Journal Pre-proof

Multiomics driven identification of glycosyltransferases in flavonoid glycoside biosynthesis in safflower

Nan Liu, Yupan Zou, Zhouqian Jiang, Lichan Tu, Xiaoyi Wu, Dan Li, Jiadian Wang, Luqi Huang, Cao Xu, Wei Gao

PII: S2468-0141(24)00238-3

DOI: https://doi.org/10.1016/j.hpj.2024.01.016

Reference: HPJ 698

To appear in: *Horticultural Plant Journal*

Received Date: 26 December 2023

Revised Date: 17 January 2024

Accepted Date: 18 January 2024

Please cite this article as: Liu N, Zou Y, Jiang Z, Tu L, Wu X, Li D, Wang J, Huang L, Xu C, Gao W, Multiomics driven identification of glycosyltransferases in flavonoid glycoside biosynthesis in safflower, *Horticultural Plant Journal*, https://doi.org/10.1016/j.hpj.2024.01.016.

Multiomics driven identification of glycosyltransferases in flavonoid glycoside biosynthesis in safflower

Nan Liu[a,b,#], Yupan Zhou[c,f,#], Zhouqian Jiang[a,d,#], Lichan Tu[a], Xiaoyi Wu[a], Dan Li[e], Jiadian Wang[a], Luqi Huang[b,*], Cao Xu[c,f,*], Wei Gao[d,a,*]

a School of Traditional Chinese Medicine, Capital Medical University, Beijing, 100069, China

b National Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, National Resource Center for Chinese Materia Medica, Chinese Academy of Chinese Medical Sciences, Beijing 100700, China

c State Key Laboratory of Plant Genomics, National Center for Plant Gene Research (Beijing), Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing, 100049, China

d Beijing Shijitan Hospital, Capital Medical University, Beijing, 100038, China

e School of Pharmaceutical Sciences, Capital Medical University, Beijing, 100069, China

f CAS-JIC Centre of Excellence for Plant and Microbial Science, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, 100049, China

# These authors contributed equally to this work.
* Corresponding author. W. G. ; C. X. (Tel:010-64803911); H-L. Q.
E-mail address: W. G. : weigao@ccmu.edu.cn; C. X. : caoxu@genetics.ac.cn; H-L. Q. : huangluqi01@163.com.
Peer review under responsibility of Chinese Society of Horticultural Science (CSHS) and Institute of Vegetables and Flowers (IVF), Chinese Academy of Agricultural Sciences (CAAS)

Https://doi.org/10.1016/j.hpj.

Multiomics driven identification of glycosyltransferases in flavonoid glycoside biosynthesis in safflower

Nan Liu[a,b,1], Yupan Zou[c,f,1], Zhouqian Jiang[a,d,1], Lichan Tu[a], Xiaoyi Wu[a], Dan Li[e], Jiadian Wang[a], Luqi Huang[b,*], Cao Xu[c,f,*], and Wei Gao[d,a,*]

[a] *School of Traditional Chinese Medicine, Capital Medical University, Beijing 100069, China*

[b] *National Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, National Resource Center for Chinese Materia Medica, Chinese Academy of Chinese Medical Sciences, Beijing 100700, China*

[c] *State Key Laboratory of Plant Genomics, National Cent*

er for Plant Gene Research (Beijing), Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100049, China

[d] *Beijing Shijitan Hospital, Capital Medical University, Beijing 100038, China*

[e] *School of Pharmaceutical Sciences, Capital Medical University, Beijing 100069, China*

[f] *CAS-JIC Centre of Excellence for Plant and Microbial Science, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100049, China*

[1] These authors contributed equally to this work.
* Corresponding author.
E-mail address: weigao@ccmu.edu.cn; caoxu@genetics.ac.cn; huangluqi01@163.com.

ABSTRACT

Safflower is an important oilseed crop that has been used in traditional Chinese medicine for thousands of years because of the clinically valuable flavonoid glycosides in its flower petals. However, the biosynthesis and molecular regulation of these compounds are still elusive due to the lack of a high-quality reference genome and scarce identification of key biosynthetic pathway genes in a medicinal safflower variety. Here we leveraged an integrative multi-omics strategy by combining genomic, comparative genomics, and tissue-specific transcriptome profiling with biochemical analysis to identify uridine diphosphate glycosyltransferases (UGTs) for flavonoid glycoside biosynthesis in safflower. We assembled and annotated a high-quality reference genome of a medicinal safflower variety, 'Yunhong3'. A comprehensive comparative genomic analysis indicated that an evolutionary whole-genome triplication event occurring in safflower contributed to gene amplification of the flavonoid biosynthetic pathway. By combining comparative transcriptome profiling with enzymatic reactions, we identified 11 novel UGTs that could catalyze the conversion of naringenin chalcone and phloretin to the corresponding O-glycosides. Moreover, we outlined the molecular pathway of hydroxysafflor yellow A (HSYA) biosynthesis featured by 17 newly identified UGTs with promising catalytic activity, laying the foundation for the synthetic production of HSYA. Our study reports systemic genome and gene expression information for flavonoid glycoside biosynthesis in medicinal safflower and provides insights into mechanisms regulating HSYA biosynthesis, which would facilitate the genetic improvement and synthetic bioengineering design for producing clinically valuable flavonoid glycosides in safflower.

Keywords: Safflower; Genomics; Transcriptome; Glycosyltransferase; Flavonoid; Hydroxysafflor yellow A

1. Introduction

Carthamus tinctorius L. (safflower) is a globally cultivated medicinal herb revered for its therapeutic properties (Ekin 2005; Dordas and Sioulas, 2008). The dried tubular safflower flower has long been utilized to

enhance blood circulation, alleviate blood stasis, and relieve menstrual pain, with its usage first recorded in *Kaibao Bencao* (Asgarpanah and Kazemivash, 2013). Modern medical research has validated safflower's extensive pharmacological activities, including the dilation of coronary arteries, as well as its anticoagulatory, antithrombotic, antioxidative, and anti-inflammatory effects (Han et al., 2009; Zhou et al., 2014; Zhang et al., 2016; Li et al., 2019a; Li et al., 2019b). The medicinal value of safflower is strongly associated with its active ingredients. To date, more than 200 compounds have been isolated from safflower plants, including flavonoids, spermidines, alkaloids, lignin, organic acids, and steroids (Kazuma et al., 2000; Yue et al., 2013; Hong et al., 2015). Years of clinical and pharmacological research have revealed that most of these compounds can affect the cardiovascular, cerebrovascular, nervous, and immune systems, making them promising drugs for various diseases (Asgarpanah and Kazemivash, 2013; Zhou et al., 2014; Zhang et al., 2016; Martin and Li, 2017).

Flavonoid derivatives are the most abundant active compounds in safflower, and they mostly exist as glycosides (Zhou et al., 2014). Based on their glycosylated forms, about two-thirds of flavonoids are classified as O-glycosides, including naringenin, kaempferol, quercetin, luteolin, and their derivatives (Hattori et al., 1992; Lee et al., 2002; Qu et al., 2015). Some other glycosides, which are associated with red and yellow pigmentation, have been designated as quinochalcone C-glycosides (Xian et al., 2022). Among them hydroxysafflor yellow A (HSYA) is the most valuable medicinal compound (Yin and He, 2000; Yue et al., 2014; Li et al., 2017; Zhang et al., 2020). Specifically, HSYA has been used as an indicator of the medical value of safflower in the *Pharmacopoeia of the People's Republic of China* (from the 2015 edition). It has also been approved as a novel drug for treating cardiovascular disease by the China State Food and Drug Administration (Li et al., 2012; Ao et al., 2018; Sun et al., 2018).

Most plant natural products, including flavonoids, undergo various post-modifications (e.g., glycosylation, methylation, acylation, and hydroxylation) to form their final complex and diverse structures. Glycosylation is one of the most common modifications (Alseekh et al., 2020). Glycosylated compounds are important for plant growth and development, metabolic regulation, maintenance of hormone homeostasis, pollination, detoxification of internal and external toxins, and defense responses (Vogt and Jones 2000; Jones and Vogt 2001; Stevenson et al., 2017; Brazier-Hicks et al., 2018). Additionally, compared with their precursors, the glycosylation of active ingredients often leads to increased polarity and pharmaceutical activities (Luzhetskyy et al., 2008). Hence, glycosylated compounds are widely used in medicine, agriculture, and other industries (Liu 2014; Hsu et al., 2018; Itkinet al., 2018; Khan et al., 2019).

Glycosyltransferases, which transfer glycosyl groups from activated glycosyl donors to acceptors, are vital for post-translational modifications that fully form flavonoid glycosides (Alseekh et al., 2020; Chung et al., 2020). Generally, glycosylation occurs at positions C-3, C-6, and C-7 of the flavonoid backbone. Differences in the types of sugars, positions of glycosidic bonds, and aglycones often give rise to structurally and functionally diverse flavonoids (Sato et al., 2006; Xie et al., 2014; Wang et al., 2020). In the flavonoid biosynthetic pathway, naringenin chalcone acts as an important precursor that undergoes a series of modifications (e.g., glycosylation) to generate

flavonoids and flavonols (Tohge et al., 2017). It has also been recognized to get involved in the biosynthesis of quinochalcone C-glycosides, including HSYA. Although glycosylation is essential for the bioactivation of flavonoid glycosides in safflower, the glycosyltransferases have rarely been cloned or identified. How the UGT enzymes could modify naringenin chalcone to produce glycosides remains unknown.

The rapid development of sequencing technology has significantly facilitated the exploration of genetic information for medicinal plants and the identification of disease-related gene loci (Teng et al., 2018). A recently published chromosome-scale safflower reference genome has provided genomic insights into flavonoid biosynthesis (Wu et al., 2021). Previous transcriptomic studies have identified many candidate genes related to flavonoid biosynthesis, including genes encoding chalcone synthases (CHSs), UGTs, and chalcone isomerases (CHIs) (Chen et al., 2020; Qiang et al., 2020; Wu et al., 2021). However, there have been no systematic analyses of the evolution and function of UGT families in safflower. Major obstacles to studying the glycosylation of flavonoid glycosides in safflower include the limited genomic resources for medicinal species and the lack of identified functional UGTs. In this study, we applied a multi-omics strategy involving comparative genomic and tissue-specific transcriptomic analyses to screen for flavonoid glycoside-specific *UGT* genes, which was followed by in *vitro* enzymatic assays to identify functional UGTs and molecular docking analysis to explore the catalytic mechanisms. We identified a series of candidate *UGT* genes, of which 17 were associated with the HYSA biosynthetic pathway. Eleven of these *UGT* genes encoded enzymes that could catalyze naringenin chalcone and phloretin to produce the corresponding glycosylated compounds.

2. Materials and methods

2.1. Genome Assembly

Fresh leaf tissues of safflower were used to extract genome DNA. A 40-kb insert PacBio SMRTbell library was prepared using SMRTBELL Express Template prep kit 2.0. The SMRTbell library was sequenced on the SMRT PacBio Sequel II platform and obtained 142.72 Gb of PacBio reads. The CANU software (version 1.8, https://github.com/marbl/canu) was utilized to preprocess and assemble the PacBio reads into contigs, with parameter “genome size=1.35G” (Koren et al., 2017). To correct and fill gaps for misassembling, the assembler HERA (version 1.0) was applied to update the contigs to supercontigs with default parameters (Du and Liang, 2019).

Fresh young leaf tissues from living plants were collected as BioNano Genomics guidelines. High-Molecular-Weight (HMW) DNA was extracted from the leaves with a BioNano Plant Tissue DNA isolation kit (BioNano Genomics) and fluorescently labeled using DpnII endonucleases (BioNano Genomics) based on the BioNano Direct Label and Stain technology. After more than two hours of staining at room temperature, the fluorescently labelled DNA was loaded on the Saphyr chips for scanning on the BioNano Genomics Saphyr System, yielding 475.7 Gb of Bionano data. This data was employed to assist the scaffold construction using Bionano Solve software (version 3.3). The resulting supercontigs and scaffolds were integrated using 202.56 Gb of Hi-C data

through Juicer and 3D-DNA (Durand et al., 2016; Dudchenko et al., 2017).

