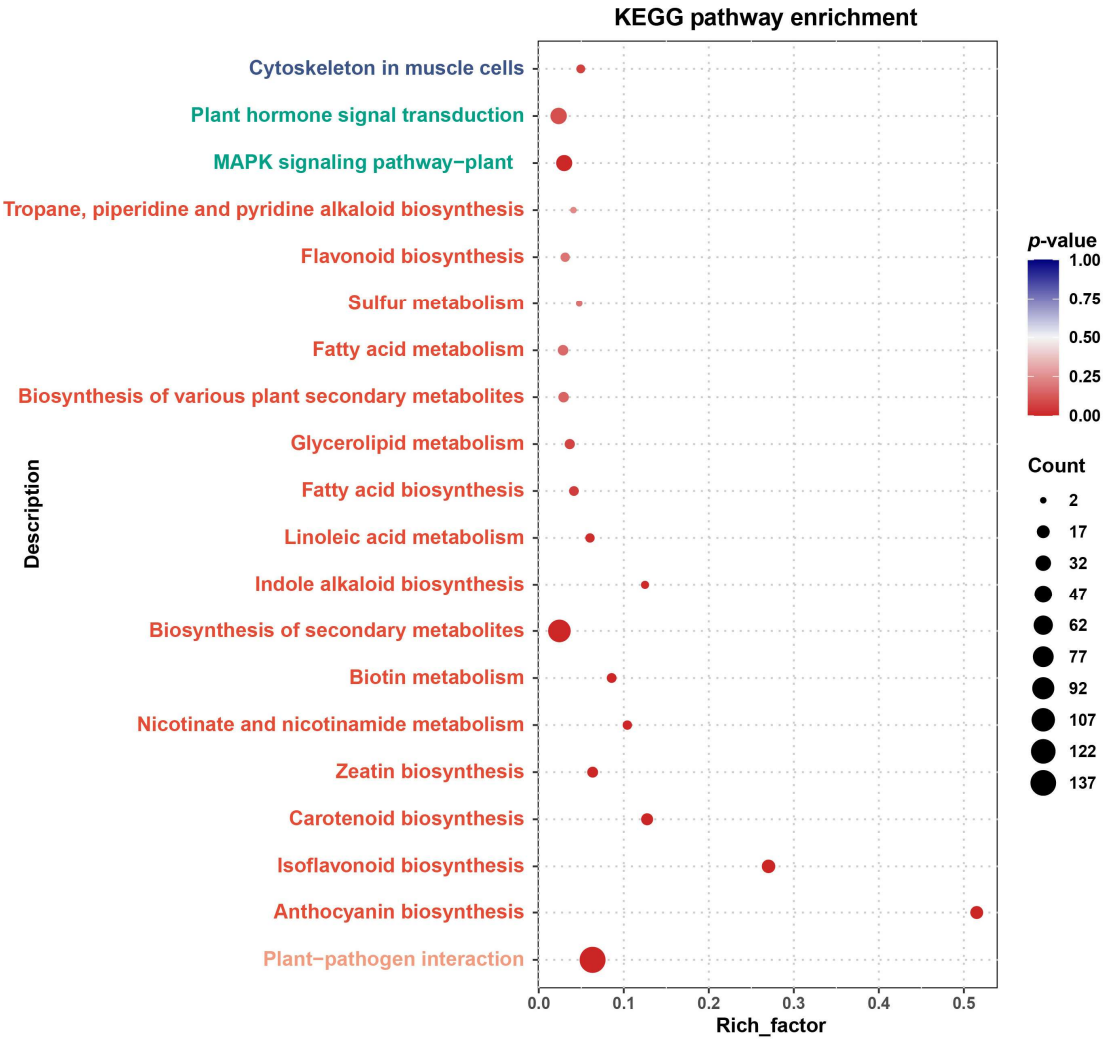
Supplementary Fig. 1 Comparison of ortholog gene families among 12 plant species. The *y*-axis displays the number of genes in homologous gene families in the genome, while the *x*-axis shows the names of 12 plant species. Single-copy orthologs: the number of single-copy homologous genes in the current species that belong to gene families shared by all species; Multiple-copy orthologs: the number of multiple-copy homologous genes in the current species that belong to gene families shared by all species; Unique paralogs: the number of homologous genes in gene families that are exclusive to the current species; Other orthologs: the number of homologous genes contained in other types of gene families; Un-clustered genes: the number of genes that are not clustered into any gene family.



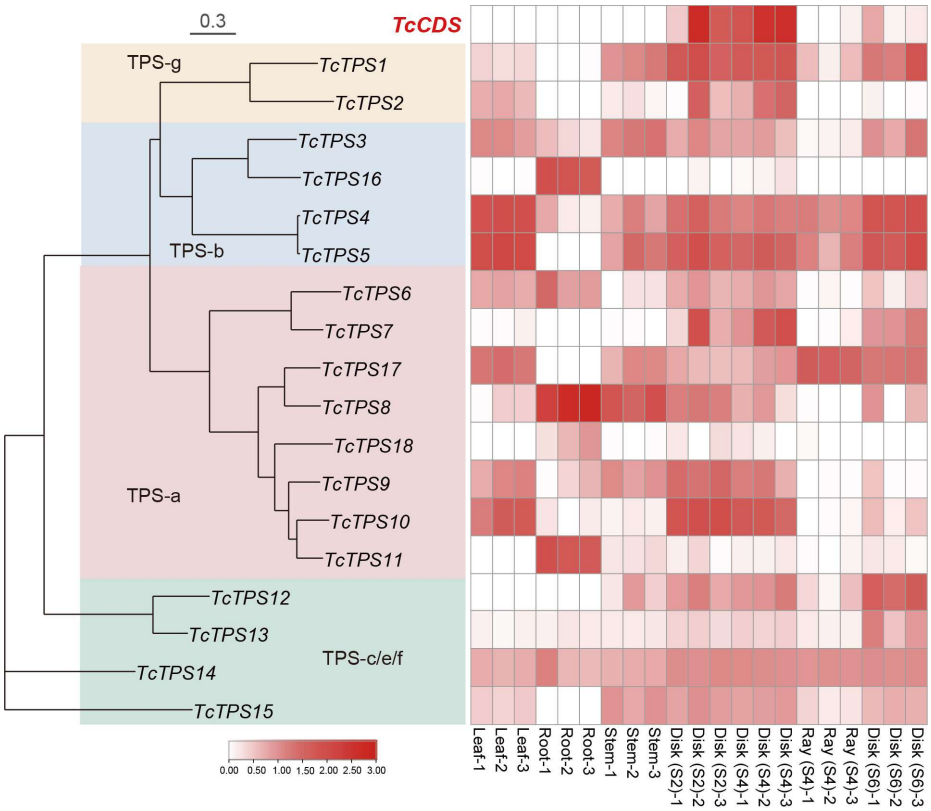
Supplementary Fig. 2 KEGG enrichment analysis of the species-specific genes of *T. cinerariifolium*. The vertical axis represents KEGG pathway information; the horizontal axis represents the corresponding gene quantity statistics.



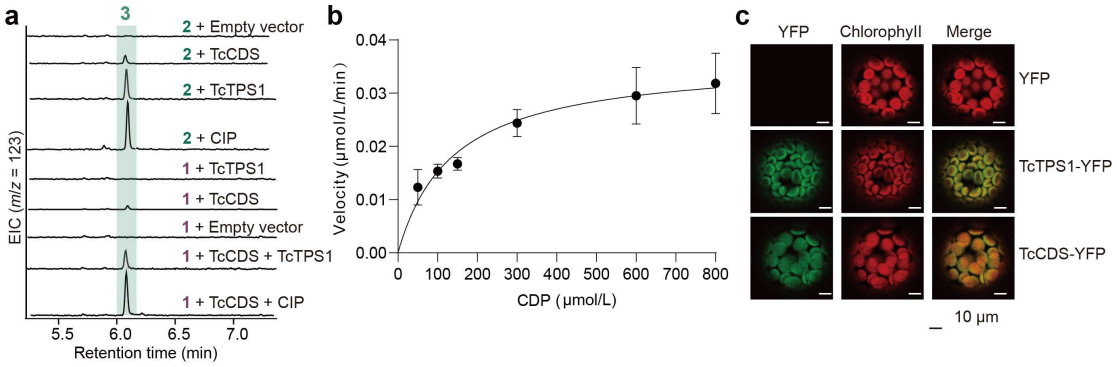
Supplementary Fig. 3 KEGG enrichment analysis of the expanded gene family in the *T. cinerariifolium* genome.



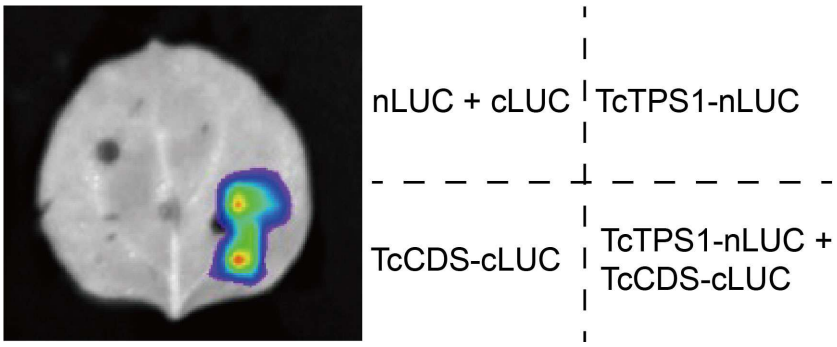
Supplementary Fig. 4 KEGG enrichment analysis of the contracted gene family in the *T. cinerariifolium* genome.



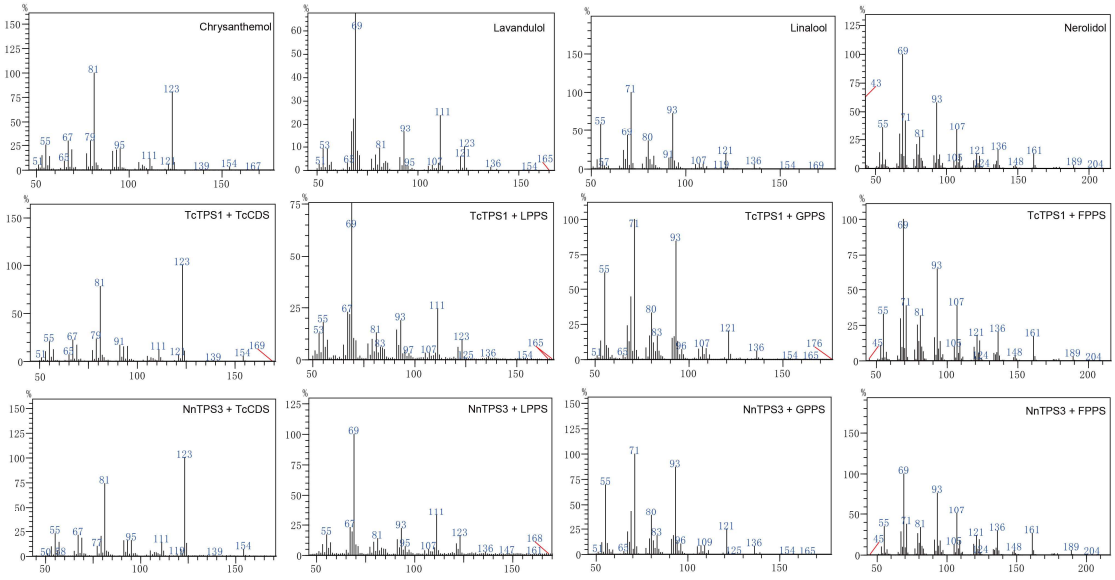
Supplementary Fig. 5 The heatmap of candidate terpene synthases (*TPSs*) genes identified in *T. cinerariifolium*. The candidate *TcTPS* genes were identified through hmm search (PF01397 and PF03936) and correlation analysis on the genome data and transcriptome data of *T. cinerariifolium*. The FPKM (fragments per kilobase of transcript per million fragments mapped) values of the identified *TPS* genes are shown in **Supplementary Table 8**. The expression levels of each gene in the figure are normalized by $\log_2$.



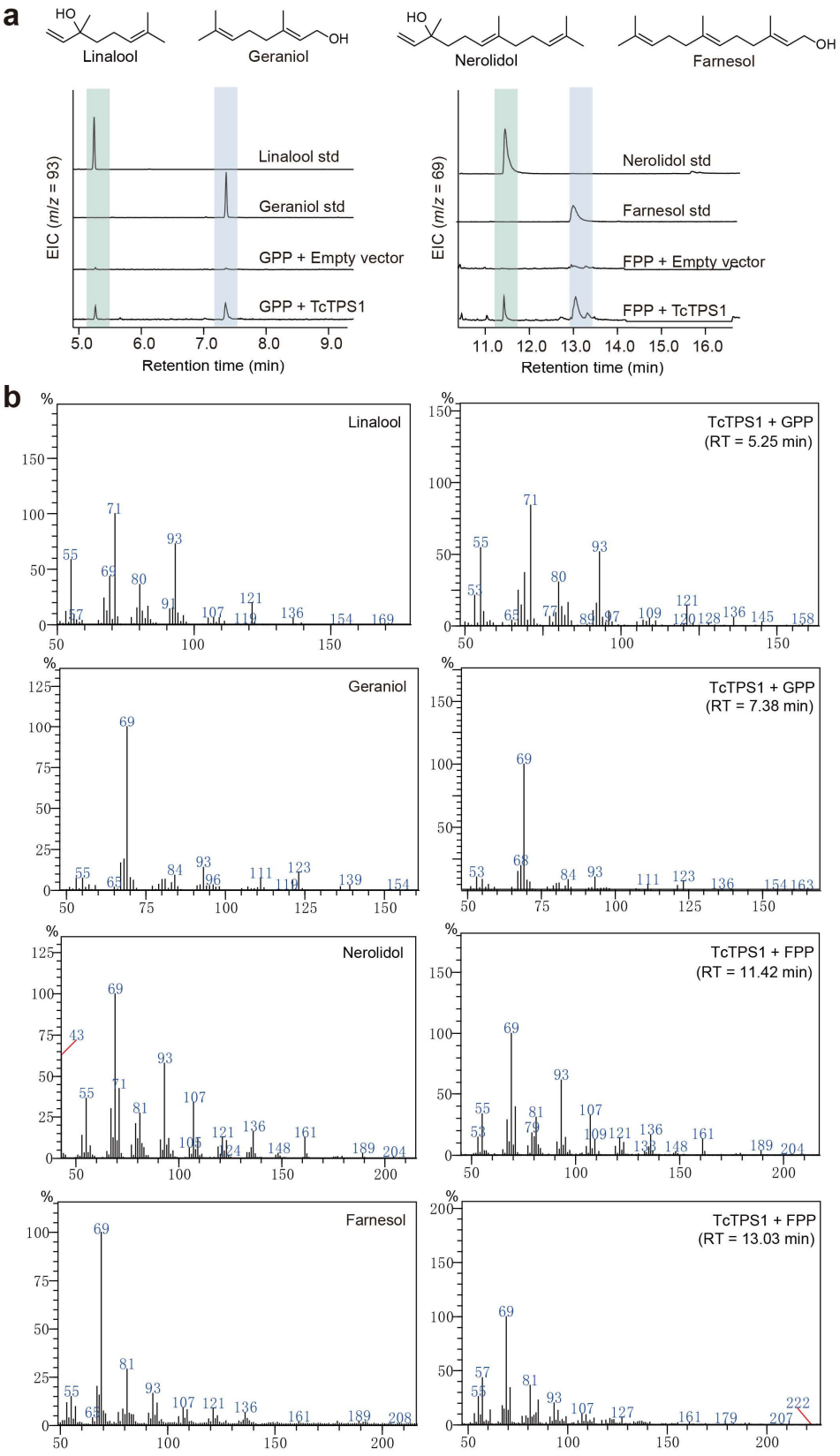
Supplementary Fig. 6 Biochemical and cellular characterization of TcCDS and TcTPS1. **a** GC-MS analysis of purified recombinant TcCDS and TcTPS1 proteins incubated with DMAPP (**1**) or CDP (**2**) in enzymatic reactions. Empty vector served as the negative control; CIP (alkaline phosphatase, calf intestinal) served as the positive control. **b** Michaelis-Menten curves of TcTPS1 for CDP dephosphorylation reaction. Error bars represent statistical differences (SDs) of three biological replicates. Steady-state kinetics analysis of TcTPS1 for CDP dephosphorylation activities reveal a catalytic efficiency (k_{cat}/K_m) of 3.15 $mM^{-1}s^{-1}$, with K_m = 138.4 μmol/L and k_{cat} = 0.44 s^{-1}. In comparison, TcCDS exhibits a catalytic efficiency of 1.01 $mM^{-1}s^{-1}$(K_m = 196 μmol/L, k_{cat} = 0.19 s^{-1})[60]. **c** Subcellular localization of TcTPS1 and TcCDS in *N. benthamiana* leaves.



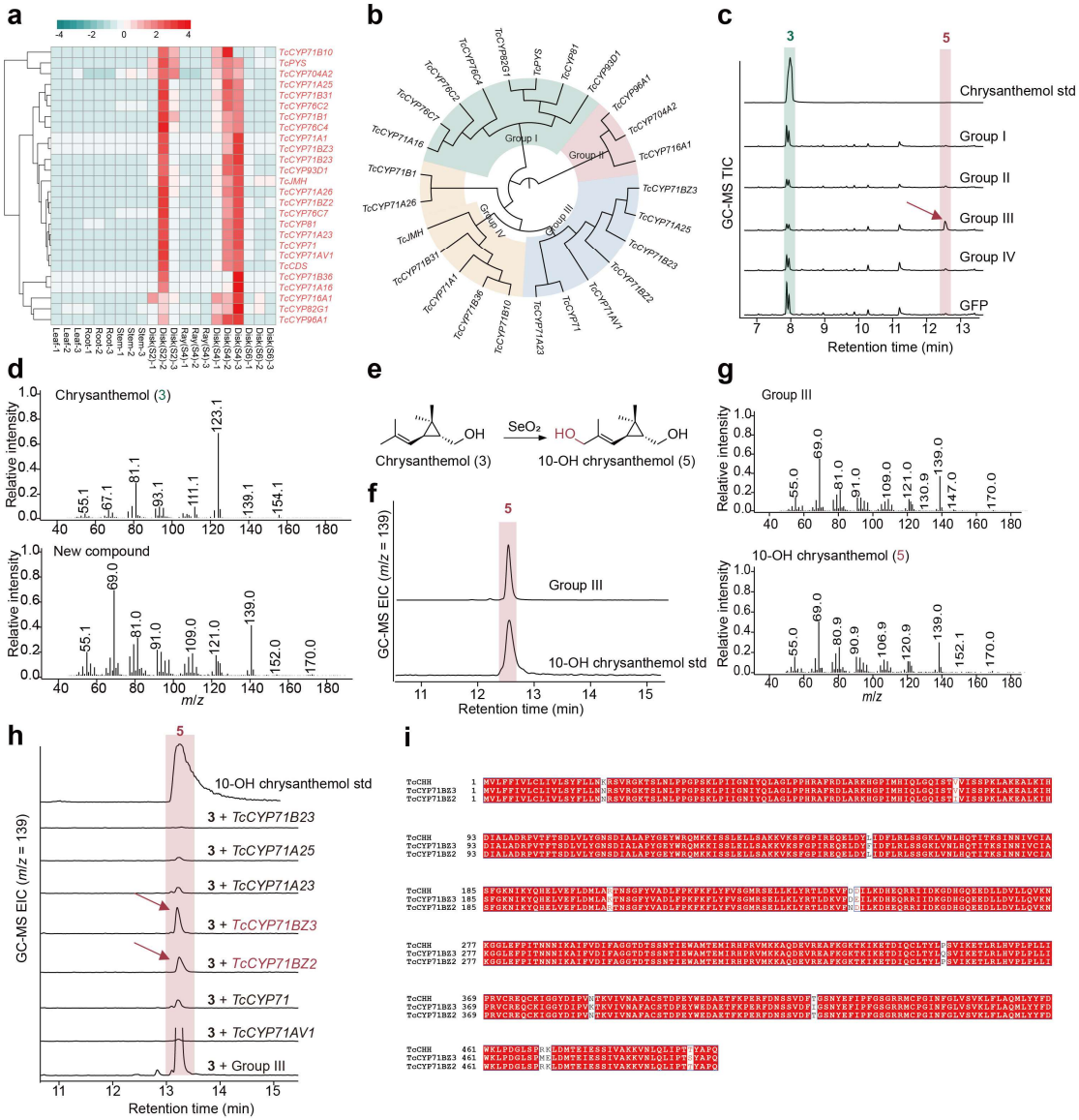
Supplementary Fig. 7 Transient dual luciferase (dual-LUC) assay showing that TcCDS interacts with TcTPS1 in *N. benthamiana* leaves. The N- or C-terminal fragment of LUC (nLUC or cLUC, respectively) was fused with the indicated proteins, and the indicated pairs were co-expressed in *N. benthamiana* leaf cells.



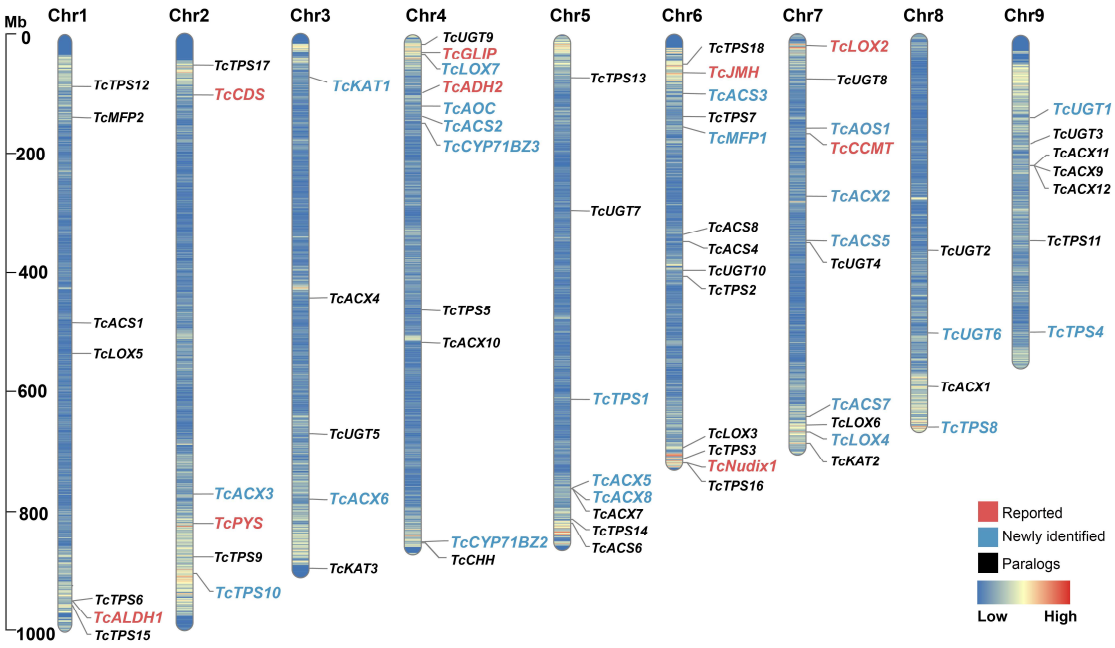
Supplementary Fig. 8 The mass spectra comparison between products of TcTPS1 or NnTPS3 co-expressed with precursor synthases and commercial standards.



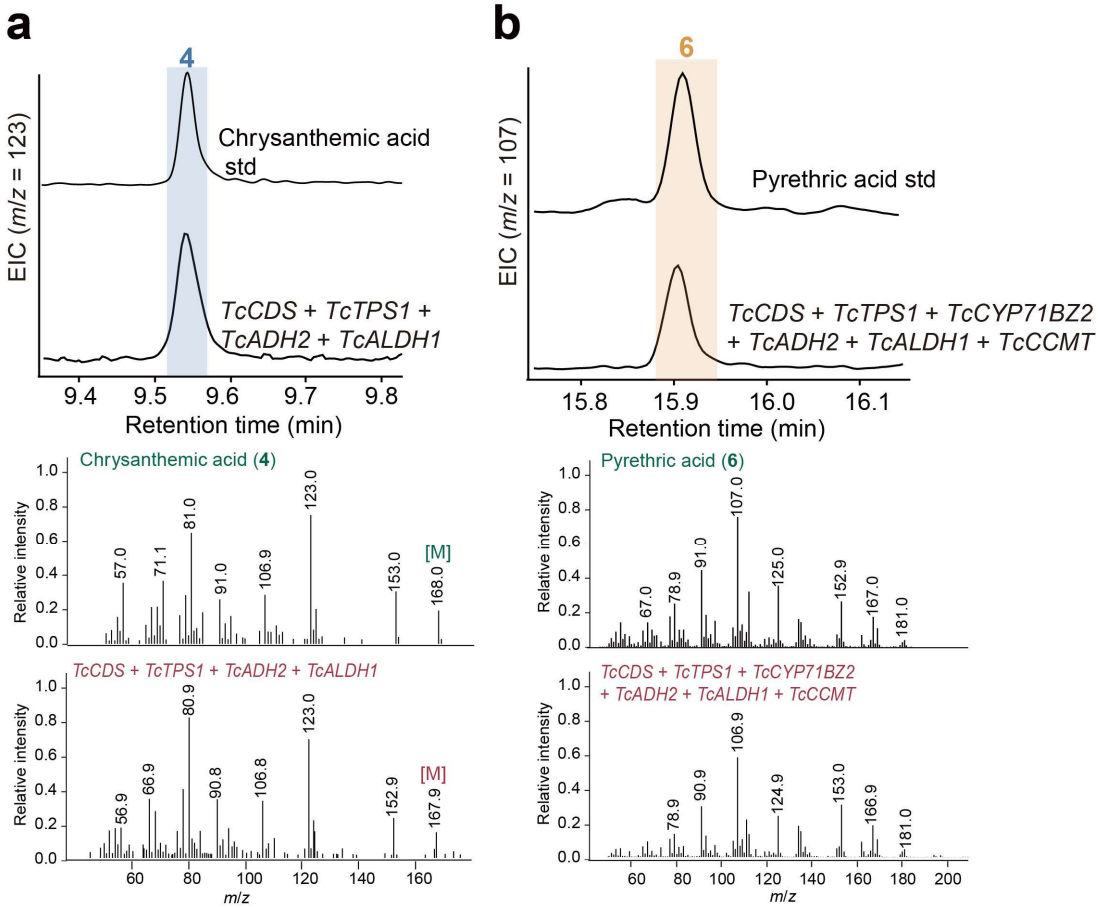
Supplementary Fig. 9 GC-MS analysis of purified TcTPS1 protein incubated with GPP or FPP in enzymatic reactions. **a** EIC chromatogram detected by GC-MS. **b** The mass spectra comparison between products of TcTPS1 and commercial standards. Empty vector served as the negative control.