2.2. Genome Annotation

Three gene prediction methods were used to annotate the safflower genome, including *ab initio*, assembled transcripts, and protein homologs. In *ab initio* prediction, four programs were utilized, including Glimmer-HMM, SNAP (version 2.3), GeneMark-ET (version 4.57), and AUGUSTUS (version3.3) (Korf, 2004; Majoros et al., 2004; Stanke et al., 2004; Ter-Hovhannisyan et al., 2008). Annotated protein sequences from *C. cardunculus*, *S. lycopersicum*, *A. thaliana*, *C. nankingense*, *H. annuus*, *V. vinifera*, *L. sativa*, and *T. kok-saghyz* were downloaded from phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html) (Goodstein et al., 2012). These protein sequences were employed for homology-based prediction by GeMoMa (version 1.6.1) software (Haas et al., 2008). All predicted gene models were combined by EvidenceModeler (EVM) into a non-redundant set of gene structures (Haas et al., 2008). In total, 35 162 genes were predicted from the safflower genome with transcript support. Gene functional annotation was performed by searching against various functional databases, such as Swiss-Prot, eggNOG (Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups), NR (Non-Redundant Protein Sequence Database), Pfam, and GO (Gene Ontology). Genome assembly and annotation completeness were assessed using BUSCO (Benchmarking Universal Single-Copy Orthologs version 3.0.2) (embryophyta_odb10) (Simao et al., 2015). We used the LTR_FINDER (x86_64-1.0.5) and LTR_retriever (v1.9) software (Xu and Wang, 2007; Ou and Jiang, 2019) to scan the long terminal repeat retrotransposons (LTR-RTs) of the safflower genome. Among them, tRNA structures were identified using tRNAscan-SE (v1.3.1) (Lowe and Eddy, 1997) software. Non-coding RNAs were annotated using the INFERNAL (v1.1.2) software (Nawrocki and Eddy, 2013). The noncoding RNAs were then classified into different types containing miRNA, snRNA, and snoRNA by searching against the Rfam database.

2.3. Analysis of Genomic Evolution and WGD Events

Paralogous gene pairs were identified using Blast-based methods, and syntenic paralogs were determined using MCscanX (version 1.0) (Wang et al., 2012). Orthologous and paralogous gene families were assigned by OrthoFinder (version 2.2.6) (Emms and Kelly, 2015) with default parameters. A total of 195 single-copy orthologous genes were selected to construct the phylogenetic tree. The protein sequences of each gene family were independently aligned by MAFFT (v7.402) (Katoh et al., 2005). The phylogenetic tree was constructed by maximum likelihood (ML) using RAxML-NG (v. 0.9.0) (Kozlov et al., 2019) employing the best-fit model GTR+G4 which was estimated by ModelTest-NG (Darriba et al., 2020). Divergence times between species were estimated using the MCMCTree programmed in the PAML package (v4.9). The calibration time of the divergence between grape and coffee (111-131 Mya) was adopted from TimeTree (http://www.timetree.org). GO and KEGG pathway enrichment analysis were conducted using the clusterProfiler package in R (version 3.6.3) (Yu et al., 2012). To estimate the timing of the whole genome duplication event in safflower, Ks values of safflower syntenic block genes were calculated using the YN model in KaKs_Calculator (version 2.0) (Wang et al., 2010). The synonymous substitution rate of 8.25×10^{-9} mutations per site per year for asterids was applied to calculate the ages of the

WGDs (Badouin et al., 2017). Gene family expansion or contraction were calculated using CAFÉ (version 4.2) (De Bie et al., 2006).

2.4. Transcriptome Sequencing and Analysis

The fresh safflower plants were subdivided into five tissues: flowers, pappus, bract, leaves, and stems. All samples were immediately frozen in liquid nitrogen, and stored at -80 °C before RNA extraction. Three biological replicates for each tissue were collected. Total RNA was extracted from these tissues using the EasyPure® RNA Kit (Trans, Beijing, China) following the manufacturer's instructions. RNA purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). The Qubit® RNA Assay Kit in Qubit® 2.0 Flurometer (Life Technologies, CA, USA) was used to measure RNA concentration. RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). The cDNA library was constructed using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina (NEB, USA), following the manufacturer's recommendations. Library preparations were sequenced on an Illumina Hiseq X Ten platform and paired-end reads were generated. RNA-seq reads were mapped to the safflower genome using Hisat2 and assembled into transcripts using StringTie (version 2.0.3) (Kim et al., 2015; Pertea et al., 2015). DESeq2 (version 1.30.1) was used for differential gene expression analysis between two tissues with biological replicates. R (v4.0.3) software was used to draw the heat maps in figures. In addition, RNA-seq data were assembled into transcripts by Trinity (Grabherr et al., 2011). These transcript evidences were mapped to the assembly, and gene models were predicted by the Program to Assemble Spliced Alignments (PASA) (version 2.0.1) (Campbell et al., 2006). Transcriptome sequencing was previously performed using the tubular flower of safflower under the MeJA treatment from the State Key Laboratory of Characteristic Chinese Medicine Resources in Southwest China, Chengdu University of Traditional Chinese Medicine, Chengdu, China (Chen et al., 2020). These transcriptome data were analyzed in the current study.

2.5. Identification, Phylogenetic and Regulation Analysis of MYB Transcription Factors

The Hidden Markov Model (HMM) profile of MYB (PF00249) from the Pfam protein family database (http://pfam.xfam.org/) was used as the query entry to perform an HMM search against the protein data of the safflower genome (El-Gebali et al., 2019). The obtained protein sequences were further confirmed using the web resource tools, Conserved Domain Search (https://www.ncbi.nlm.nih.gov/cdd/) (Lu et al., 2020) and matched with annotated proteins in the Pfam database. Based on the reads of each gene in different transcriptome data, egdeR was used to analyze the differential expression genes, and the groups for comparison were Floret-Phyllary, Floret-Stem, Floret-Leaf, and Floret-Pappus. Then we used SPSS software to calculate the Pearson correlation coefficient between MYB TF genes and pathway genes to construct co-expression regulatory networks to trace back to the specific *MYB* genes in the pathway. The network was displayed through Cytoscape (Shannon, 2003). To determine whether the *MYB* genes identified were flavonoid biosynthesis-related, all the known *MYB* genes of *A. thaliana*, *Malus domestica*, *Gentiana trifloral*, *Solanum lycopersicum*, and *P. hybrida* were downloaded from the GenBank database. The protein sequences were aligned using CLUSTAL-X (Larkin et al., 2007) and the

neighbor-joining tree was constructed using MEGAX (Kumar et al., 2018) under the *p*-distance mode with 1 000 bootstrap samples and complete deletion. The online website ITOL (https://itol.embl.de/) was employed to visualize the tree (Letunic and Bork, 2019).

2.6. Screening and Functional Verification of Candidate UGT Genes

According to the genes functional annotation information and the tissue expression information, we screened 187 full-length candidate *UGT* sequences with conserved domain and the FPKM > 5. Among them, 23 *UGT* genes were narrowed down for further analysis due to their high expression level in flowers. We performed in *vitro* functional identification of these 23 *UGT* genes and found that 8 of them had glycosyltransferase activities. In addition, we also conducted homologous sequence alignment of the candidate *UGT* genes regarding a variety of glycosyltransferases that had been reported in the previous literature (Modolo et al., 2007; Brazier-Hicks et al., 2009; Schnable et al., 2009; Chen et al., 2015; Ito, Fujimoto et al., 2017; He et al., 2019; Mashima, et al., 2019; Gao, et al., 2020). A total of 11 candidate genes were screened, and finally, 34 glycosyltransferases were identified.

Multiple sequence alignments were generated using DNAMAN to visualize the conserved motifs. For phylogenetic tree analysis, the amino acid sequences of UGTs from other species were downloaded from the National Center for Biotechnology Information (NCBI) database and aligned using ClustalW. Then, a neighbor-joining tree was built using MEGA X software with 1 000 bootstrap iterations. The safflower cDNA was prepared using the PrimeScript 1st Strand cDNA Synthesis Kit (Takara, Dalian, China). After designing primers, we cloned a total of 34 *UGT* genes, and the PCR products were ligated into the N-terminal MBP fusion expression vector HIS-MBP-pET28a (HIS, histidine; MBP, maltose-binding protein) according to the protocol of the Seamless Cloning Kit (Beyotime, Shanghai, China). We transformed the successfully sequenced positive strains into *E. coli* BL21 (DE3) (Transgen Biotech, Beijing, China) and maintained the cultures in Luria Bertani liquid medium with kanamycin (50 $\mu g \cdot mL^{-1}$) at 37 °C in a shaking incubator until the optical density at 600 nm (OD_{600}) reached 0.6–0.8. Then, isopropyl β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 $mmol \cdot L^{-1}$, and cultures were induced at 16 °C and 200 $r \cdot min^{-1}$ for 16 h to allow expression of recombinant proteins. pET28a-transformed *E. coli* BL21 (DE3) cells were treated in parallel as a control.

To investigate the biochemical properties of CtUGT10 in *vitro*, recombinant CtUGT10 was purified by Ni-NTA affinity chromatography. The recombinant cells were harvested by centrifugation at 10 000 *g* and 4 °C, then resuspended in 100 $mmol \cdot L^{-1}$ Tris-HCl buffer (pH 8.0) that contained 1 $mmol \cdot L^{-1}$ phenylmethanesulfonylfluoride and sonicated in an ice-water bath for 10 min (lysed for 4 s, paused for 6 s). The sample lysates were centrifuged for 40 min at 15 000 *g* and 4 °C to separate crude enzymes from cell debris. A UGT activity assay was performed in a total volume of 100 µL that contained 100 $mmol \cdot L^{-1}$ crude enzyme buffer (pH 8.0), 1 $mmol \cdot L^{-1}$ UDP-glucose, and 0.1 $mmol \cdot L^{-1}$ acceptor substrate for 12 h in 37 °C and was terminated by the addition of 200 µL methanol. Precipitated proteins were removed by centrifugation (10 000 *g* for 10 min) and filtered through 0.22 µm filters before injection.

The experimental methods to explore the influence of different conditions on enzyme activity are as follows:

Effect of pH value: The final concentrations employed (total volume, 100 μL) were as follows: UDPG, 2 mmol · L^{-1}, naringin chalcone,0.2mmol · L^{-1}, incubated with 20 μg purified protein, and 50 mmol · L^{-1} buffer. Buffers preparation of different pH values: 50 mmol · L^{-1}citric acid-sodium citrate buffer (pH: 4.0, 5.0, 6.0); 50 mmol · L^{-1} sodium dihydrogen phosphate buffer (pH: 6.0, 7.0, 8.0); 50 mmol · L^{-1} Tris-HCl buffer (pH: 7.0, 8.0, 9.0); 50 mmol · L^{-1} sodium bicarbonate-sodium carbonate buffer (pH: 9.0, 10.0, 11.0). The reaction was performed at 37 °C for 12 h, and each reaction was repeated three times in parallel.

Temperature effects: UDPG and naringin chalcone were added to 50 mmol · L^{-1} Tris-HCl buffer (pH 8.0) to a final concentration of 2 mmol · L^{-1} and 0.2 mmol · L^{-1}, respectively. Then, 20 μg purified protein was added to the mixture. The total volume of the reaction was 100 μL. Different temperature conditions (25, 30, 37, 40, 45, 50, 55, 65 °C) were set for 12 h, and each reaction group was repeated three times in parallel.

Effects of divalent metal ions: UDPG and naringin chalcone were added to 50 mmol · L^{-1} Tris-HCl buffer (pH 8.0) to a final concentration of 2 mmol · L^{-1} and 0.2 mmol · L^{-1}, respectively. Then, 20 μg purified protein was added to the mixture. The total volume of the reaction was 100 μL, and the final concentration of divalent metal ion was 5 mmol · L^{-1}. Different divalent metal ions (Ba^{2+}, Mg^{2+}, Ca^{2+}, EDTA) were added to the buffer and reacted for 12 h, and each reaction group was repeated three times in parallel.

Effect of reaction time: UDPG and naringin chalcone were added to 50 mmol · L^{-1} Tris-HCl buffer (pH 8.0) to a final concentration of 2 mmol · L^{-1} and 0.2 mmol · L^{-1}, respectively. Then, 20 μg purified protein was added to the mixture. The total volume of the reaction was 100 μL. The reaction was terminated by adding 100 μL methanol at 10 min, 30 min, 1 h, 2 h, 4 h, 8 h, 12 h, 24 h, 48 h and 72 h, respectively. Each reaction was repeated three times in parallel.

The enzymatic kinetic assay was carried out under the following conditions: (1) 4 mmol · L^{-1} UDPG (sugar donor), naringin chalcone (substrate) at different concentrations and 20 μg purified protein were added into 50 mmol · L^{-1} Tris-HCl buffer (pH 8.0) to incubate at 37 °C. The total volume of the reaction was 100 μL.. The final concentrations of naringenin chalcone were 2.5, 5, 10, 20, 60, 100, 150, 200, 250, 300 μmol · L^{-1}, the reaction was terminated after 5 h by the addition of 100 μL methanol. Each reaction was repeated three times in parallel. (2) 4 mmol · L^{-1} naringin chalcone as substrate, UDPG of different concentrations as sugar donor, incubated with 20 μg purified protein at 37 °C, adding 50 mmol · L^{-1} Tris-HCl buffered (pH 8.0) to 100 μL total volume. The final concentrations of UDPG were 4, 8, 20, 60, 100, 150, 200, 250, and 300 μmol · L^{-1}, the reaction was terminated after 5 h by the addition of 100 μL methanol. Each reaction was repeated three times in parallel. Glycosylated products were detected using ultra-high-performance liquid chromatography coupled with quadrupole time-of-flight mass spectrometry (UPLC/Q-TOF-MS, Waters, Milford, MA) using a Waters ACQUITY UPLC HSS T3 analytical column (2.1 × 100 mm, 1.8 mm). Data analysis was performed using MassLynx software (version 4.1). The K_m and V_{max} of naringenin chalcone and UDPG were calculated by GraphPad Prism 7.04 software. Standards of flavonoid glycoside compounds and UDP-glucose were purchased from Yuanye Bio-Technology (Shanghai, China).

2.7. Molecular Docking

The homology structures of UGTs were predicted by SWISS-MODEL. Molecular docking was conducted with Autodock software (Version 4.2) (Morris et al., 2009). UDPG was docked into the sugar-binding pocket by referring to the binding conformation of the complex crystal structures of UGTs/UDPG. The docking results were visualized using PyMOL2 software (Schrödinger et al., 2020).

3. Results

3.1. Genome sequencing, assembly, and annotation of a medicinal safflower accession

To comprehensively explore the *UGT* genes related to flavonoid glycoside biosynthesis in medicinal safflower, we selected 'Yunhong3', an excellent medicinal safflower variety in China (Fig. 1, A-C), for an integrative multi-omics investigation by combining genomic, comparative genomics and tissue-specific transcriptome profiling with biochemical analysis (Fig. 1, D). 'Yunhong 3' is a long-day crop native to arid environments with seasonal rainfall (Guo et al., 2009). Its growth period is about 3–5 months (sowing to harvest maturity), which varies slightly depending on the location, sowing time, and environmental conditions. In China, 'Yunhong 3' is mainly cultivated in the northwestern region, and its petals harvested as valuable materials for traditional Chinese medicine. The safflower flowering phase is relatively short, usually lasting 1 to 2 days. The petals collected from tubular flowers gradually change from yellow to red during the drying and storage periods (Fig. 1, C). Unlike its seeds, which are commonly used as a source of edible oil, 'Yunhong3' petals are primarily employed in Chinese medicine to promote blood circulation and eliminate blood stasis. The tubular flowers contain many active components, including flavonoids, alkaloids, polyalkynes, lignans, and polysaccharides. Notably, the dried petals of 'Yunhong3' contain high concentrations of HSYA.

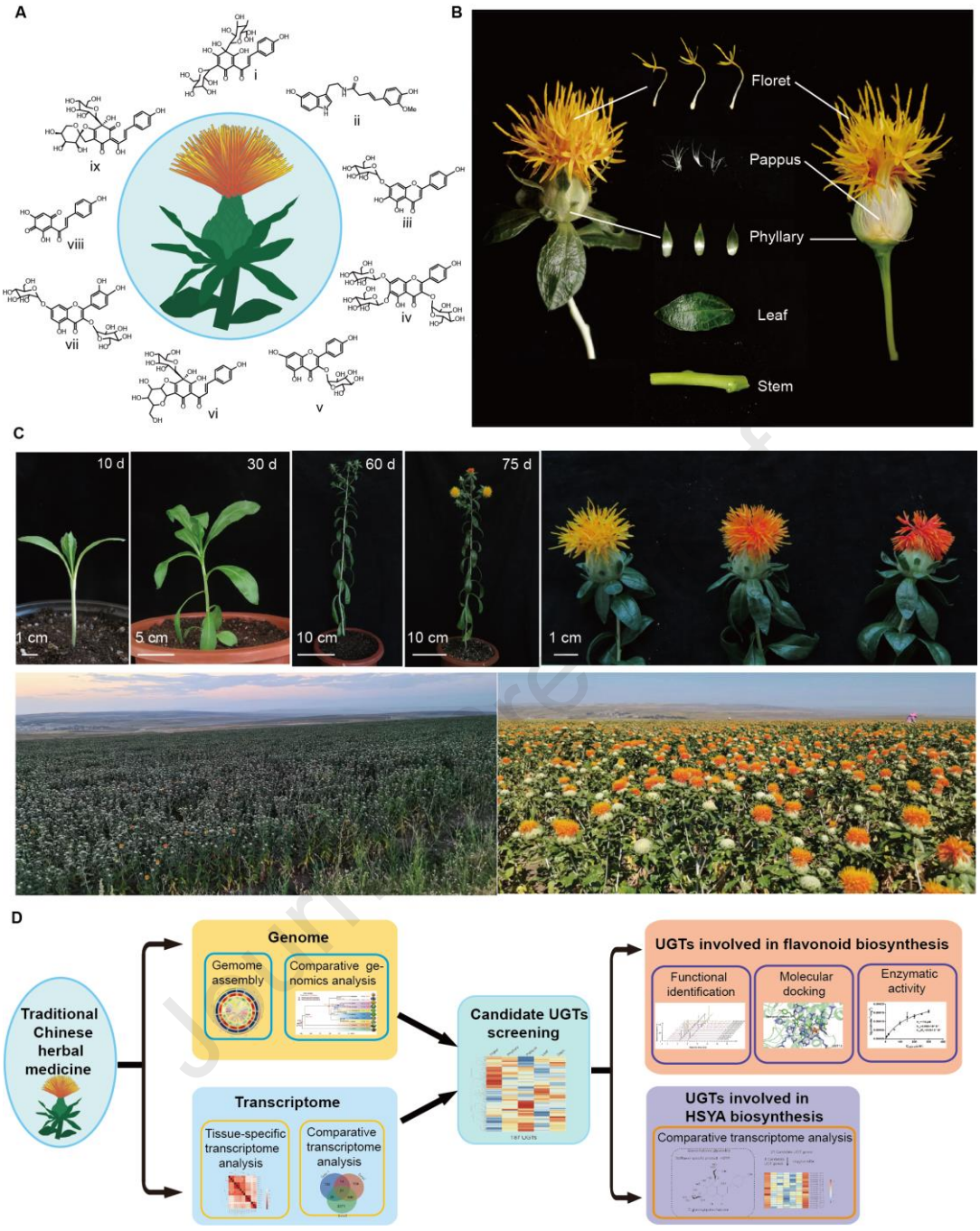


Fig. 1 The morphological and developmental characteristics of 'Yunhong3'

(A) Active ingredients in *C. tinctorius*. The following representative chemical structures are provided: (i) hydroxysafflor yellow A (HSYA); (ii) N-[2-(5-hydroxy-1H-indol-3-yl) ethyl] ferulamide; (iii) scutellarin; (iv) 6-hydroxykaempferol-3,6,7-*O*-*β*-D-glucoside; (v) kaempferol-3-*O*-*β*-D-glucoside; (vi) safflor yellow A; (vii) quercetin-3,7-*O*-*β*-D-glucoside; (viii) carthamone; (ix) saffloquinoside A. (B**)** Anatomical details of *C. tinctorius*, including floret, pappus, phyllary, leaf and stem. (C) Plant morphological characteristics of *C. tinctorius* at different stages and in the field**.** (D) The roadmap for identification of glycosyltransferases in flavonoid glycoside biosynthesis in safflower.

To generate a high-quality reference genome for 'Yunhong 3', we combined PacBio sequencing, and Bio-nano with Hi-C analyses. A total of 135 Gb of PacBio long reads, 475.5 Gb of Bio-nano reads, and 203.2 Gb of Hi-C data were generated, resulting in approximately 123.9×, 436.2×, and 186.4× coverage of the *C. tinctorius* genome,

respectively (Fig. S1, A and Table S1). The final assembled genome (1.09 Gb) consisted of 134 super scaffolds, with a scaffold N50 of 88.49 Mb and a contig N50 of 58.25 Mb (Fig. 2, A and Table 1). Subsequently, the assembled sequence was anchored to 12 pseudochromosomes (66.38–109.56 Mb), which accounted for 96.42% of the assembled genome sequence (Fig. S1, B and Table S2). The GC content of the *C. tinctorius* genome was 37.13% (Table S3). Based on Benchmarking Universal Single-Copy Orthologs (BUSCO), we evaluated the completeness of the genome assembly, which revealed that 92.8% of the plant sets in the *C. tinctorius* genome were complete (1 497 of 1 614 BUSCO). Accordingly, the genome described herein represented the most complete safflower genome assembled to date (Table S4) (Wu et al., 2021). The results of the *de novo* prediction analysis indicated that 65.15% of the assembled *C. tinctorius* genome comprised repetitive elements, of which long terminal repeat (LTR) retrotransposons being the most abundant, accounting for approximately 34.34% of the genome (Table S5). Specifically, LTR Gypsy and LTR Copia were identified as the two primary types, accounting for 28.12% and 16.61% of the genome, respectively.