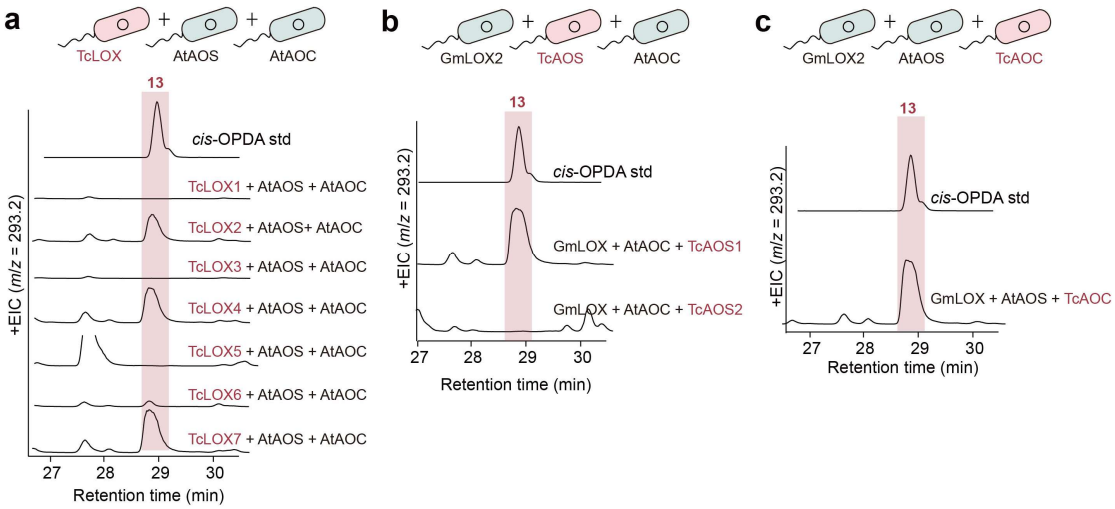
Supplementary Fig. 10 Discovery of genes related to the biosynthesis of 10-OH chrysanthemol. a The heatmap of 25 candidate *CYP450* genes correlated to *TcCDS* in the *T. cinerariifolium*. The FPKM (fragments per kilobase of transcript per million fragments mapped) values of the identified *TcCYP450* genes are shown in **Supplementary Table 9**. The expression levels of each gene in the figure are normalized by log2. **b** The phylogenetic tree of 25 *TcCYP450* genes was constructed, and they were divided into 4 groups based on the branches. **c** Transient expression of these 4 group genes in *N. benthamiana* with co-infiltration of chrysanthemol (**3**). GFP served as the negative control. **d** Comparison of the MS spectra of chrysanthemol (**3**) and the new compound. **e** The synthesis route map of 10-OH chrysanthemol (**5**). **f** GC-MS chromatograms of group III product and 10-OH chrysanthemol (**5**). **g** The corresponding MS spectra of group III product and 10-OH chrysanthemol (**5**). **h** Transient expression of 7 genes in group III in *N. benthamiana* with co-infiltration of chrysanthemol (**3**). The product indicated by the red arrow is 10-OH chrysanthemol (**5**). **i** The amino acid sequence alignments among TcCHH[61], TcCYPBZ2, and TcCYP71BZ3.



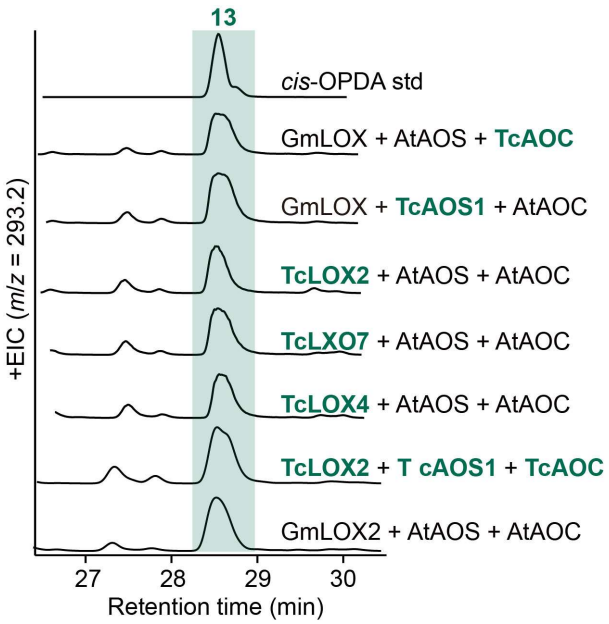
Supplementary Fig. 11 Chromosome position of genes involved in pyrethrin biosynthesis in the *T. cinerariifolium* genome. The color of chromosomes indicates the density of genes. The genes marked in red, blue and black are respectively those that have been reported previously, and the newly identified genes in this study, as well as their homologous genes.



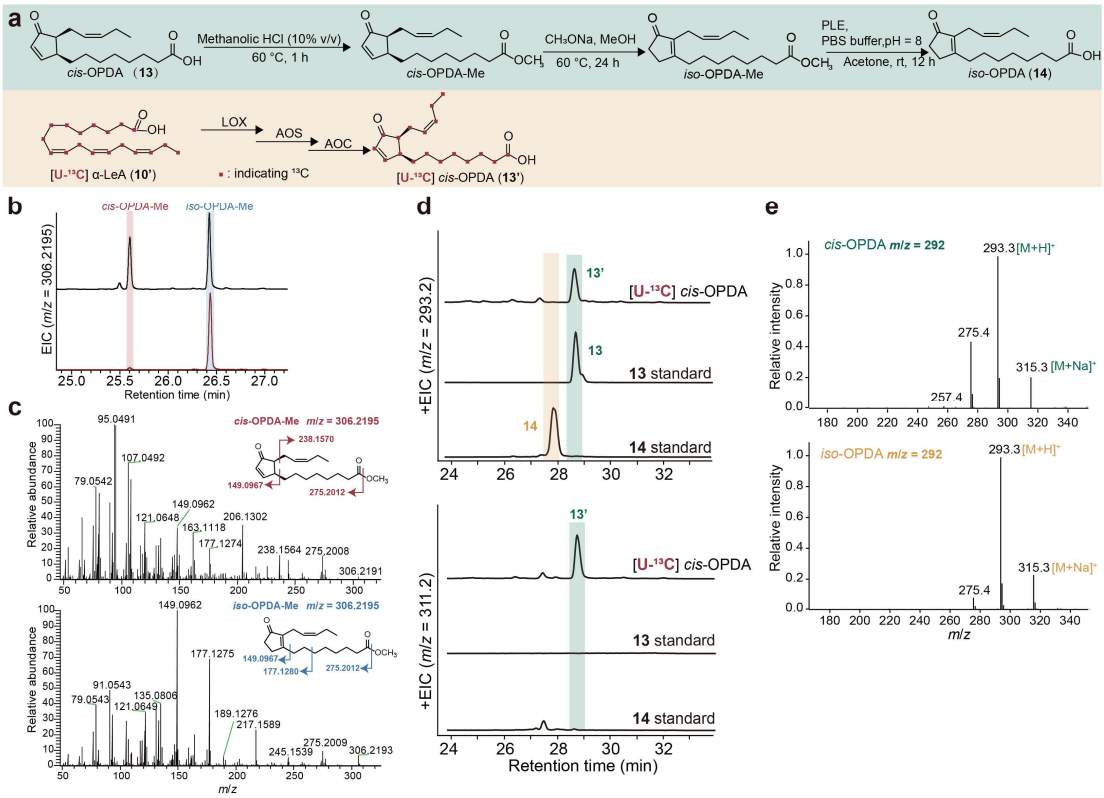
Supplementary Fig. 12 ***De novo*** **biosynthesis of the acid moieties in** ***N. benthamiana*****. a-b** Production of chrysanthemic acid (**4**) and pyrethric acid (**6**) by transient co-expression of the indicated gene combinations in *N. benthamiana* leaves. For each panel, the upper traces show GC-MS EICs of the products co-eluting with authentic standards; the lower panels show the corresponding mass spectra, confirming identity by match with the standard spectra.



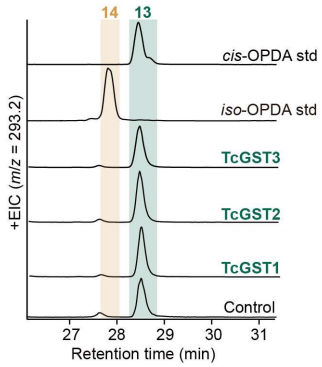
Supplementary Fig. 13 Functional characterization of *TcLOX*, *TcAOS*, *TcAOC* genes involved in *cis*-OPDA biosynthesis in *T. cinerariifolium*. a–c The screening process for candidate TcLOX, TcAOS, TcAOC enzymes in crude *E. coli* proteins with α-LeA (**10**) as the substrate. The enzymes marked in red are those from *T. cinerariifolium*. The product peak is labeled as **13**, which is consistent with the peak time (28.95 min) of the *cis*-OPDA standard (+EIC *m*/*z* $[M+H]^+$ = 293.2).



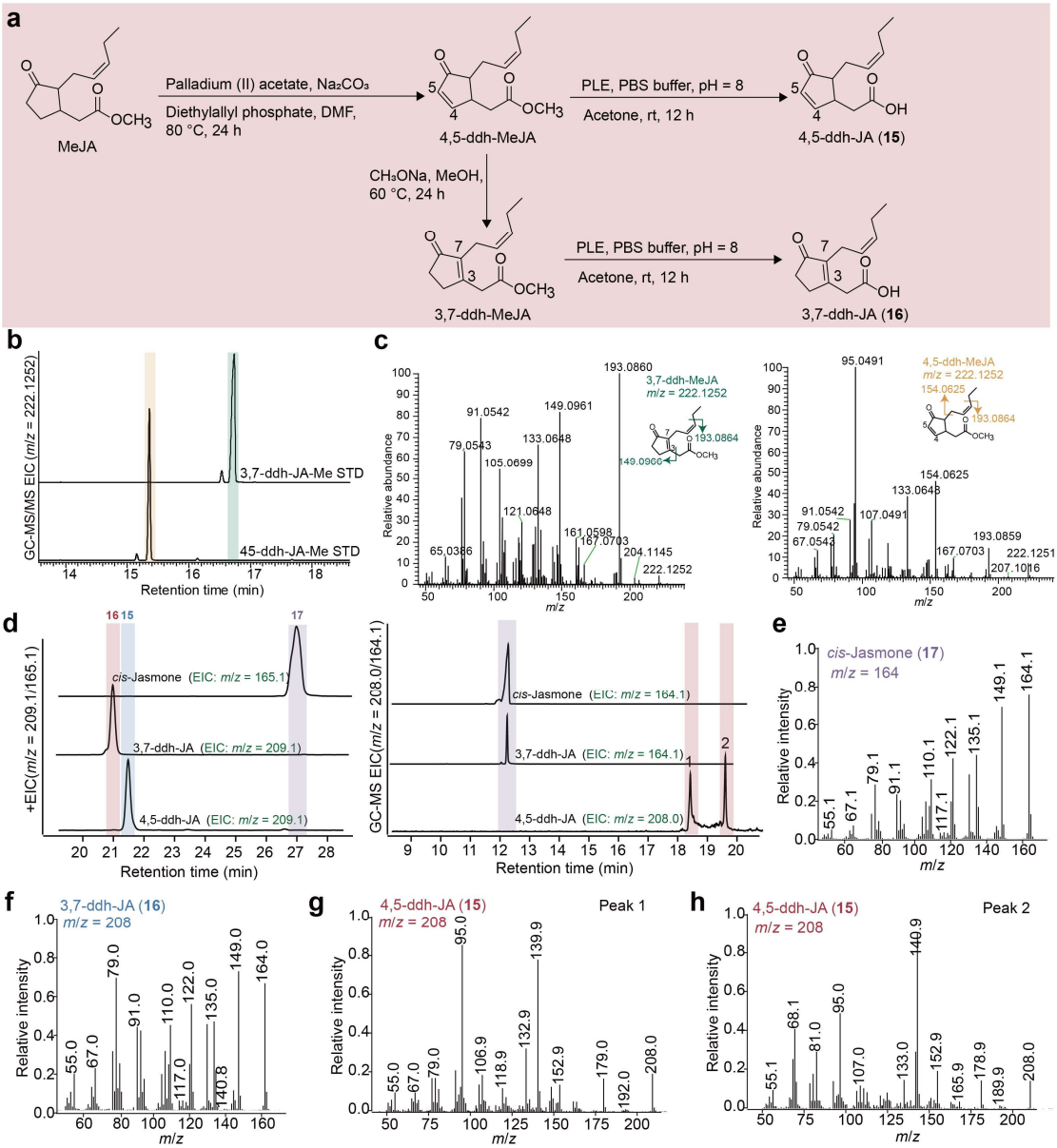
Supplementary Fig. 14 LC-MS analysis of recombinant LOX, AOS, and AOC proteins incubated with α-LeA (**10**), resulting in the production of *cis*-OPDA (**13**), (+EIC *m*/*z* $[M+H]^+$ = 293.2).