Table 1 Details regarding the final *C. tinctorius* genome assembly and comparison with a published version

Feature	SafflowerRS1	*C. tinctorius*
Genome assembly		
Number of contigs	128	159
Contig N50 (Mb)	21.23	58.25
Longest contig (Mb)	57.98	104.11
Assembly size (Gb)	1.06	1.09
Repeat region of assembly (%)	60.13	66.96
Gene annotation		
Number of protein-coding genes	33 343	35 162
Average exons per gene	6.54	5.5
Mean exon length (bp)	269.59	300
Average CDS length (bp)	1 265.89	1 380
Reference	Wu et al., 2021	This study

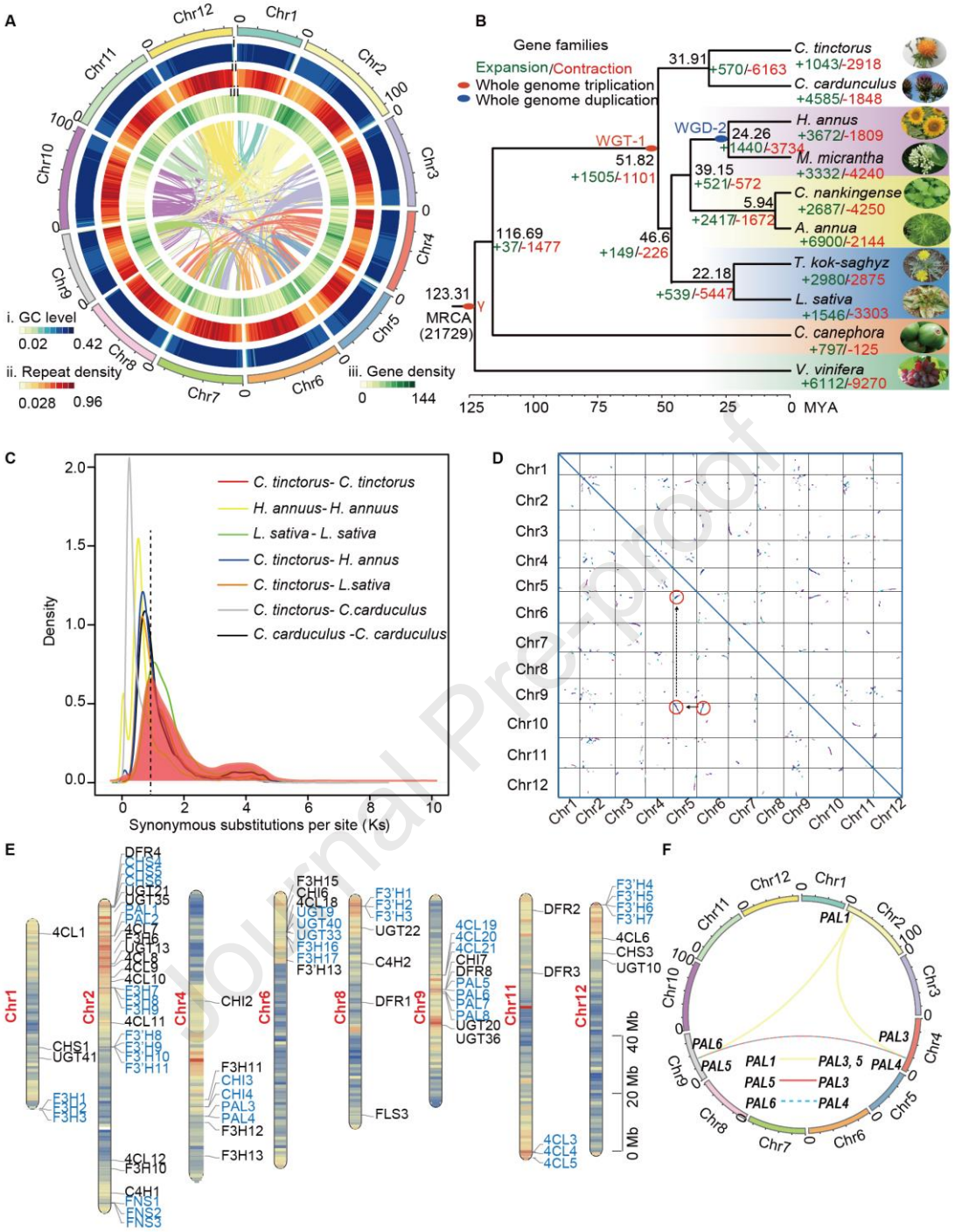


Fig. 2 Evolution of the *C. tinctorius* genome

(A) Genome characteristics and chromosomal locations. Landscape of the *C. tinctorius* genome (from outside to inside): chromosome number, GC content, repeat density, gene density, and genomic synteny. The GC content is presented in a 1 Mb window (color gradient indicates 0.02–0.42). The repeat density is presented in a 1 Mb window (color gradient indicates 0.028–0.96). The gene density is presented in a 1 Mb window (color gradient indicates 0–144). (B) Phylogenetic tree comprising 195 single-copy genes from 10 plant species. Gene family expansions and contractions are marked in green and red, respectively. Whole-genome duplication (WGD) and whole-genome triplication (WGT) events are marked on the tree. (C) Synonymous substitution rate (Ks) distributions in syntenic blocks in *C. tinctorius* and other species are marked with different colored lines. (D) Dot plots of paralogues in the *C. tinctorius* genome illustrating WGT (1-3 chromosomal relationships in red circles) events. (E) Chromosomal locations of flavonoid biosynthetic pathway genes affected by the tandem duplication event. The genes were marked in blue. (F) Distribution of PALs affected by WGD events. The links among chromosomes represent genes in the collinear blocks.

We subsequently employed an integrated strategy to predict the protein-coding genes in the *C. tinctorius* genome. A total of 35 162 genes were annotated, with an average coding sequence length of 1 380 bp and an average of 5.5 exons per transcript (**Table S6**). Additionally, 32,431 genes (92.23%) were functionally classified by a BLAST search of various databases (**Table S7**). We further annotated the noncoding RNA genes in the safflower genome, resulting in the identification of 212 microRNA genes, 1 277 transfer RNA genes, 4 709 ribosomal RNA genes, and 573 small nuclear RNA genes (Table S8).

3.2. Genome evolution and comparative genomic analysis

To explore the evolution of the safflower genome, we compared our assembled genome with those of nine plant species: *Cynara cardunculus* (globe artichoke), *Taraxacum kok-saghyz* (Russian dandelion), *Lactuca sativa* (lettuce), *Chrysanthemum nankingense* (chrysanthemum), *Artemisia annua* (sweet wormwood or Qinghao), *Helianthus annus* (sunflower), *Mikania micrantha* (climbing hempweed) from Asteraceae, *Coffea canephora* from Rubiaceae, and *Vitis vinifera* from Vitaceae (Fig. S2, A and Table S9). A total of 31 234 *C. tinctorius* genes were clustered into 15 274 gene families, which included 11 951 gene families shared by the five Asteraceae species (Fig. S2, B).We selected 195 single-copy genes from the aforementioned 10 species to construct a phylogenetic tree. The phylogenetic data showed that the eight Asteraceae species were clustered into one group, reflecting their close evolutionary relationships. In addition, *C. tinctorius* was most closely related to *C. cardunculus*, which was consistent with the similarities in their morphological characteristics. We estimated they diverged approximately 31.91 million years ago (Fig. 2, B).

We also compared the 10 plant species with their most recent common ancestor to analyze gene family expansions and contractions. In *C. tinctorius*, 1 043 gene families underwent expansion, while 2 918 gene families experienced contraction (Fig. 2, B). The main Gene Ontology (GO) terms assigned to the genes in the contracted families were nuclear nucleosome, plasmodesma (cellular component) (Fig. S3, A), protein heterodimerization activity, terpene synthase activity, and ATPase-coupled transmembrane transporter activity (molecular function) (Fig. S3, B), plant-type cell wall organization, geranyl diphosphate metabolic process, abscisic acid-activated signaling pathway (biological process) (Fig. S3, C). On the other hand, the expanded families were enriched in GO terms associated with chloroplast thylakoid membrane, photosystem II (cellular component) (Fig. S3, D), macromolecule metabolic process, primary metabolic process, DNA metabolic process (biological process) (Fig. S3, E), ADP binding, chlorophyll binding, transferase activity, and transferring hexosyl groups (molecular function) (Fig. S3, F). Furthermore, the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis revealed that the genes in the contracted families were associated with the ubiquitin system and plant hormone signal transduction, while the genes in the expanded families were related to energy metabolism and photosynthesis proteins (Fig. S3, G, H) (Kanehisa et al., 2000).

To explore the whole-genome expansion events that occurred during the evolution of *C. tinctorius*, we analyzed the synonymous substitution rate (Ks). Previous studies suggested that Asteraceae species shared a whole genome triplication (WGT-1) event (Badouin et al., 2017a; Shen et al., 2018; Song et al., 2018). The Ks

distribution of paralogous genes in safflower detected an obvious peak at 0.86 (approximately 52.1 million years ago), indicating a WGT event that also occurred in chrysanthemum, lettuce, artichoke, and other Asteraceae species (Fig. 2, C). The occurrence of this event was also supported by the triple collinearity in the safflower genome revealed by the chromosome scatter plot (Fig. 2, D). Notably, by analyzing the distribution and the collinear relationship of genes involved in the flavonoid biosynthetic pathway, we identified several genes affected by tandem duplication events and whole genome duplication (WGD) events, such as *F3H* (flavanone-3β-hydroxylase), *4CL* (4-coumarate: coenzyme A ligase), *F3'H* (flavonoid 3'-hydroxylase), and *PAL* (phenylalanine ammonia-lyase) (Fig. 2, E, F and Fig. S4, A). Subsequently, we performed a Ks analysis of flavonoid biosynthetic pathway genes to examine the associated duplication events in the safflower genome. We observed that the Ks distributions of these gene pairs were consistent with the timing of the γ-WGT and WGT-1 events (Fig. S4, B) (Badouin et al., 2017), indicating the positive impact of WGT on flavonoid biosynthesis in *C. tinctorius*.

3.3. Transcriptome analysis and sifting of transcription factors responsive to flavonoid biosynthesis

To comprehensively understand the expression pattern of genes involved in flavonoid glycoside biosynthesis, a transcriptome profiling was conducted using various safflower tissues, including floret, pappus, phyllary, leaf, and stem, collected from 1-year-old plants (Fig. 1, B and Fig. S5). We carried out a comparative transcriptome analysis and annotated these enzyme-encoding genes according to our high-quality genome. We divided the genes into the following three main groups: upstream (precursors), mid-downstream (intermediates), and downstream (final or near-final products). Interestingly, these genes had distinct expression patterns. Most of the upstream genes, such as *PAL*, *C4H* (cinnamate-4-hydroxylase), and *4CL*, were found to be highly expressed in the stem and phyllary but were expressed at low levels in the leaf. The *CHS* family genes were highly and usually specifically expressed in the floret. Furthermore, most of the mid-downstream genes, including *CHI*, *F3H*, and *FNS*, were more highly expressed in the floret or pappus than in the leaf and stem tissues, which was consistent with the tissue-specific distribution of flavonoids (Fig. 3).

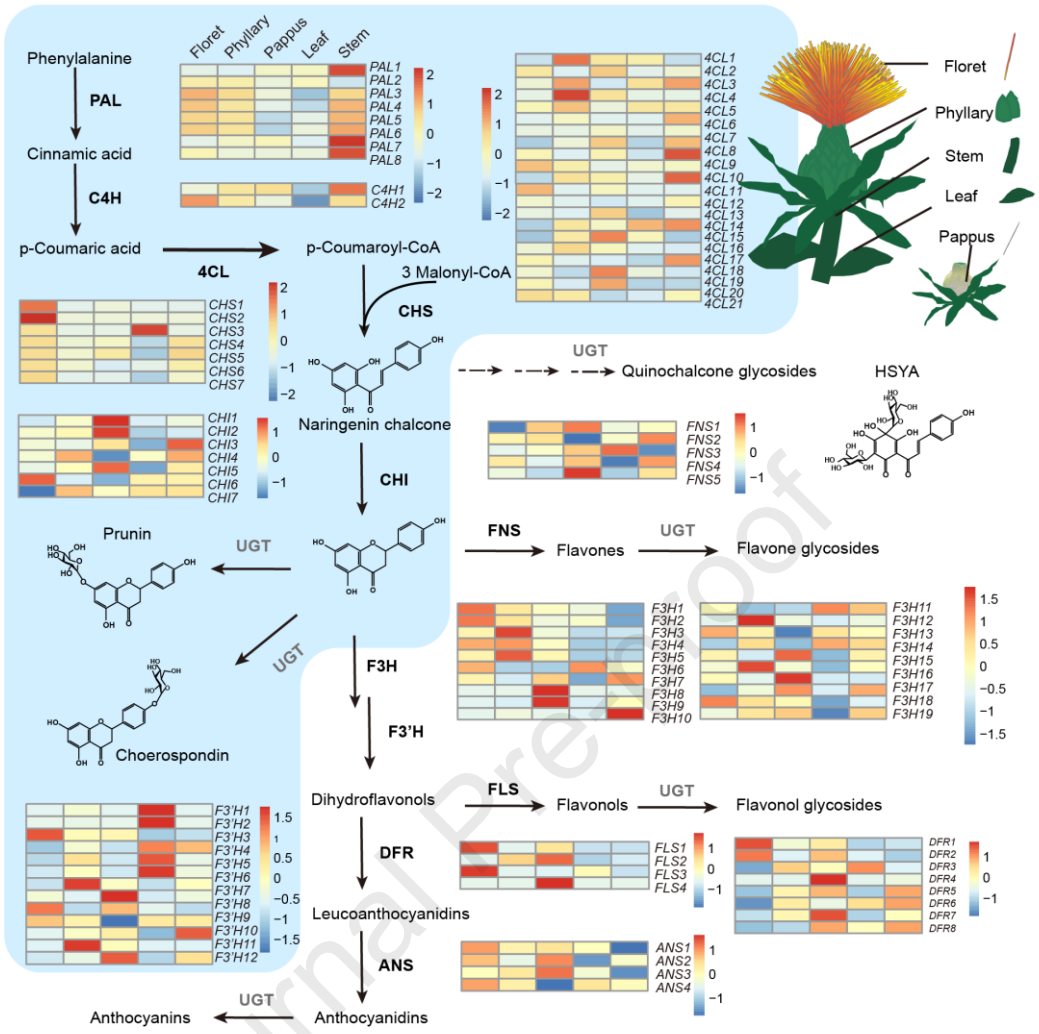


Fig. 3 Expression patterns of genes encoding key enzymes in the flavonoid biosynthetic pathway