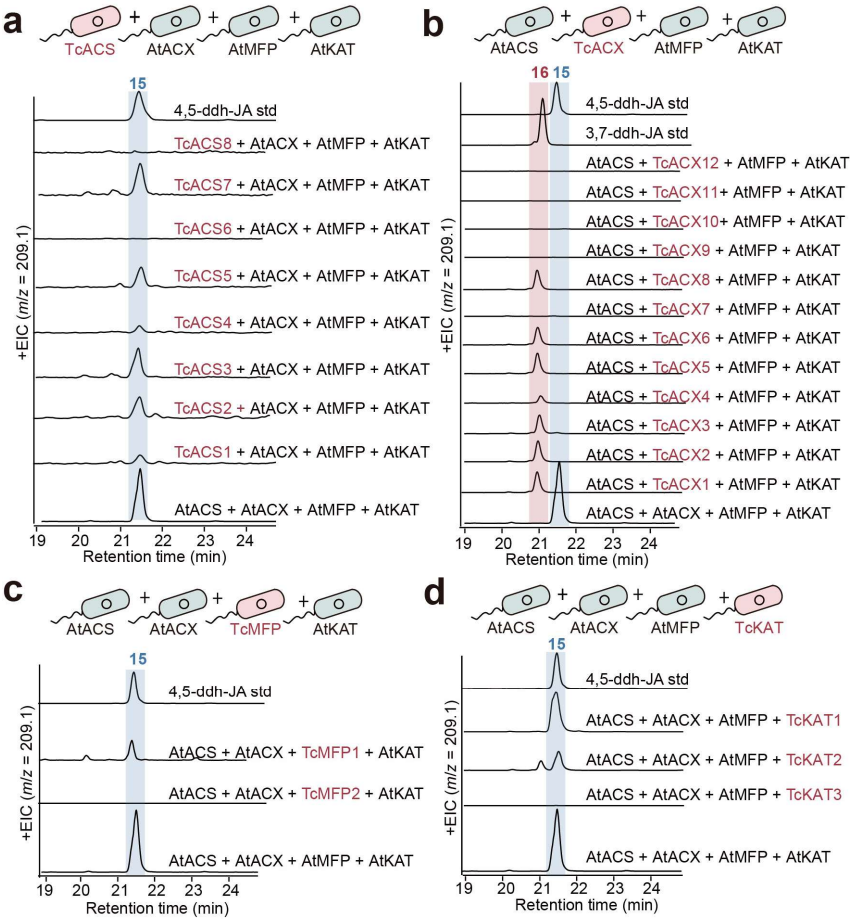
Supplementary Fig. 15 Synthesis of *iso*-OPDA and [U-^{13}C] *cis*-OPDA. **a** The synthesis route map of iso-OPDA (**14**) and [U-^{13}C] *cis*-OPDA (**13'**). **b** GC-MS/MS chromatograms of *cis*-OPDA-Me and *iso*-OPDA-Me. EIC for both (*m*/*z*, $[M]^+ = 306.2195$). **c** The MS spectra of *cis*-OPDAMe and *iso*-OPDAMe respectively. The chemical structures of *cis*-OPDAMe and *iso*-OPDAMe, along with their corresponding cleavage patterns were shown. **d** LC-MS chromatograms of *cis*-OPDA (**13**), *iso*-OPDA (**14**) (+EIC, *m*/*z* $[M+H]^+ = 293.2$), and [U-^{13}C] *cis*-OPDA (**13'**) (+EIC, *m*/*z* $[M+H]^+ = 311.2$). No [U-^{13}C] *iso*-OPDA (**14'**) was detected in the synthesized [U-^{13}C] *cis*-OPDA (**13'**). **e** The MS spectra of *cis*-OPDA (**13**) and *iso*-OPDA (**14**).



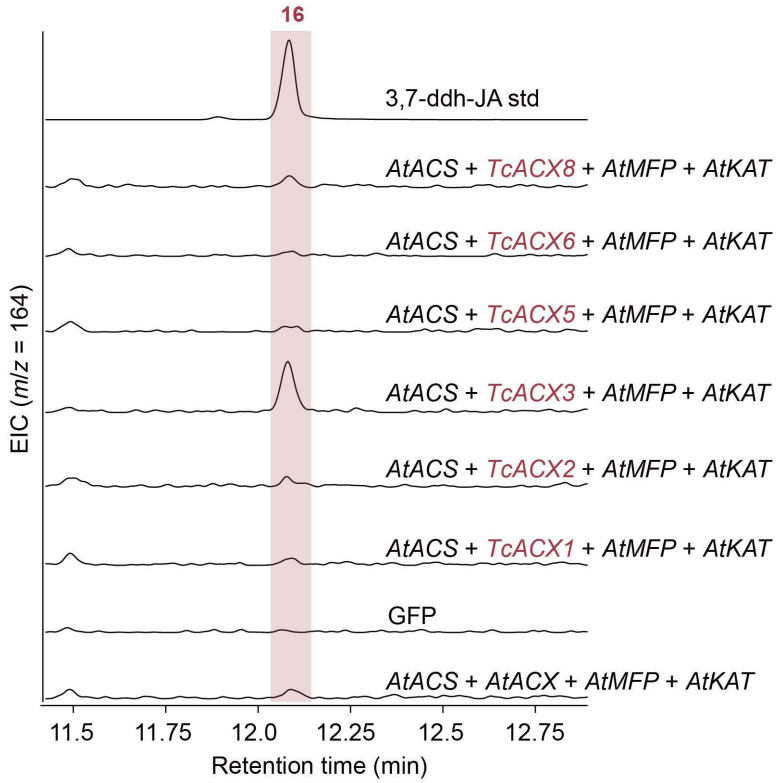
Supplementary Fig. 16 *In vitro* screening of glutathione transferases (GSTs) for isomerization of *cis*-OPDA. The reaction mixtures (100 μL) contain 0.1 mmol/L Tris-HCl (pH 7.5), 5 mmol/L $MgCl_2$, 1 mmol/L reduced glutathione (GSH), 1 mmol/L *cis*-OPDA (**13**) substrate. Assays were carried out at 28 ℃ for 2.5 h and were stopped by the addition of 100 μL ethyl acetate. The reaction product was analyzed using LC-MS and extracted ion chromatograms (EIC) was shown (+EIC *m*/*z* $[M+H]^+$ =293.2).



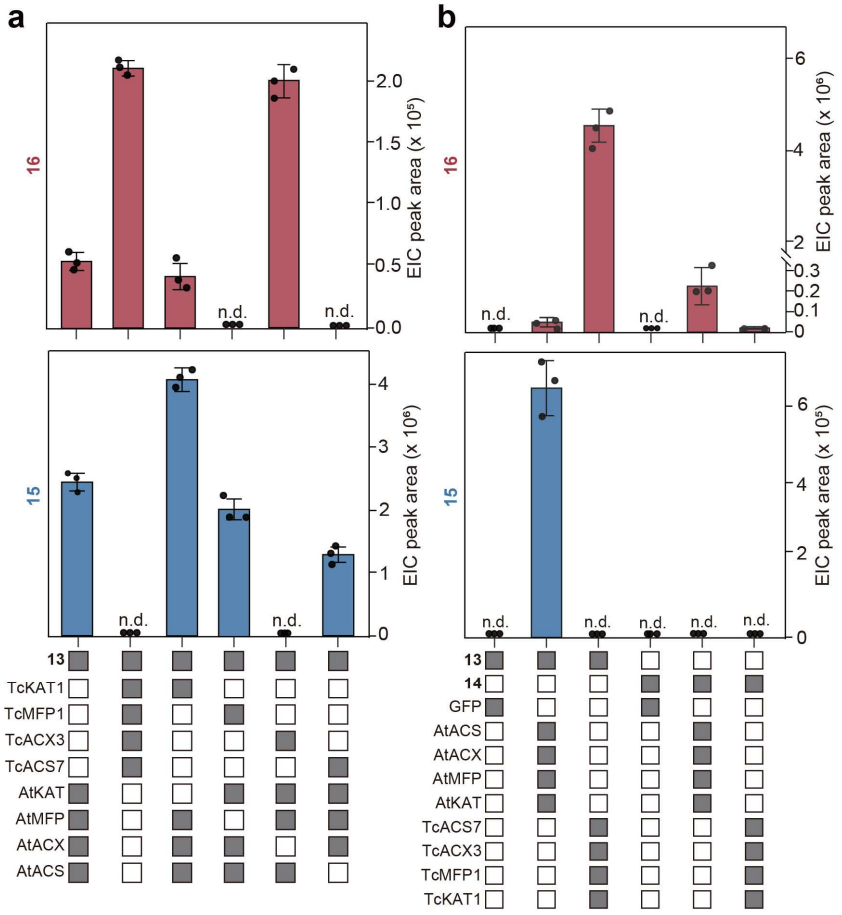
Supplementary Fig. 17 Synthesis of 4,5-ddh-JA and 3,7-ddh-JA. a The synthesis route map of 4,5-ddh-JA (**15**) and 3,7-ddh-JA (**16**). **b** GC-MS chromatogram of 3,7-ddh-MeJA and 4,5-ddh-MeJA ([M] ion; *m/z* = 222.1252). **c** The MS spectra of 4,5-ddh-MeJA and 3,7-ddh-MeJA. **d** The LC-MS and GC-MS chromatograms of 3,7-ddh-JA (**16**), 4,5-ddh-JA (**15**) and *cis*-jasmone (**17**), respectively. **e-h** MS spectra of *cis*-jasmone (**17**), 3,7-ddh-JA (**16**) and 4,5-ddh-JA (**15**). 3,7-ddh-JA (**16**) decarboxylates under high-temperature conditions to form *cis*-jasmone (**17**) during GC-MS detection. Peak 1 and 2 are presumed to be two isomers of 4,5-ddh-JA (**15**) under high-temperature conditions.



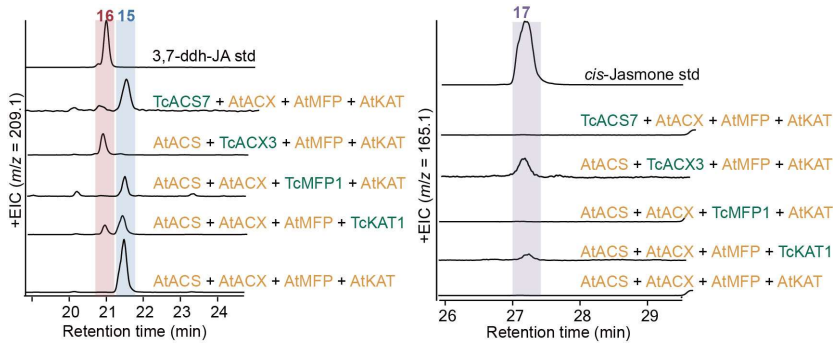
Supplementary Fig. 18 Functional characterization of candidate *ACS*/*ACX*/*MFP*/*KAT* genes involved in β-oxidation in *T. cinerariifolium*. a–**d** The screening process for candidate TcACS, TcACX, TcMFP and TcKAT enzymes in crude *E. coli* proteins with *cis*-OPDA (**13**) as the substrate. The enzymes marked in red are those from *T. cinerariifolium*.