In the gene expression heatmap, increases in the intensity of the red coloration reflect increases in the expression level, whereas increases in the intensity of the blue coloration reflect decreases in the expression level. PAL, phenylalanine ammonia-lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumarate: coenzyme A ligas; CHS, chalcone synthase; CHI, chalcone isomerase; FNS, flavone synthase; F3H, flavanone-3β-hydroxylase; F3'H, flavonoid 3'-hydroxylase; FLS, flavonol synthase; DFR, dihydroflavonol 4-reductase; ANS, anthocyanidin synthase; UGT, uridine diphosphate glycosyltransferase.

Given that transcription factors, such as MYB and basic helix-loop-helix (bHLH) family members, were important regulators of the flavonoid biosynthetic pathway in plants (Hichri et al., 2011; Ravaglia et al., 2013; Xu et al., 2014; Liu et al., 2019; Sun et al., 2020), we then analyzed transcription factors from different families in safflower to provide data support for further exploring their regulatory impacts. A total of 1,928 annotated transcription factors from 56 families were identified. Among them, AP2/ERF (Apetala2/Ethylene Response Factor), bHLH, FAR1 (FAR-RED Impaired Response 1), and MYB transcription factor families were the most represented (Table S10). Based on transcriptome data, we used gene expression values to construct the correlation network between flavonoid biosynthetic pathway genes and transcription factors (Fig. S6). The network diagram showed that MYB, bHLH, ERF, FAR1, C2H2-ZF (Cys2-His2 zinc-finger), and WRKY transcription factors had stronger relation with flavonoid biosynthetic pathway genes and might attribute as the predominant regulators. Previous research indicated that MYB transcription factors were also involved in the regulation of anthocyanin biosynthesis in other

medicinal plants (Zhou et al., 2015; Jian et al., 2019; Sun et al., 2020; Li et al., 2021). Intriguingly, during the colour-transition, which was often used as an indicator of the maturation of flower petals with medicinal glycosides, not only HSYA and carthamin showed a significant increase but also anthocyanins were detected a big difference (Ren et al., 2022).

Hence, anthocyanin synthesis and medicinal glycoside production may be correlated and MYB transcription factors as the "mediator" may play an important role in regulating the expression of these two compounds. We identified 148 *MYB* genes in total and according to the comparative transcriptomic analysis we screened out 30 genes that expressed at relatively higher levels in flowers (Table S11). We evaluated the correlations between these 30 *MYB* genes and the flavonoid and anthocyanin biosynthetic pathway genes by calculating their correlation coefficients, as shown in the regulatory network diagram (Fig. S7). Six genes (*MYB18*, *MYB2*, *MYB28*, *MYB26*, *MYB19*, and *MYB17*) had been found closely related to flavonoid and anthocyanin biosynthesis in the tubular flower, suggestive of the regulatory effects of the encoded transcription factors on tubular flower coloration and glycoside quality in *C. tinctorius*. To further explore the phylogenetic relationships among the 30 safflower *MYB* genes identified in this study and other known anthocyanin-related *MYB* genes, we constructed a neighbor-joining phylogenetic tree following the alignment of multiple protein sequences (Fig. S8 and Table S12). From the phylogenetic tree, we found that *CtMYB2* and *CtMYB10* clustered with *AtMYB112* (Lotkowska et al., 2015), a regulator that promotes anthocyanin accumulation in *Arabidopsis*; while *CtMYB28* had a closer relationship with *PnMYB27*, a suppressor of anthocyanin pigmentation. Apart from this, we found one clade that was specific to *C. tinctorius* containing *CtMYB16*, *CtMYB17*, and *CtMYB18*. Overall, we provided these useful candidate gene data and more experimental verification would be performed in the future.

3.4. Identification of UGT genes involved in flavonoid biosynthesis

The fact that most flavonoids exist as glycosides implies that UGT*s* play a critical role in the flavonoid biosynthetic pathway. Consequently, we conducted an in-depth exploration of the highly expressed *UGT* genes in floral tissues, leading to the identification of 187 highly expressed full-length candidate *UGT* genes (Fig. 4, A). Subsequently, a phylogenetic analysis was performed (Fig. S9 and Table S13) by aligning the encoded amino acid sequences with the reported functional genes from *A. thaliana*, *Medicago truncatula*, *Glycine max*, *Vitis riparia*, and other species. All of the selected genes have glycosylation functions, and some of them are flavonoid glycosyltransferase genes. These *UGT* genes were classified into different subfamilies, including previously reported ones, such as *UGT71* (Dong et al., 2014), *UGT72* (Lim et al., 2005), *UGT75* (Sun et al., 2018), *UGT79* (Li et al., 2017), and *UGT83* (Modolo et al., 2007). This preliminary classification of *UGT* genes was not only useful for a comprehensive understanding of the UGT family in safflower, but was also used to select candidate genes that might have the function of flavonoid glycosylation. Considering the properties of the glycosyltransferase, such as substrate promiscuity and functional diversity, a local BLAST search was additionally conducted based on all annotated UGT genes in the safflower genome, using previously reported flavonoid biosynthesis related *UGT* genes as queries to identify homologous candidate genes. Combining the tissue properties of flavonoids in

safflower, we then selected the genes that were predominantly expressed in florets according to our tissue-specific transcriptome profiling data. Finally, 34 genes were narrowed down as our candidate genes for further analysis. (Fig. 4, A and Table S14).

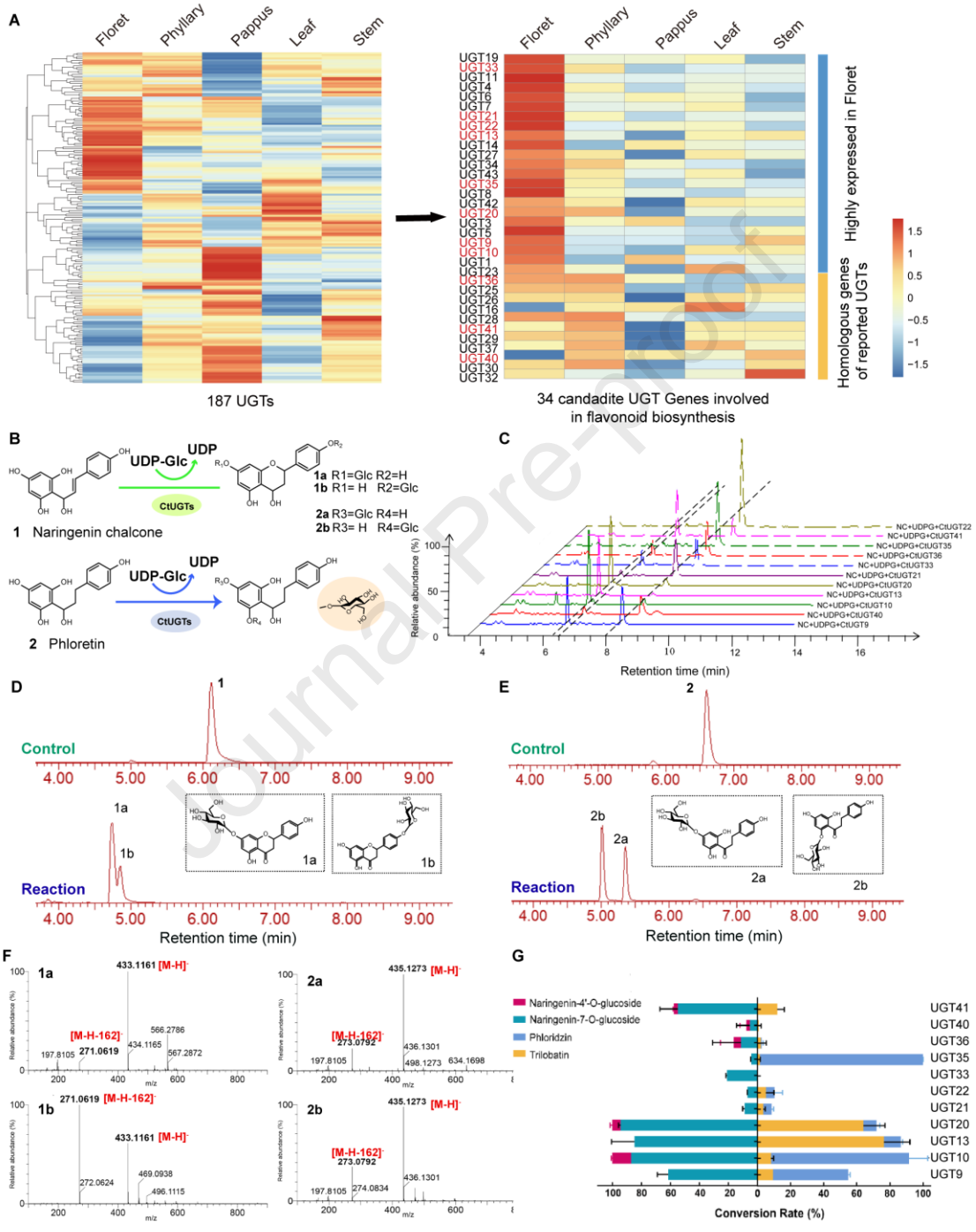


Fig. 4 Functional characterization of the recombinant UGTs involved in the flavonoid biosynthetic pathway

(A) Candidate *UGT* genes involved in flavonoid biosynthesis. A total of 187 *UGT* genes were identified in the safflower genome, of which 23 *UGT* genes were highly expressed in the floret. Additionally, 11 homologs of *UGT* genes reported involved in flavonoid biosynthesis were screened out as the flavonoid biosynthesis-related candidate genes in safflower. (B) UGTs catalyzed the O-glycosylation of naringenin chalcone (1) and phloretin (2) in this study. (C) UPLC chromatograms of the enzymatic reactions of 11 UGTs, using naringenin chalcone as the substrate. (D) UPLC-Q-TOF/MS analysis of naringenin chalcone and the

products of the enzymatic reactions (1a, 1b). The chemical structures of the products are presented in the dashed boxes. The UPLC chromatograms were recorded at 290 nm. (E) UPLC-Q-TOF/MS analysis of phloretin and the products of the enzymatic reactions (2a, 2b). The chemical structures of the products are presented in dashed boxes. The UPLC chromatograms were recorded at 290 nm. (F) Molecular ion peak extracted from the products of the enzyme-catalyzed reactions (1a, 1b, 2a, 2b). (G) Conversion rates (%) for the glycosylated products for different UGTs, using UDPG as the sugar donor. The conversion rates were calculated based on the UPLC peak area ratio. All experiments were performed in triplicate (n = 3).