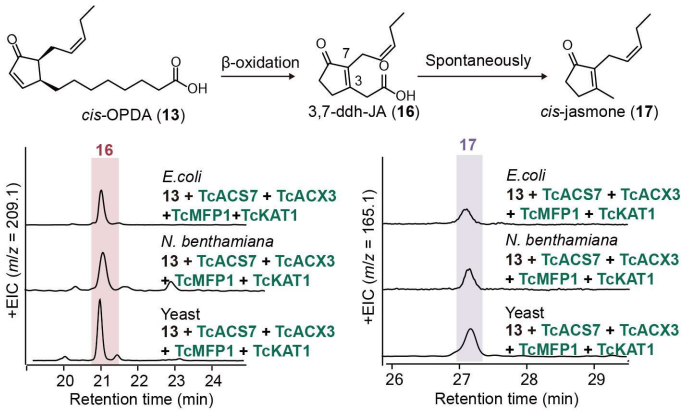
Supplementary Fig. 19 Functional characterization of *TcACXs* genes in *N. benthamiana*. Transient expression of *TcACXs* along with *AtACS*, *ATMFP* and *ATKAT* in *N. benthamiana* with co-infiltration of *cis*-OPDA (**13**). GFP served as the negative control. EIC for 3,7-ddh-JA (**16**) (*m*/*z* = 164), was shown by the GC-MS.



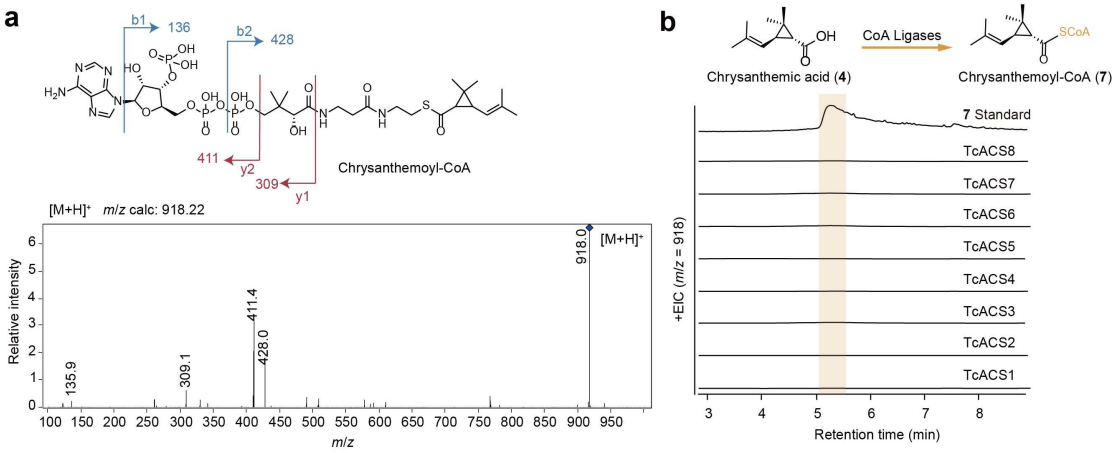
Supplementary Fig. 20 Functional characterization of β-oxidation genes indentified from *T. cinerariifolium*. **a** Comparison of EIC peak areas for compounds produced by in *vitro* enzymatic assays characterizing pyrethrum β-oxidation genes. **b** Comparison of EIC peak areas for compounds produced by transient expression of pyrethrum and Arabidopsis derived β-oxidation genes in *N. benthamiana*. Numbers **15** and **16** denote products 4,5-ddh-JA and 3,7-ddh-JA, respectively. Co-expressed genes in each experiment are indicated by gray boxes, with numbers **13** and **14** representing the reaction substrates *cis*-OPDA and *iso*-OPDA respectively. Data represent means of three biological replicates; error bars denote standard deviation (SD).



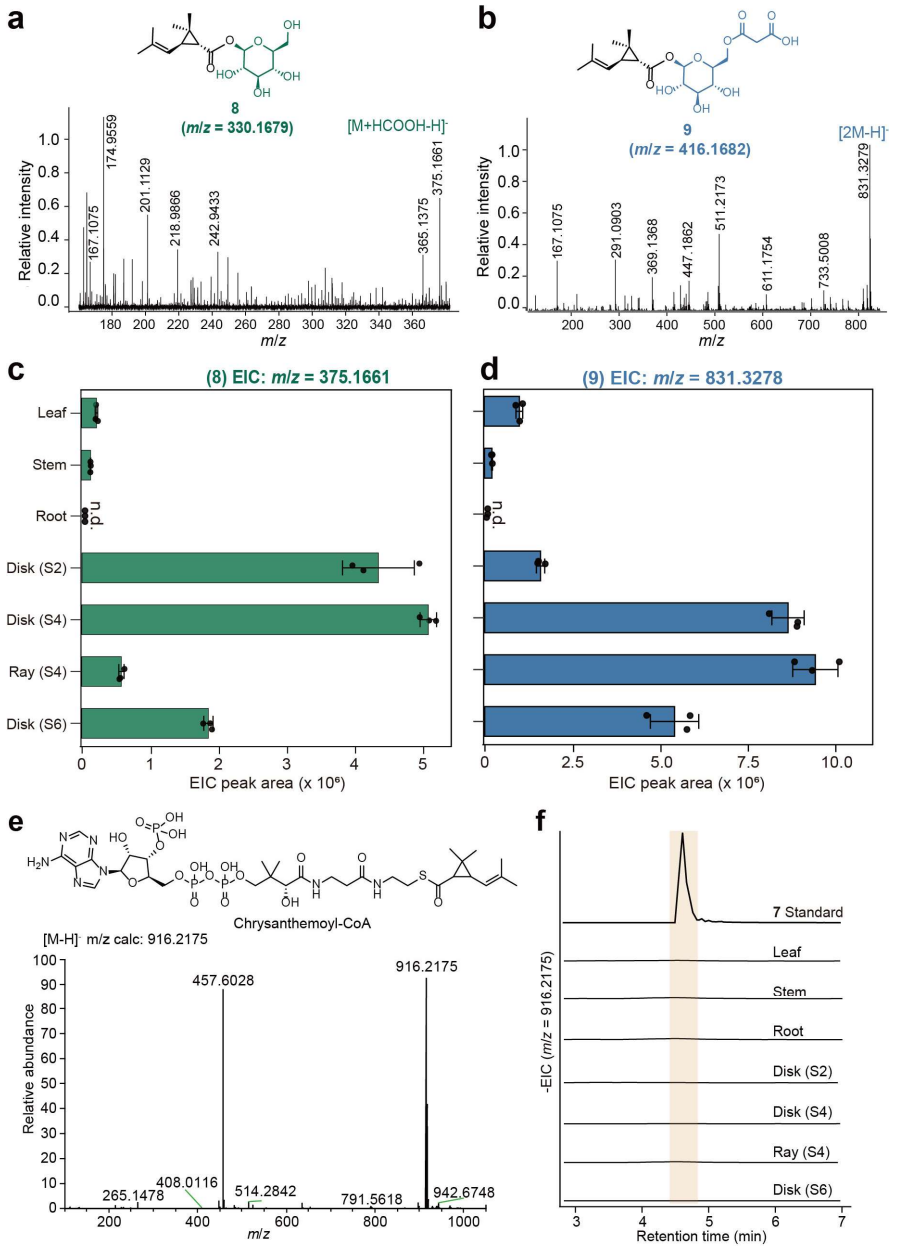
Supplementary Fig. 21 Detection of spontaneous decarboxylation of 3,7-ddh-JA. Extracted ion chromatograms for 3,7-ddh-JA (**16**) (*m*/*z* $[M+H]^+ = 209.1$, left) and *cis*-jasmone (**17**) (*m*/*z* $[M+H]^+ = 165.1$, right) *in vitro* assays without supplemented decarboxylase.



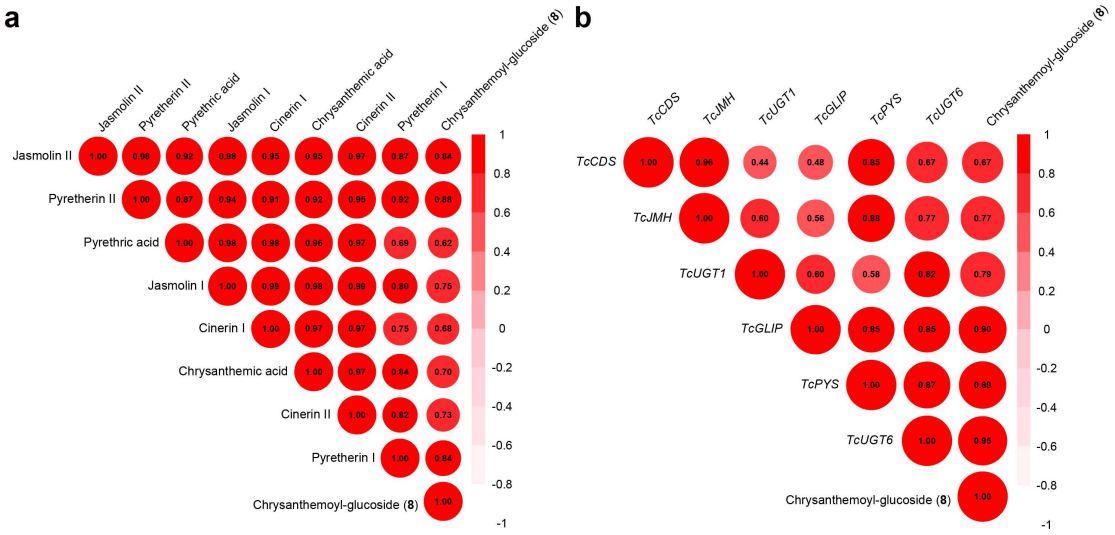
Supplementary Fig. 22 Detection of spontaneous decarboxylation of 3,7-ddh-JA in *E. coli*, *N. benthamiana* and yeast. Extracted ion chromatograms for 3,7-ddh-JA (**16**) (*m*/*z* $[M+H]^+ = 209.1$, left) and *cis*-jasmone (**17**) (*m*/*z* $[M+H]^+ = 165.1$, right) *in vitro* assays, *N. benthamiana*, and yeast systems without supplemented decarboxylase.



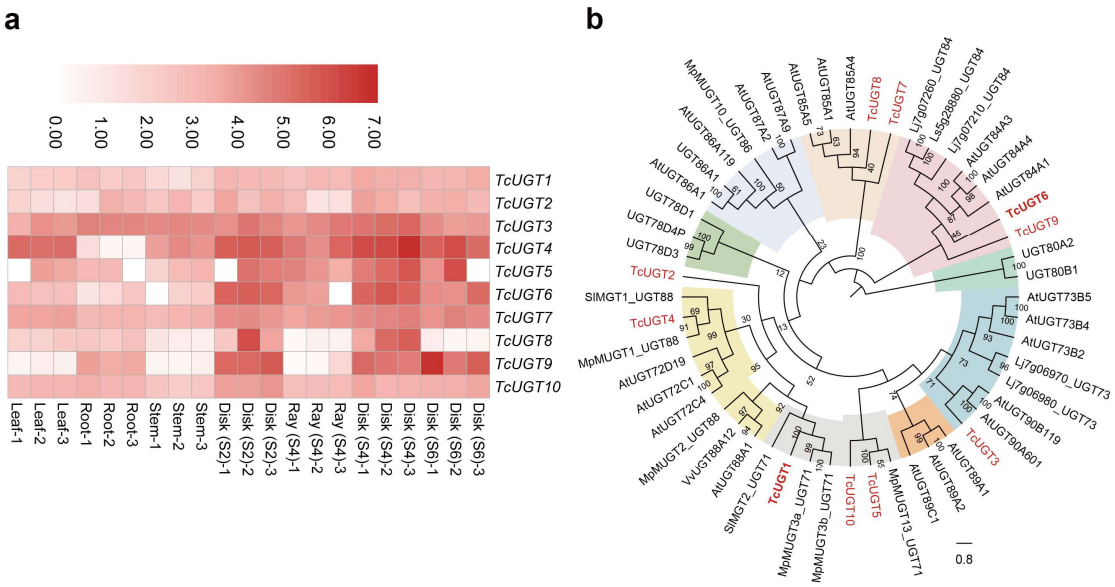
Supplementary Fig. 23 LC-MS analysis of chrysanthemoyl-CoA synthesized in this study. a Predicted fragments and MS/MS spectra were shown. A C18 column (Agilent EclipsePlus C18, 2.1 mm × 100 mm, 1.8 μm) was used. The gradient (solvent A, water + 0.1% formic acid; solvent B, Acetonitrile + 0.1% formic acid) program was set as follows at a flow of 0.3 mL/min: 0–3 min, a linear gradient from 5% to 30% of B; 3–9 min, a linear gradient from 30% to 75% of B, 9–10 min, a linear gradient from 75% to 100% of B, 10–11 min, 100% of B, 11–11.5 min, a linear gradient to 95% of A and equilibration for 7.5 min before starting the next sample. The conditions of the mass spectrometer are set as follows: full-scan mode, scan range: 100–1000 *m/z*; scan time: 500 ms; Fragmentor: 135 V; CAV: 5 V; Ion mode: positive; Capillary voltage: 4000 V; Nozzle voltage: 1500 V; Gas Temperature: 300°C; Sheath Gas Temperature: 250 °C. **b** *In vitro* screening of chrysanthemoyl-CoA (**7**) synthases. The reaction mixtures (100 μL) containing 0.1 mmol/L Tris-HCl (pH 7.5), 5 mmol/L $MgCl_2$, 1 mmol/L CoA, 1 mmol/L ATP, 1 mmol/L Chrysanthemic acid (**4**) substrate. Assays were carried out at 28 ℃ for 2.5 h and were stopped by the addition of 100 μL methanol. The reaction product was analyzed using LC-MS and extracted ion chromatograms (EIC) was shown.