To identify the functional UGTs, we expressed and purified soluble recombinant UGT proteins using *Escherichia coli* and conducted enzyme assays (Fig. S10). We used naringenin chalcone (Fig. 4, B), a vital substrate for flavonoid glycoside biosynthesis, for the in *vitro* enzymatic reactions. The catalytic products were extracted and analyzed using an ultra-performance liquid chromatography-quadrupole-time of flight mass spectrometry (UPLC-Q-TOF/MS) system. We identified 11 novel UGTs that could catalyze the conversion of naringenin chalcone to a new compound 1a (Fig. 4, B, C, D, and F). The genes encoding these functional safflower UGTs were named as follows: *CtUGT9*, *CtUGT10*, *CtUGT13*, *CtUGT20*, *CtUGT21*, *CtUGT22*, *CtUGT33*, *CtUGT35*, *CtUGT36*, *CtUGT40*, and *CtUGT41* (Fig. S12, A, C). Product 1a was purified from a preparative-scale reaction and identified as prunin (naringenin-7-O-glucoside) by an NMR (Nuclear Magnetic Resonance) analysis and a comparison with an authentic reference standard (Fig. S13, A and Fig. S14). Notably, CtUGT10, CtUGT20, CtUGT36, CtUGT40, and CtUGT41 were revealed to add a glucosyl group to the hydroxyl group at the C4′ position of naringenin chalcone to generate 1b, which was identified as choerospondin (naringenin-4′-O-β-glucopyranoside) based on a comparison with an authentic reference standard (Fig. 4, B, C, D, F, Fig. S12, A, D and Fig. S13, A). Intriguingly, when naringenin chalcone was employed as the substrate in the reaction, it initially underwent a spontaneous cyclization in the buffer, which converted the chalcone skeleton into a dihydroflavone in a reaction reportedly catalyzed by CHI in plants (Liu et al., 2015; Elarabi et al., 2021).

We then used the common flavonoid precursor phloretin (Fig. 4, B), which was structurally similar to naringenin chalcone, as a substrate to identify whether these enzymes had the same catalytic functions. Surprisingly, 10 of the 11 UGTs (i.e., all except CtUGT33) catalyzed the addition of glucose to the hydroxyl group at the C4′ position of the dihydrochalcone skeleton to generate 2a, which was identified as trilobatin (phloretin-4′-O-glucoside) following a comparison with the standards (Fig. 4, B, E, and F and Fig. S12, B, Fig. S13, B, Fig. S13, C and Fig. S12, F). Some of the UGTs catalyzed reactions that modified phloretin at two positions. Furthermore, CtUGT9, CtUGT10, CtUGT13, CtUGT20, CtUGT21, CtUGT22, CtUGT35, and CtUGT36 also mediated the addition of glucose to the hydroxyl group at the C2′ position to generate 2b, which was identified as phlorizin (phloretin-2′-O-glucoside) based on a comparison with an authentic reference standard (Fig. 4, B, D, F and Fig. S12, B, E and Fig. S13, B, C). The catalytic functions of these UGTs are reminiscent of the functions of F7GAT in *Erigeron breviscapus* (Liu et al., 2018) and OcUGT1 in *Ornithogalum caudatum* (Yuan et al., 2018), but these are the first reported safflower UGTs. Overall, these enzymes can glycosylate multiple sites (e.g., 7,4′-OH of the flavonoid skeleton and 2′,4′-OH of the chalcone skeleton), suggesting that different glycosyltransferases possess diverse catalytic activities, allowing them to modify specific substrates. This diversity may be related to the active pockets and binding sites of proteins.

3.5. Molecular docking and enzymatic kinetic analyses of UGTs

The molecular docking approach has emerged as an increasingly vital and effective computational strategy for drug discovery due to its ability to model the interaction between a small molecule and a protein at the atomic level (McConkey et al., 2002). To characterize the substrate behavior after binding to UGTs and to elucidate fundamental biochemical processes, we selected CtUGT9, CtUGT10, CtUGT13, and CtUGT20 for the molecular docking analysis (Fig. S10, B) because of their high activities (i.e., glycosylation of naringenin chalcone and phloretin; Fig. 4, G). Naringenin chalcone was used as the feeding substrate and uridine diphosphate glucose (UDPG) served as the sugar donor. During the docking analysis, naringenin chalcone and UDPG were detected in the binding pocket of CtUGT10 and CtUGT13 (Fig. 5, A, B). The amino acid residues surrounding the substrate determined the shape of the binding pocket. The docking results for CtUGT10 revealed that Trp332, Cys333, Leu336, and Ser355 of the plant secondary product glycosyltransferase (PSPG) box formed hydrogen bonds with naringenin chalcone, whereas His350, Trp353, Asp374, and Gln375 formed hydrogen bonds with UDPG (Fig. 5, A). Previous studies have reported that a conserved histidine that was often present in the active site of UGTs was positioned close to the glucose moiety of the sugar donor and the acceptor (Wang, 2009). A multiple sequence alignment identified His19 of CtUGT10 as a conserved histidine among the different UGTs **(**Fig. S11**)**. Furthermore, a hydrogen bond between His19 and UDPG was detected. Similarly, in CtUGT13, Trp362, Asn363, and Ser364 of the PSPG box formed hydrogen bonds with naringenin chalcone, while Ala342, Gln344, Ala345, and Glu367 formed hydrogen bonds with UDPG (Fig. 5, B). Additionally, many amino acids in the PSPG box of CtUGT9 and CtUGT20 were observed to form hydrogen bonds with the two substrates (Fig. S15). These amino acids surrounding the binding site may be crucial for catalytic activities and could be targeted and modified to optimize UGT catalytic functions.

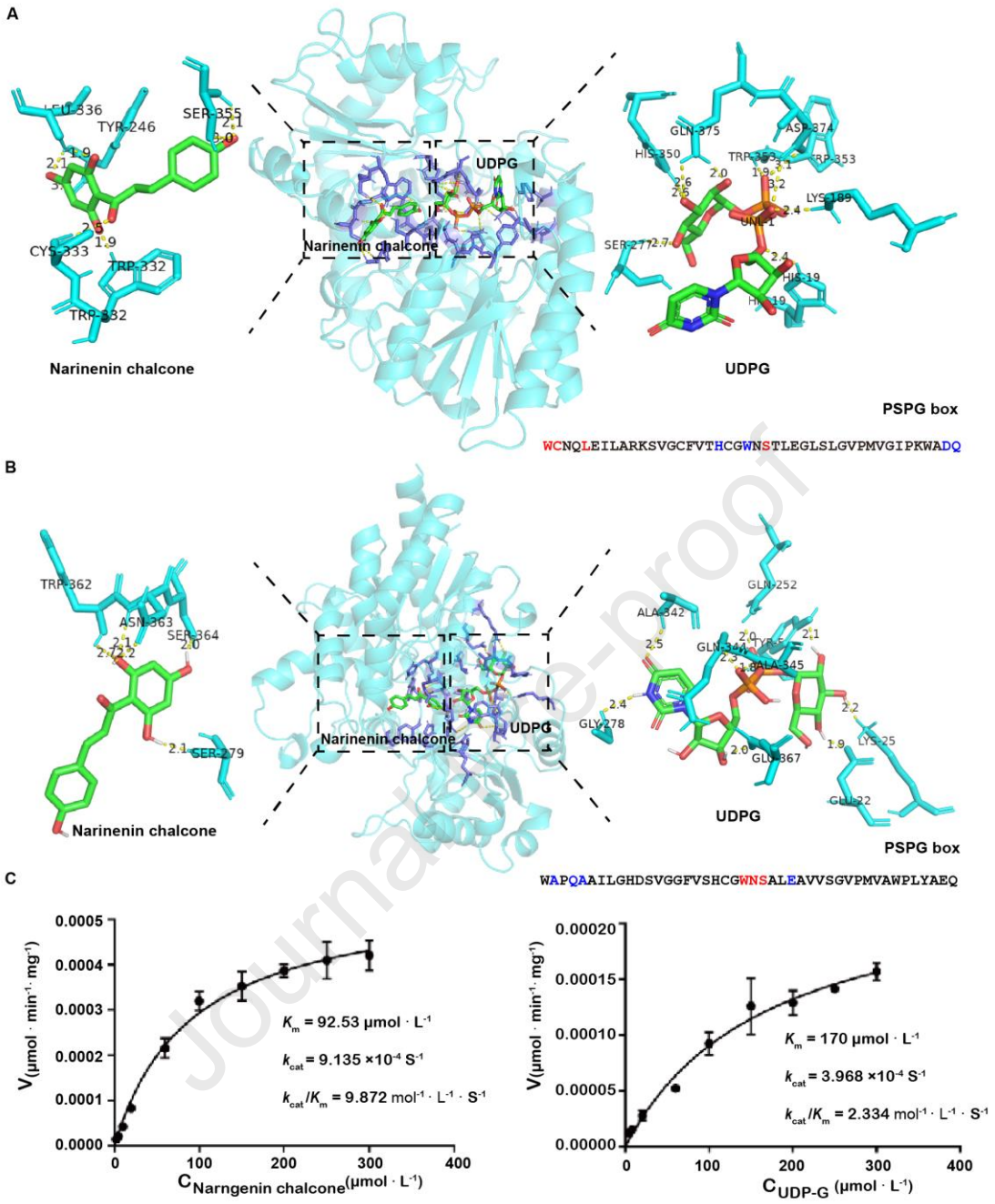


Fig. 5 Molecular docking and enzymatic kinetic analyses of UGTs with catalytic functions

(A) CtUGT10 (using 5u6m as a template) binds naringenin chalcone at the active site for the addition of glucose. The CtUGT10 protein structure is presented, with the residues around the substrate (within 3 Å) marked by violet sticks. Hydrogen bonds are indicated by a yellow dashed line, with distances marked in black. The residues in the PSPG box binding to naringenin chalcone are in red, whereas the residues binding to UDPG are in blue. (B) CtUGT13 (using 7q3s as the template) binds naringenin chalcone at the active site for the addition of glucose. (C) Enzymatic kinetic parameter investigation for CtUGT10. The K_m value was calculated using naringenin chalcone and UDPG as the substrate and sugar donor, respectively.

From our analyses of CtUGT9, CtUGT10, CtUGT13, and CtUGT20, CtUGT10 emerged as the most active enzyme (i.e., up to 100% and 84% glycosylated naringenin chalcone and phloretin, respectively, according to the UPLC analysis). Furthermore, it catalyzed the reactions leading to all four products (Fig. 4, G). These findings prompted us to select it for the subsequent enzymatic kinetic analysis. We expressed and purified recombinant CtUGT10 proteins for the in *vitro* enzymatic kinetic assay, using naringenin chalcone as the substrate and UDPG as

the sugar donor. The recombinant CtUGT10 protein exhibited the highest activity at pH 8.0 (50 mmol · L^{-1} Tris-HCl buffer) and 37 °C. Its activity did not require divalent metal ions (Fig. S16). The kinetic parameters of CtUGT10 were calculated. The maximum reaction rates for naringenin chalcone and UDPG were 5.64×10^{-4} and 2.45×10^{-4} μmol · min^{-1} with K_m and k_{cat} values of 92.53 μmol · L^{-1} and 9.14×10^{-4} s^{-1}, and 170 μmol · L^{-1} and 3.97×10^{-4} s^{-1}, respectively, reflecting the adequate affinity of CtUGT10 for naringenin chalcone and UDPG (Fig. 5, C).