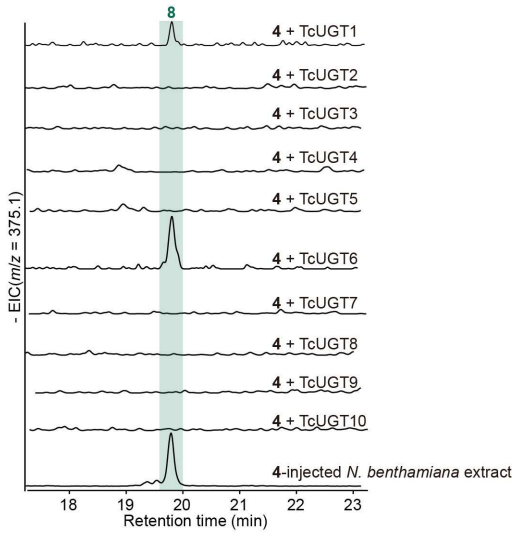
Supplementary Fig. 24 Analysis of chrysanthemic acid glycosylated derivatives and CoA conjugates in *T. cinerariifolium* tissues. a, **b** MS/MS spectra of chrysanthemoyl-glucoside and chrysanthemoyl-malonylglucoside, respectively. **c**, **d** Comparison of EIC peak areas for chrysanthemoyl-glucoside (**8**) ($[M+HCOOH-H]^-$ ion; m/z = 375.1661) and chrysanthemoyl-malonylglucoside (**9**) ($[2M-H]^-$ ion; m/z = 831.3278) across *T. cinerariifolium* tissues (leaf, stem, root, floral disk). **e** MS/MS spectrum of chemically synthesized chrysanthemoyl-CoA. **f** EIC chromatograms of chrysanthemoyl-CoA in *T. cinerariifolium* tissues. Data in **c** and **d** are presented as mean ± SD (n = 3 biological replicates).



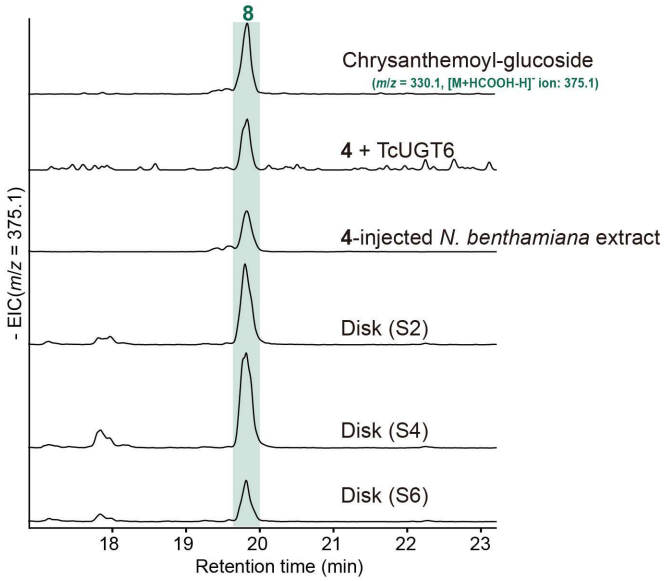
Supplementary Fig. 25 Correlation heatmaps of chrysanthemoyl-glucoside (8) accumulation with pyrethrins, their intermediates, and pathway gene expression in pyrethrum tissues. a Pearson correlation heatmap of chrysanthemoyl-glucoside (**8**) with acid moieties and pyrethrins. **b** Pearson correlation heatmap of chrysanthemoyl-glucoside (**8**) with pyrethrin biosynthesis-related gene expression. The peak area of chrysanthemoyl-glucoside (**8**) was divided by 10^6 to normalize the scale prior to analysis. Color scales indicate correlation strength.



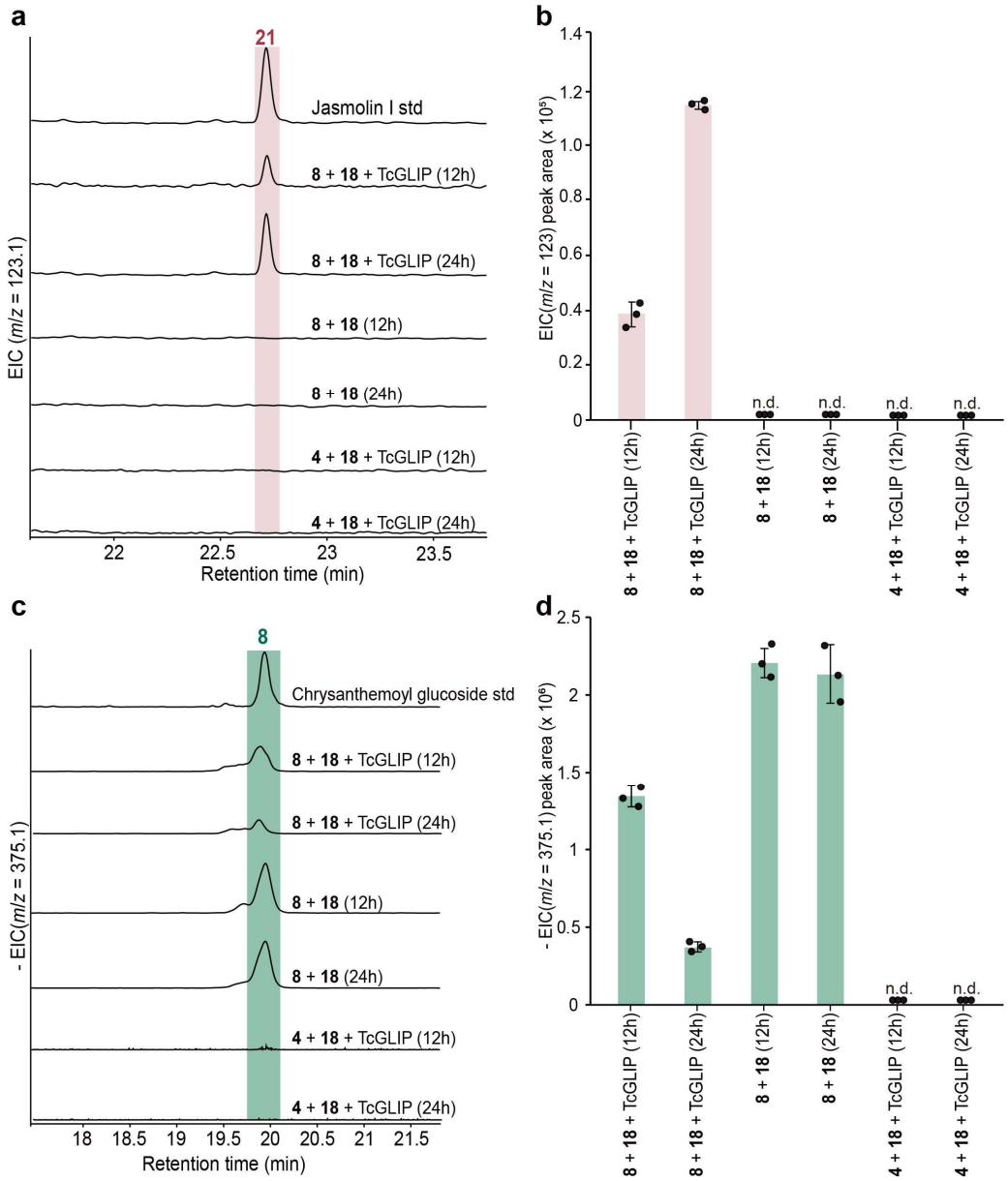
Supplementary Fig. 26 Phylogenetic analysis of selected UGT proteins from *T. cinerariifolium, L. japonicas, Arabidopsis* and others. a The heatmap of candidate *UGT* genes identified in the *T. cinerariifolium*. The candidate *TcUGT* genes were identified through hmm search (PF00201) and correlation analysis on the genome data and transcriptome data of *T. cinerariifolium*. The FPKM values of the identified *TcUGT* genes are shown in **Supplementary Table 10**. The expression levels of each gene in the figure are normalized by $\log_2$. **b** The phylogenetic tree of *UGT* genes, the clades were annotated based on *AtUGT* family classification[62, 63]. Different groups are shown with different colours and *UGTs* from *T. cinerariifolium* are shown in red. Sequences of UGTs used for phylogenetic analysis in this study are shown in **Supplementary Table 11**.



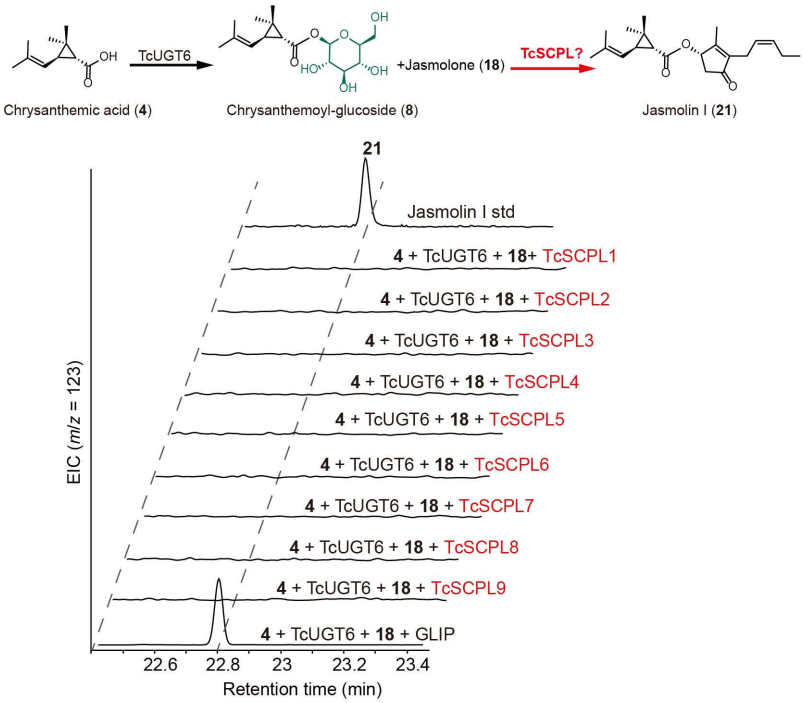
Supplementary Fig. 27 Functional characterization of selected UGT proteins from *T. cinerariifolium in vitro*. The reaction mixtures contain 0.1 mmol/L Tris-HCl (pH 7.5), 5 mmol/L $MgCl_2$, 1 mmol/L UDP-Glucose and 1 mmol/L chrysanthemic acid (**4**) substrate. The reaction product was analyzed using LC-MS and extracted ion chromatograms (EIC) was shown.



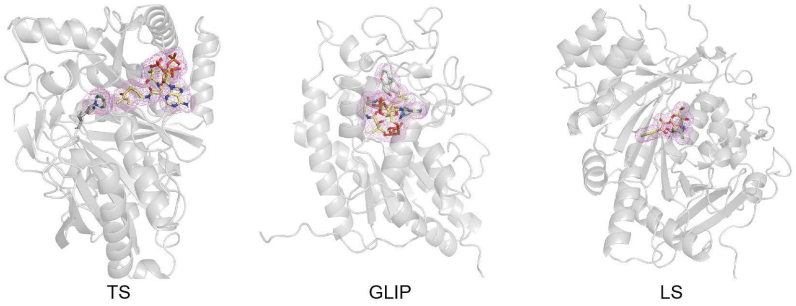
Supplementary Fig. 28 Analysis of chrysanthemoyl-glucoside (8). From top to bottom, the traces show the authentic chrysanthemoyl-glucoside (**8**), the product of the TcUGT6 reaction, the extract from *N. benthamiana* leaves fed with chrysanthemic acid (**4**), and flower-disk extracts from *T. cinerariifolium* at developmental stages S2, S4 and S6. EIC was acquired in negative-ion mode monitoring the $[M+HCOOH-H]^-$ ion (m/z = 375.1).