3.6. Identification of key genes involved in HSYA biosynthesis

Hydroxysafflor yellow A is the most medicinally valuable flavonoid glycoside in safflower, but there are a few reports on glycosyltransferase involved in the biosynthesis of HSYA (Xie et al., 2014). The plant hormone methyl jasmonate (MeJA) promotes the production of quinone chalcones, such as HSYA, by up-regulating the expression of the upstream genes in the flavonoid biosynthetic pathway and down-regulating the expression of specific downstream genes (Chen et al., 2020). We conducted a comparative transcriptomic analysis by integrating the publicly available RNA-seq data for safflower flowers treated with or without MeJA with our tissue-specific transcriptome data to screen for candidate genes encoding proteins with key functions associated with HSYA formation. In the initial step of the HSYA biosynthetic pathway, PAL catalyzes the conversion of phenylalanine to cinnamic acid. The expression levels of seven of the eight *PAL* genes annotated in our safflower genome were up-regulated after the MeJA treatment (Fig. 6, A). Notably, seven of the 21 annotated *4CL* genes, which encoded enzymes that converted cinnamic acid to p-coumaroyl CoA, had MeJA-induced up-regulated expression levels. Moreover, four of seven *CHS* genes, which were responsible for the conversion of p-coumaroyl CoA and three malonyl-CoA molecules to naringenin chalcone, had MeJA-induced up-regulated expression levels.

A recent study reported that the HSYA concentration in safflower inflorescences increased substantially in response to exogenous MeJA (Chen et al., 2020), which compelled us to perform an in-depth comparative analysis of *UGT* genes. Of the examined genes, the expression levels of 21 candidate *UGT* genes were up-regulated by a MeJA treatment (2-fold higher than the control expression level) and were higher in the florets than in the other tissues [$\log_2$(fold-change) ≥ 1] (Fig. 6, B). Furthermore, four candidate *UGT* genes encoded proteins that were observed to catalyze O-glycosylations rather than C-glycosylations in this study. Therefore, we finally identified 17 candidate UGTs that may catalyze C-glycosylations during the production of HSYA (Fig. 6 and Table S15).

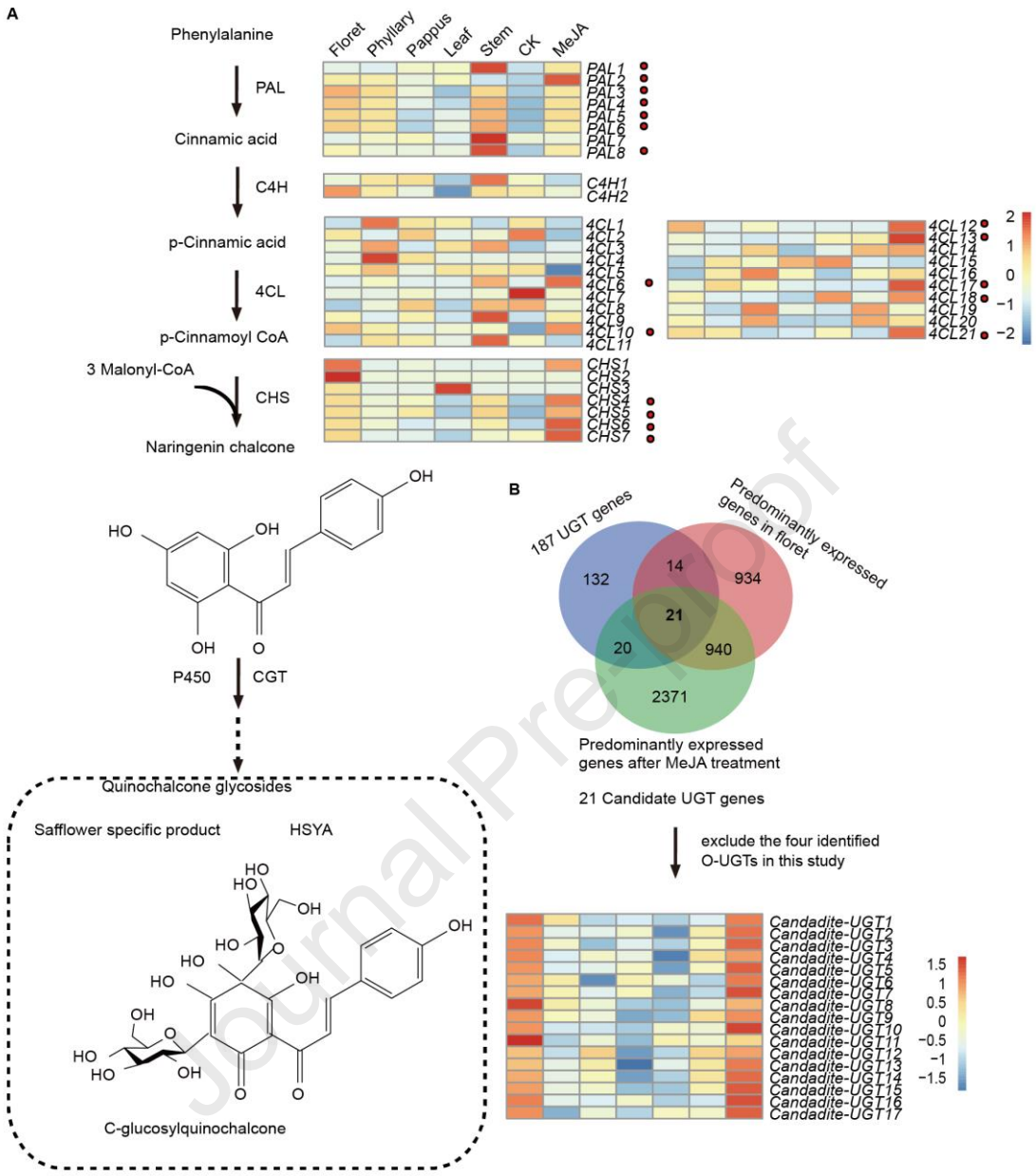


Fig. 6 Expression patterns of candidate genes encoding key enzymes in the HSYA biosynthetic pathway

(A) In the gene expression heatmap, increases in the intensity of the red coloration reflect increases in the expression level, whereas increases in the intensity of the blue coloration reflect decreases in the expression level. CK represents the flower treated without MeJA, whereas MeJA represents the flower treated with MeJA. Genes marked with a red circle encode the key enzymes in the HSYA biosynthetic pathway. PAL, phenylalanine ammonia-lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumarate: coenzyme A ligas; CHS, chalcone synthase. (B) Venn diagram of the overlap genes among different gene sets.

4. Discussion

Safflower is one of the oldest known cultivated crops. It was first used as a clothing and food dye and now a cooking oil and an important Chinese herb. The dried tubular corolla derived from its flower has been used in traditional Chinese medicine for thousands of years (Delshad et al., 2018). Modern pharmacological research demonstrated that flavonoids contributed to its significant clinical activities. Notably, the most representative compound HSYA is potentially useful for medical research and the development of therapeutic agents. However,

the biosynthesis and regulation of these flavonoids remain relatively uncharacterized in safflower. And the lack of comprehensive molecular genetics and gene-editing systems hampered the development of safflower. In this study, we conducted a multi-omics analysis to clarify the biosynthesis of flavonoids and generate data regarding valuable medicinal safflower germplasm resources. Using high-throughput sequencing technology, we constructed a high-quality chromosome-level reference genome for 'Yunhong3', which was an excellent medicinal safflower variety, as well as a tissue-specific transcriptome. Based on the substantial abundance of genetic data produced in this study, we conducted a comparative genomic analysis to explore the phylogenetic relationships between safflower and other plant species as well as the effects of evolutionary processes on flavonoid biosynthesis. A WGT event was shared between safflower and other Asteraceae plants, which positively affected the flavonoid biosynthesis in safflower. Our presented findings may be used as supporting material in future systematic investigations of the evolution and divergence of Asteraceae plant species.

In addition, we compared the expression patterns of pathway genes in different tissues and found that partial copies of *PAL*, *C4H* and *4CL* genes in the upstream of the pathway were highly expressed in stem, leaf, bract and pappus, and a few were highly expressed in flowers, but *CHS*, *CHI*, *FNS*, *F3H*, and *FLS* genes were specifically expressed in flowers, including some post-modification *UGT* genes. Based on the above results, we speculated that the precursors of flavonoids might be firstly synthesized in the stems or leaves and then transferred to the flowers or finally modified in flowers, which needed further research contents to support. Besides the pathway genes, we were also interested in the transcription factors of safflower and focused on *MYB* family genes for analysis. Many reports suggest that MYB transport factors regulate the biosynthesis of secondary metabolites, for example, tanshinones and phenolic acid from *Salvia miltiorrhiza* (Deng et al., 2020; Zhou et al., 2021; Liu et al., 2022; Liu et al., 2023). By comparing the expression levels of genes in different tissues and their correlations with flavonoids and anthocyanin biosynthesis pathway genes, a series of candidate *MYB* genes were screened out, which might be involved in the formation of safflower color and active ingredients. We focused on elucidating the process leading to flavonoid formation in safflower, particularly the key glycosyltransferase genes. Based on our annotated genome and the tissue-specific transcriptome data, we first identified 187 *UGT* genes. By conducting BLAST analysis using published functional UGTs as queried and selecting highly expressed genes in florets, we narrowed down the initial 187 *UGT* genes to 34 candidate genes. The full-length sequences of these *UGT* genes were used to express recombinant proteins in *E. coli* for in *vitro* enzyme assays. Mass spectrometry and NMR data showed that 11 of these novel UGTs catalyzed the production of prunin (naringenin-7-O-glucoside) from naringenin chalcone. In addition, five of these 11 UGTs also produced another glycoside, choerospondin (naringenin-4'-O-β-glucopyranoside). Notably, the spontaneous cyclization of the chalcone backbone was observed during this process, which was consistent with the reported structural changes for other similar reactions, such as isoflavone glucosyltransferase from *Pueraria thomsonii* (Duan et al., 2022), indicative of the complexity and unpredictability of flavonoid biosynthesis. When phloretin was used as a substrate, 10 of the 11 UGTs catalyzed the production of trilobatin (phloretin-4'-O-glucoside). Moreover, eight UGTs also produced

another glycoside, phlorizin (phloretin-2'-O-glucoside), reflecting the diversity in enzyme functions, which explained the structural variability among glycosidesin safflower. To further explore the catalytic mechanism of the functional UGTs, we selected four highly active enzymes for simulated molecular docking experiments, which identified candidate amino acid residues that may form key active sites for substrates or glycosyl donors. The generated molecular-level data provide the foundation for future studies on directed evolution and enzyme modifications.

To further explore the enzymes involved in the biosynthesis of HSYA, we conducted a comparative transcriptomic analysis of our data and publicly available data regarding the effects of MeJA. We screened a series of genes that may contribute to the formation of HSYA (e.g., 3 *4CL* genes, 7 *CHS* genes, and 21 *UGT* genes). These genes were evaluated in terms of whether their expression was up-regulated by an exogenous application of MeJA. Four of these 21 candidate *UGT* genes were determined in this study. Therefore, we identified *UGT* genes responsible for the biosynthesis of flavonoids in safflower, while also screened out promising candidate *UGT* genes involved in the HSYA biosynthetic pathway. The gene resources described herein could be used in future research on the sustainable development of safflower (e.g., synthetic biology and the breeding and cultivation of high-quality varieties). The strategy used in this study will also give examples in the investigation of natural active ingredients from other valuable medicinal plants.

Data availability

The raw sequencing data were submitted to Genome Sequence Archive (GSA) under the accession number PRJCA005368. The safflower genome, gene annotation, and nucleotide and protein sequences were stored in the Genome Warehouse (GWH) under the same accession number.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the ability establishment of sustainable use for valuable Chinese medicine resources (Grant No. 2060302), the National Key R&D Program of China (Grant No. 2018YFA0900603), the National Key R&D Program of China (Grant No. 2020YFA0908000) and the crosswise task based on DEYUANTANG pharmacy Co., Ltd. Shanxi, China (Grant No. DYTKY180725).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:

References

Albert, N.W., Davies, K.M., Lewis, D.H., Zhang, H., Montefiori, M., Brendolise, C., Boase, M.R., Ngo, H., Jameson, P.E., and Schwinn, K.E., 2014. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. Plant Cell, 26: 962-980.