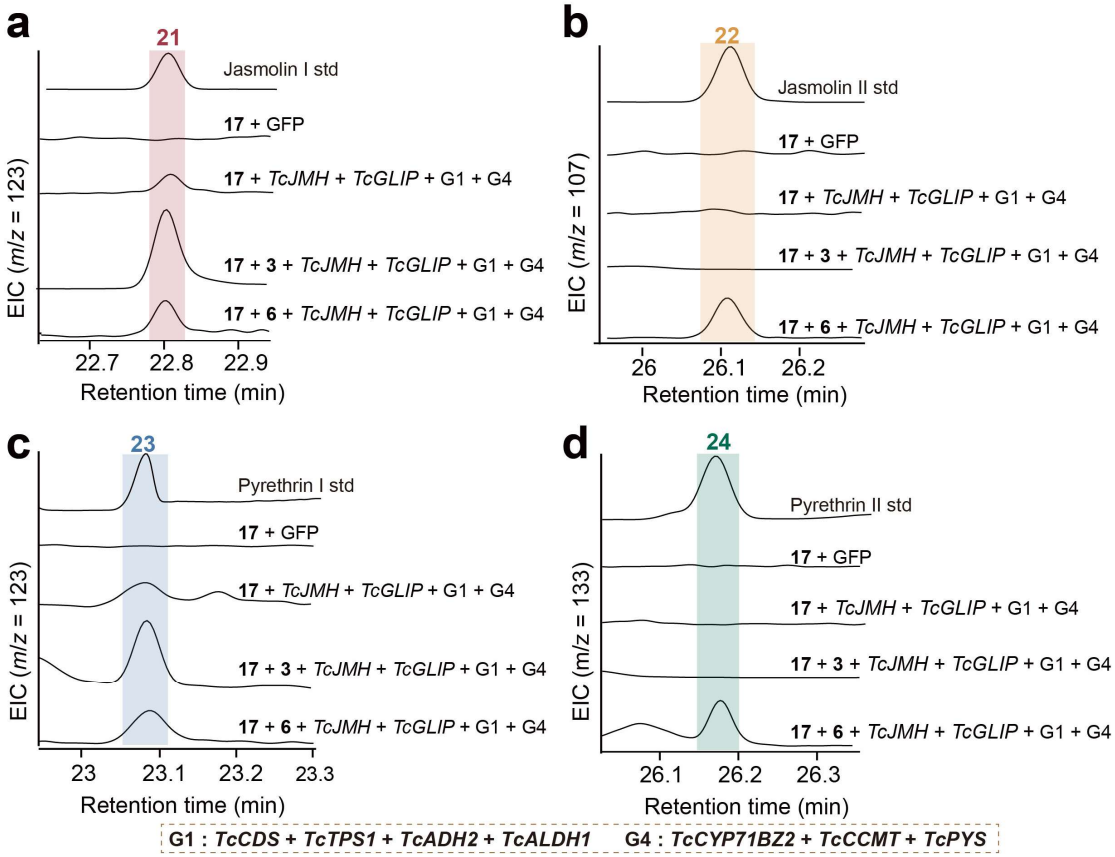
Supplementary Fig. 29 Time-course of jasmolin I (21) formation and chrysanthemoyl-glucoside (8) consumption across different reaction combinations. a GC-MS extracted ion chromatograms (EIC, *m*/*z* = 123.1) showing jasmolin I (**21**) accumulation over time. **b** Bar charts comparing jasmolin I (**21**) peak areas across different combinations. **c** LC-MS EIC ($[M+HCOOH-H]^-$ ion; *m*/*z* = 375.1) showing consumption of chrysanthemoyl-glucoside (**8**) over time. **d** Bar charts comparing chrysanthemoyl-glucoside (**8**) peak areas across different combinations. Data are presented as mean ± SD (*n* = 3 independent reactions).



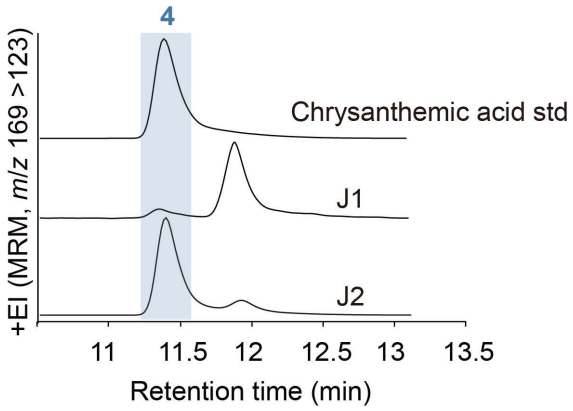
Supplementary Fig. 30 Functional characterization of selected SCPL proteins from *T. cinerariifolium in vitro*. The reaction mixtures contain 0.1 mmol/L Tris-HCl (pH 7.5), 5 mmol/L $MgCl_2$, 1 mmol/L UDP-Glucose, 1 mmol/L chrysanthemic acid (**4**) and 1 mmol/L jasmolone (**18**). The reaction product was analyzed using LC-MS and EIC was shown.



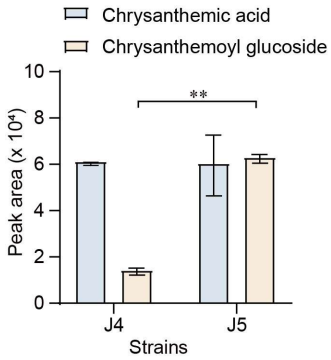
Supplementary Fig. 31 Comparison of active-site pocket volumes among TS (left), TcGLIP (middle) and LS (right). Pockets were predicted with DoGSiteScorer implemented in the ProteinsPlus server (https://proteins.plus) [64, 65].



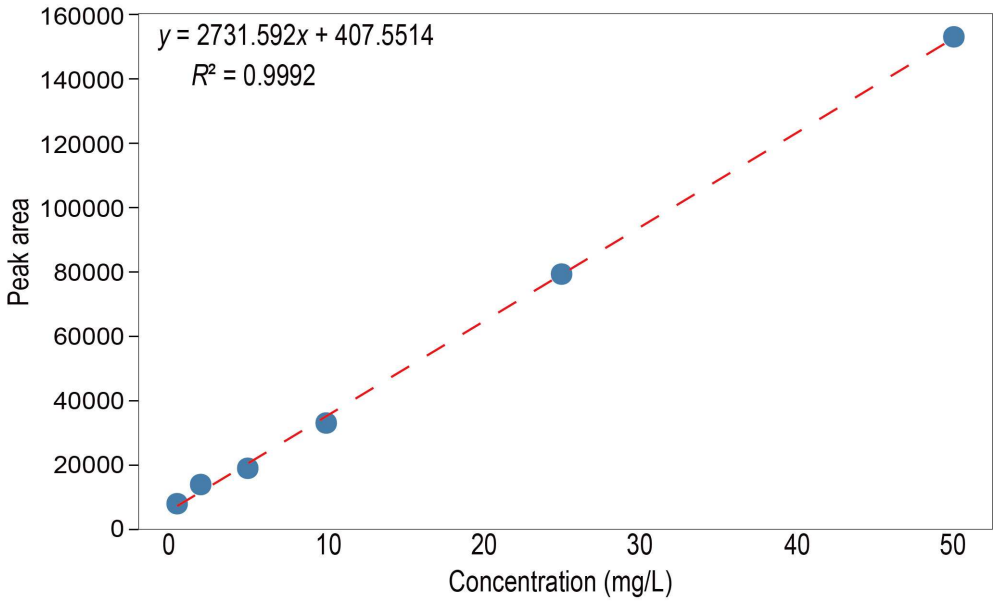
Supplementary Fig. 32 GC-MS chromatograms showing the formation of **a** jasmolin I (**21**), **b** jasmolin II (**22**), **c** pyrethrin I (**23**), and **d** pyrethrin II (**24**) through different combinations of gene sets and substrates.



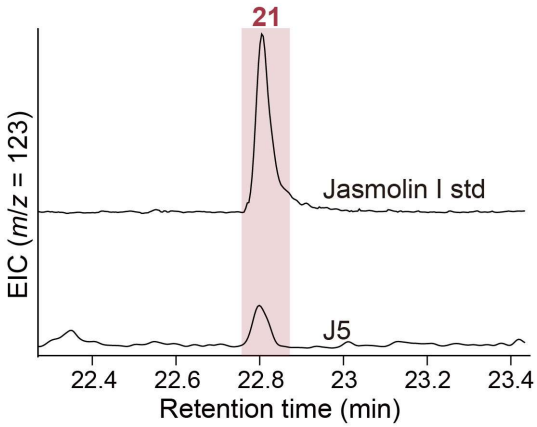
Supplementary Fig. 33 Production of chrysanthemic acid in *S. cerevisiae*. The multiple reaction monitoring (MRM) mode was used with the precursor ion (*m*/*z* = 169) and product ion (*m*/*z* = 123).



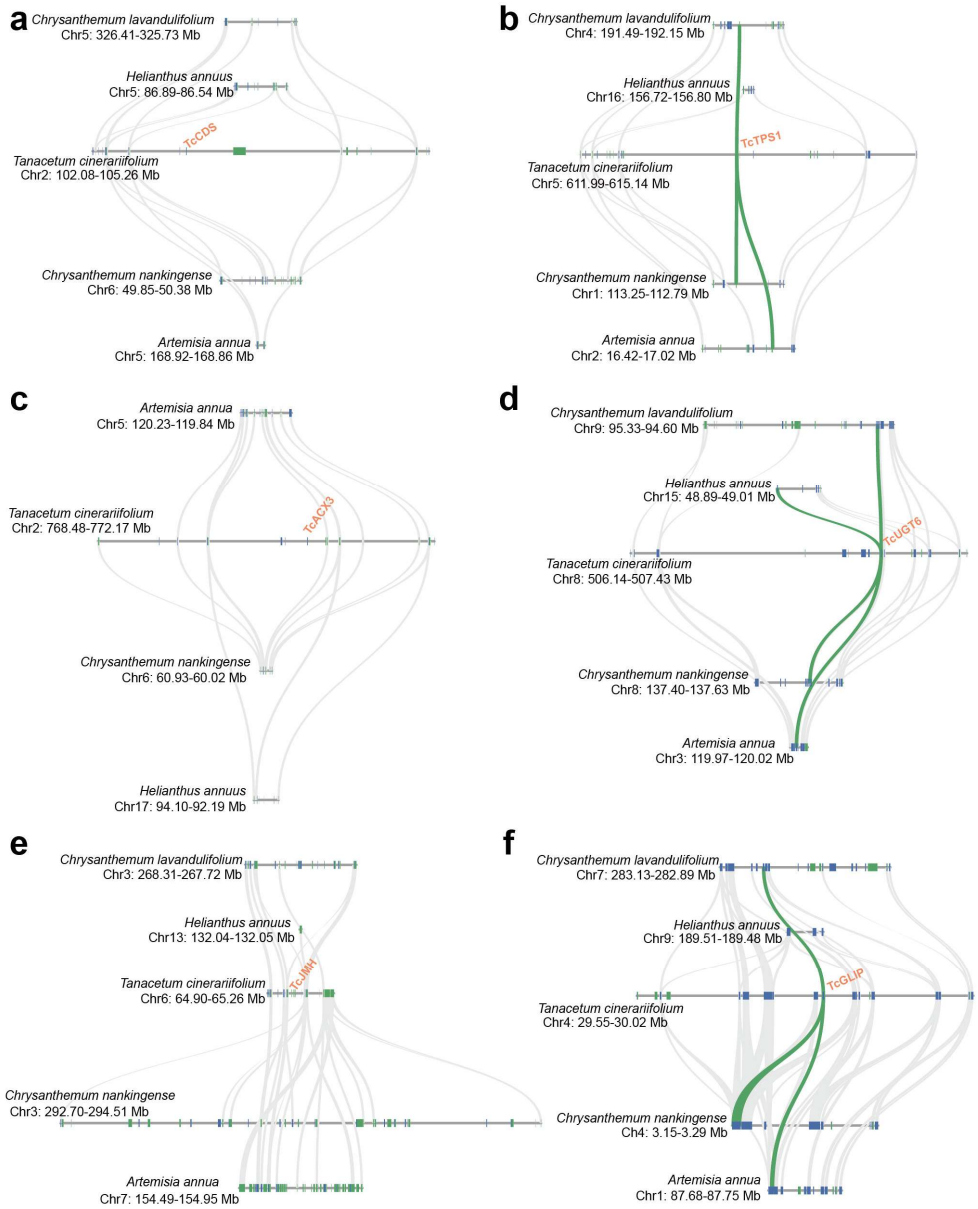
Supplementary Fig. 34 Improving *S. cerevisiae* UDP-glucose supply optimizes the titer of chrysanthemoyl-glucoside (8). Statistical significance was determined by Student's *t*-test ($n = 3$, $^{**}p < 0.01$).



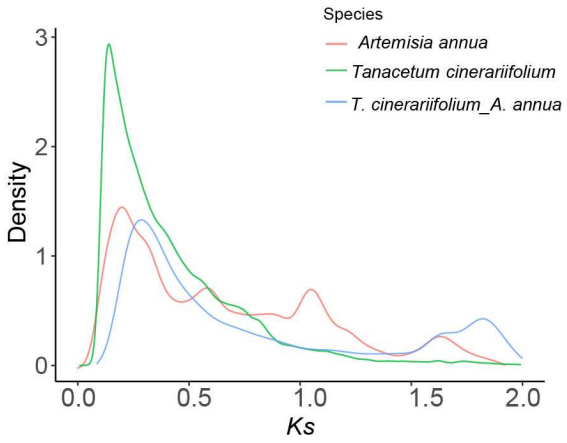
Supplementary Fig. 35 Standard curve for the quantification of jasmolin I (21) production in yeast strain J5 and J6.



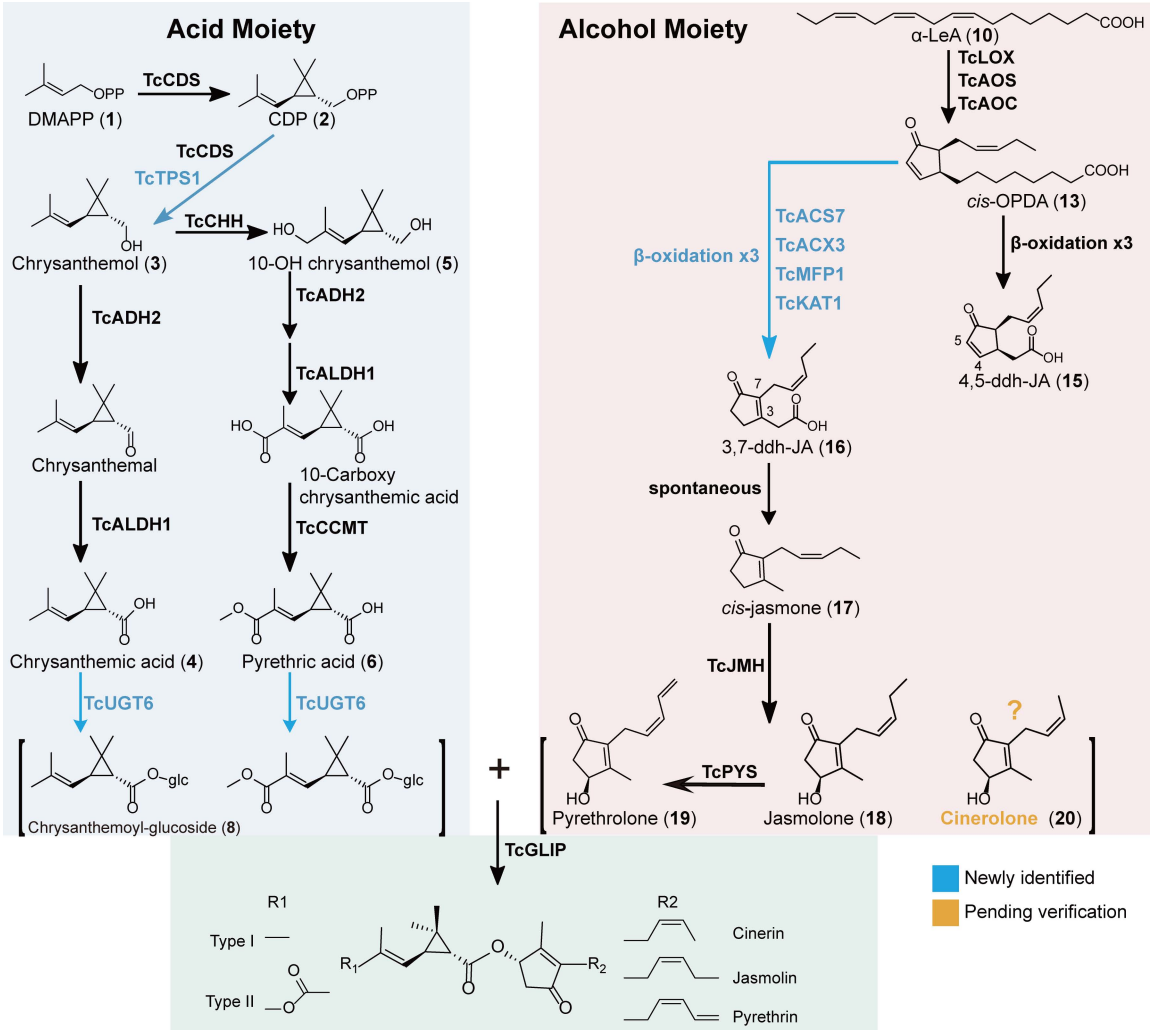
Supplementary Fig. 36 GC-MS chromatograms showing the formation of jasmolin I (21) in yeast strain J5.



Supplementary Fig. 37 Microsynteny plot genes involved in the biosynthesis of pyrethrins. The genes marked in orange are the target genes. **a**–**f** represent *TcCDS*, *TcTPS1*, *TcACX3*, *TcUGT6*, *TcJMH*, *TcGLIP* respectively**.**



Supplementary Fig. 38 The *Ks* distribution of orthologous gene pairs of *T. cinerariifolium* and *Artemisia annua*. Orthologous gene pairs for each species were identified using the FASTKs pipeline. The *Ks* (synonymous substitution rate) values and *Ka* (nonsynonymous substitution rates) were subsequently estimated with PAML. Finally, *Ks* frequency distributions were visualized using the R packages ggplot2 and Mclust.



Supplementary Fig. 39 Summary of the pyrethrins biosynthetic pathway. Core genes newly identified in this study are highlighted in blue, while previously characterized pathway genes are shown in black; the biosynthesis of cinerolone remains unresolved (indicated in orange).

References

1. Cheng, H., Concepcion, G., Feng, X., Zhang, H. & Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18** (2), 170-175 (2021).
2. Walker, B., *et al*. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**(11): e112963 (2014).
3. Durand, N. C. *et al*. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* **3**, 95–98 (2016).
4. Dudchenko, O. *et al*. *De novo* assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
5. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes, *Mol Bio Evol* **38**, 4647–4654 (2021).
6. Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., & Gurevich, A. Versatile genome assembly evaluation with QUAST-LG, *Bioinformatics* **34** (13), i142-i150 (2018).
7. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
8. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, CC3094–3100 (2018).
9. Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
10. Guizard, S., *et al*. nf-core/isoseq: simple gene and isoform annotation with PacBio Iso-Seq long-read sequencing. *Bioinformatics* **39**(5): btad150 (2023).
11. Flynn, J.M., *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci* U S A **117**(17):9451–9457 (2020).
12. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* **35**, W265–268 (2007).
13. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics **9**, 18 (2008).
14. Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol* **176**, 1410–1422 (2018).
15. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
16. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. *Bioinformatics* Chapter 4, 4.10.11–14.10.14 (2009).
17. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573–580 (1999).
18. Lomsadze, A., Burns, P. D. & Borodovsky, M. Integration of mapped RNA-Seq

reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res* **42**, e119 (2014).
19. Stanke, M. *et al*. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* **34**, W435–439 (2006).
20. Pertea, M. *et al*. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290–295 (2015).
21. Haas, B.J., *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* **9**(1), R7 (2008).
22. Lagesen, K. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**, 3100–3108 (2007).
23. Lowe, T.M. & Eddy, S.R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955–964 (1997).
24. Nawrocki, E. P. & Eddy, S.R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
25. Kalvari, I. *et al*. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res* **49**, D192–D200 (2021).
26. Boeckmann, B. et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* **31**, 365–370 (2003).
27. Jones, P. *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
28. Emms, D.M., & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238 (2019).
29. Shen, Q. *et al*. The Genome of *Artemisia annua* Provides Insight into the Evolution of Asteraceae Family and Artemisinin Biosynthesis. *Mol Plant* **11**(6), 776–788 (2018).
30. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
31. DePamphilis, C. W. *et al*. The *Amborella* genome and the evolution of flowering plants. *Science* **342** (6165), 1241089 (2013).
32. Wen, X. *et al*. The *Chrysanthemum lavandulifolium* genome and the molecular mechanism underlying diverse capitulum types. *Hortic Res* **9**, uhab022 (2022).
33. Song, C. *et al*. The *Chrysanthemum nankingense* Genome Provides Insights into the Evolution and Diversification of Chrysanthemum Flowers and Medicinal Traits. *Mol Plant* **11** (12), 1482–1491 (2018).
34. Badouin, H. *et al*. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* **546** (7656), 148–152 (2017).
35. Firl, A. *et al*. Chromosome-scale assembly of apple mint *(Mentha suaveolens). BMC Genom Data* **25** (1), 96 (2024).
36. Chen, W. *et al.* The complete genome assembly of *Nicotiana benthamiana* reveals the genetic and epigenetic landscape of centromeres. *Nat Plants* **10**, 1928–1943 (2024).
37. Lin, S. *et al*. Extensive sequence divergence between the reference genomes of Taraxacum kok-saghyz and *Taraxacum mongolicum*. Sci *China Life Sci* **65**, 515–

528 (2022).
38. Wang, P. *et al*. Genomes of single- and double-petal jasmines (*Jasminum sambac*) provide insights into their divergence time and structural variations. *Plant Biotechnol J* **20**, 1232–1234 (2022).
39. Shi, X. *et al*. The complete reference genome for grapevine (*Vitis vinifera* L.) genetics and breeding, *Hortic Res* **10**, uhad061 (2023).
40. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies *Bioinformatics* **30**, 1312–1313 (2014).
41. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
42. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Bio Evo* **24**, 1586–1591 (2007).
43. Wang, J. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**, e49 (2012).
44. Tang, H. *et al*. JCVI: A Versatile Toolkit for Comparative Genomics Analysis. *iMeta* **3**, e211 (2024).
45. Langfelder, P., & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **29**, 559 (2008).
46. Eddy, S. R. What is a hidden Markov model? *Nat Biotechnol* **22**, 1315–1316 (2004).
47. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059–3066 (2002).
48. Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268–274 (2015).
49. Ma, T., *et al*. Lipid engineering combined with systematic metabolic engineering of Saccharomyces cerevisiae for high-yield production of lycopene. *Metab Eng* **52**, 134–142 (2019).
50. Gietz, R.D., & Schiestl, R.H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2**(1), 31–34 (2007).
51. Zhi, Y., *et al*. Gene-Directed In Vitro Mining Uncovers the Insect-Repellent Constituent from Mugwort (*Artemisia argyi*). *J Am Chem Soc*. **146**(45), 30883–30892 (2024).
52. Sheen, J. Signal transduction in maize and Arabidopsis mesophyll protoplasts. *Plant Physiol* **127**(4), 1466–1475 (2001).
53. Abramson, J., *et al*. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**(8016), 493–500 (2024).
54. Trott, O. & Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455–461 (2010).
55. Riley, H.L., Morley, J.F., & Friend, N.A.C. Selenium dioxide, a new oxidising agent. Part I. Its reaction with aldehydes and ketones. *J Chem Soc* **255**, 1875–1883 (1932).

56. Nickel, A., *et al*. Total synthesis of ingenol. *J Am Chem Soc* **125** (6), 1498–1500. (2003).
57. Hideyuki, M., Fumihiro, O., Masatomo, K., Akira, S., Teruhiko, Y., Qualitative and Quantitative Analysis of Endogenous Jasmonoids in Potato Plant (*Solanum tuberosum* L.), *Biosci Biotechnol Biochem*, **64** (11), 2380–2387 (2000).
58. Dang, H.T., *et al*. New jasmonate analogues as potential anti-inflammatory agents. *Bioorg Med Chem* **16**(24), 10228–10235 (2008).
59. Kajiwara, A., Abe, T., Hashimoto, T., Matsuura, H., & Takahashi, K. Efficient synthesis of (+)-cis-12-oxo-phytodienoic acid by an in vitro enzymatic reaction. *Biosci Biotechnol Biochem* **76,** 2325–2328 (2012).
60. Yang, T., et al. Chrysanthemyl Diphosphate Synthase Operates in Planta as a Bifunctional Enzyme with Chrysanthemol Synthase Activity. *J Biol Chem* **289**, 36325–36335 (2014).
61. Xu, H., *et al*. Pyrethric acid of natural pyrethrin insecticide: complete pathway elucidation and reconstitution in *Nicotiana benthamiana*. *New Phytol* **223**(2), 751–765 (2019).
62. Li, P., *et al*. Multiomics analyses of two Leonurus species illuminate leonurine biosynthesis and its evolution. *Mol Plant* **17**(1), 158–177 (2024).
63. Karlova, R., *et al*. Detoxification of monoterpenes by a family of plant glycosyltransferases. *Phytochemistry* **203**, 113371 (2022).
64. Volkamer, A., Griewel, A., Grombacher, T., & Rarey, M. Analyzing the topology of active sites: on the prediction of pockets and subpockets. *J Chem Inf Model* **50**(11), 2041–2052 (2010).
65. Volkamer, A., Kuhn, D., Grombacher, T., Rippmann, F., & Rarey, M. Combining global and local measures for structure-based druggability predictions. *J Chem Inf Model* **52**(2), 360–772 (2012).