Albert, N.W., Lewis, D.H., Zhang, H., Schwinn, K.E., Jameson, P.E., Davies, K.M., 2011. Members of an R2R3-MYB transcription factor family in Petunia are developmentally and environmentally regulated to control complex floral and vegetative pigmentation patterning. Plant J, 65: 771-784.

Alseekh, S., Perez de Souza, L., Benina, M., Fernie, A.R., 2020. The style and substance of plant flavonoid decoration; towards defining both structure and function. Phytochemistry, 174: 112347.

Ao, H., Feng, W., Peng, C., 2018. Hydroxysafflor yellow a: a promising therapeutic agent for a broad spectrum of diseases. Evid Based Complement Alternat Med, 2018: 8259280.

Aoki, K., Yano, K., Suzuki, A., Kawamura, S., Sakurai, N., Suda, K., Kurabayashi, A., Suzuki, T., Tsugane, T., Watanabe, M., Ooga, K., Torii, M., Narita, T., Shin-I, T., Kohara, Y., Yamamoto, N., Takahashi, H., Watanabe, Y., Egusa, M., Kodama, M., Shibata, D., 2010. Large-scale analysis of full-length cDNAs from the tomato (*Solanum lycopersicum*) cultivar Micro-Tom, a reference system for the Solanaceae genomics. BMC Genomics, 11: 210

Asgarpanah, J., Kazemivash, N., 2013. Phytochemistry, pharmacology and medicinal properties of *Carthamus tinctorius L.* Chin J Integr Med, 19: 153-159.

Badouin, H., Gouzy, J., Grassa, C.J., Murat, F., Staton, S.E., Cottret, L., Lelandais-Briere, C., Owens, G.L., Carrere, S., Mayjonade, B., Legrand, L., Gill, N., Kane, N.C., Bowers, J.E., Hubner, S., Bellec, A., Berard, A., Berges, H., Blanchet, N., Boniface, M.C., Brunel, D., Catrice, O., Chaidir, N., Claudel, C., Donnadieu, C., Faraut, T., Fievet, G., Helmstetter, N., King, M., Knapp, S.J., Lai, Z., Le Paslier, M.C., Lippi, Y., Lorenzon, L., Mandel, J.R., Marage, G., Marchand, G., Marquand, E., Bret-Mestries, E., Morien, E., Nambeesan, S., Nguyen, T., Pegot-Espagnet, P., Pouilly, N., Raftis, F., Sallet, E., Schiex, T., Thomas, J., Vandecasteele, C., Vares, D., Vear, F., Vautrin, S., Crespi, M., Mangin, B., Burke, J.M., Salse, J., Munos, S., Vincourt, P., Rieseberg, L.H., Langlade, N.B., 2017. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. Nature, 546: 148-152.

Borevitz, J.O., Xia, Y.J., Blount, J., Dixon, R.A., Lamb, C., 2000. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell, 12: 2383-2393.

Brazier-Hicks, M., Edwards, R., 2005. Functional importance of the family 1 glucosyltransferase UGT72B1 in the metabolism of xenobiotics in *Arabidopsis thaliana*. Plant J, 42: 556-566.

Brazier-Hicks, M., Evans, K.M., Gershater, M.C., Puschmann, H., Steel, P.G., Edwards, R., 2009. The C-glycosylation of flavonoids in cereals. J Biol Chem, 284: 17926-17934.

Brazier-Hicks, M., Gershater, M., Dixon, D., Edwards, R., 2018. Substrate specificity and safener inducibility of the plant UDP-glucose-dependent family 1 glycosyltransferase super-family. Plant Biotechnol J, 16: 337-348.

Campbell, M.A., Haas, B.J., Hamilton, J.P., Mount, S.M., Buell, C.R., 2006. Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. BMC Genomics, 7: 327-328.

Chen, D.W., Sun, L.L., Chen, R.D., Xie, K.B., Yang, L., Dai, J.G., 2016. Enzymatic synthesis of acylphloroglucinol 3-C-glucosides from 2-O-glucosides using a C-glycosyltransferase from *Mangifera indica*. Chem Eur J 22: 5873-5877.

Chen, D., Chen, R., Wang, R., Li, J., Xie, K., Bian, C., Sun, L., Zhang, X., Liu, J., Yang, L., Ye, F., Yu, X., Dai, J., 2015. Probing the catalytic promiscuity of a regio- and stereospecific C-glycosyltransferase from *Mangifera indica*. Angew Chem Int Ed Engl, 54: 12678-12682.

Chen, J., Wang, J., Wang, R., Xian, B., Ren, C., Liu, Q., Wu, Q., Pei, J., 2020 Integrated metabolomics and transcriptome analysis on flavonoid biosynthesis in safflower (*Carthamus tinctorius L.*) under MeJA treatment. BMC Plant Biol, 20: 353.

Chung, S.Y., Seki, H., Fujisawa, Y., Shimoda, Y., Hiraga, S., Nomura, Y., Saito, K., Ishimoto, M., Muranaka, T., 2020. A cellulose synthase-derived enzyme catalyses 3-O-glucuronosylation in saponin biosynthesis. Nat Commun, 11: 5664.

Dai, X., Zhuang, J., Wu, Y., Wang, P., Zhao, G., Liu, Y., Jiang, X., Gao, L., Xia, T., 2017. Identification of a flavonoid glucosyltransferase involved in 7-OH site glycosylation in tea plants (*Camellia sinensis*). Sci Rep, 7: 5926.

Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., Flouri, T., 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol, 37: 291-294.

De Bie, T., Cristianini, N., Demuth, J.P., Hahn, M.W., 2006. CAFE: a computational tool for the study of gene family evolution. Bioinformatics, 22: 1269-1271.

Deng, C., Wang, Y., Huang, F., Lu, S., Zhao, L., Ma, X., Kai, G., 2020. SmMYB2 promotes salvianolic acid biosynthesis in the medicinal herb *Salvia miltiorrhiza*. J Integr Plant Biol, 62: 1688–1702.

Delshad, E., Yousefi, M., Sasannezhad, P., Rakhshandeh, H., Ayati, Z., 2018. Medical uses of *Carthamus tinctorius* L. (Safflower): a comprehensive review from

Traditional Medicine to Modern Medicine. Electron Physician, 10: 6672-6681.

Di, S., Yan, F., Rodas, F. R., Rodriguez, T. O., Murai, Y., Iwashina, T., Sugawara, S., Mori, T., Nakabayashi, R., Yonekura-Sakakibara, K., Saito, K., Takahashi, R., 2015. Linkage mapping, molecular cloning and functional analysis of soybean gene Fg3 encoding flavonol 3-O-glucoside/galactoside (1→2) glucosyltransferase. BMC Plant Biol. 15: 126.

Dong, T., Xu, Z.Y., Park, Y., Kim, D.H., Lee, Y., Hwang, I., 2014. Abscisic acid uridine diphosphate glucosyltransferases play a crucial role in abscisic acid homeostasis in *Arabidopsis*. Plant Physiol, 165: 277-289.

Dordas, C.A., Sioulas, C., 2008. Safflower yield, chlorophyll content, photosynthesis, and water use efficiency response to nitrogen fertilization under rainfed conditions. Ind Crops Prod, 27: 75-85.

Duan, H.Y., Wang, J., Zha, L.P., Peng, H.S., Zhao, Y.P., Yuan, Y., Huang, L.Q., 2022. Molecular cloning and functional characterization of an isoflavone glucosyltransferase from *Pueraria thomsonii*. Chin J Nat Med, 20: 133-138.

Du, H.L., Liang, C.Z., 2019. Assembly of chromosome-scale contigs by efficiently resolving repetitive sequences with long reads. Nat Commun, 10: 5360.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., Aiden E.L., 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science, 356: 92-95.

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., Aiden, E.L., 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst, 3: 95-98.

Ekin, Z., 2005. Resurgence of safflower (*Carthamus tinctorius* L.) utilization: a global view. J Agron 4: 83-87.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., Finn, R.D., 2019. The Pfam protein families database in 2019. Nucleic Acids Res, 47: D427-D432.

Elarabi, N.I., Abdelhadi, A.A., Sief-Eldein, A.G.M., Ismail, I.A., Abdallah, N.A., 2021. Overexpression of chalcone isomerase a gene in *Astragalus trigonus* for stimulating apigenin. Sci Rep, 11: 24176.

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol, 16: 157.

Espley, R. V., Brendolise, C., Chagné, D., Kutty-Amma, S., Green, S., Volz, R., Putterill, J., Schouten, H. J., Gardiner, S. E., Hellens, R. P., Allan, A. C., 2009. Multiple repeats of a promoter segment cause transcription factor autoregulation in red apples. Plant Cell, 21: 168-183.

Faust, B., Hoffmeister, D., Weitnauer, G., Westrich, L., Haag, S., Schneider, P., Decker, H., Kunzel, E., Rohr, J., Bechthold, A., 2000. Two new tailoring enzymes, a glycosyltransferase and an oxygenase, involved in biosynthesis of the angucycline antibiotic urdamycin A in Streptomyces fradiae Tu2717. Microbiology, 146: 147-154.

Ford, C.M., Boss, P.K., Hoj, P.B., 1998. Cloning and characterization of *Vitis vinifera* UDP-glucose : flavonoid 3-O-glucosyltransferase, a homologue of the enzyme encoded by the maize Bronze-1 locus that may primarily serve to glucosylate anthocyanidins *in vivo*. J Biol Chem, 273: 9224-9233.

Funaki, A., Waki, T., Noguchi, A., Kawai, Y., Yamashita, S., Takahashi, S., Nakayama, T., 2015. Identification of a Highly Specific Isoflavone 7-O-glucosyltransferase in the soybean (*Glycine max* (L.) Merr.). Plant and Cell Physiol, 56: 1512-1520.

Gao, J., Ma, B., Lu Y, Zhang, Y., Tong, Y., Guo, S., Gao, W., Huang, L., 2020. Investigating the Catalytic Activity of Glycosyltransferase on Quercetin from *Tripterygium wilfordii*. ACS Omega, 5: 1414-1421.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., Rokhsar, D. S., 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res, D1178–D1186.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q.D., Chen, Z.H., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol, 29: 644-652.

Guo, F., Xu, N., Yang, J., Zhang, X., Liu, X., 2009. The character and the cultivation technique of new safflower variety "Yunhong No.3". J Agric Sci, 000: 38-39,41.(in Chinese)

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R., 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol, 9: R7.

Han, S.Y., Li, H.X., Ma, X., Zhang, K., Ma, Z.Z., and Tu, P.F., 2009. Protective effects of purified safflower extract on myocardial ischemia in vivo and in *vitro*. Phytomedicine, 16: 694-702.

Hansen, E.H., Osmani, S.A., Kristensen, C., Møller, B.L., Hansen, J., 2009. Substrate specificities of family 1 UGTs gained by domain swapping. Phytochemistry, 70: 473-482.

Hattori, M., Huang, X.L., Che, Q.M., Kawata, Y., Tezuka, Y., Kikuchi, T., Namba, T., 1992. 6-hydroxykaempferol and its glycosides from *Carthamus tinctorius* petals. Phytochemistry, 31: 4001-4004.

He, J.B., Zhao, P., Hu, Z.M., Liu, S., Kuang, Y., Zhang, M., Li, B., Yun, C.H., Qiao, X., Ye, M., 2019. Molecular and structural characterization of a promiscuous c-glycosyltransferase from *Trollius chinensis*. Angew Chem Int Ed Engl, 58: 11513-11520.

Hirade, Y., Kotoku, N., Terasaka, K., Saijo-Hamano, Y., Fukumoto, A., Mizukami, H., 2015. Identification and functional analysis of 2-hydroxyflavanone C-glucosyltransferase in soybean (*Glycirte max*). Febs Letters, 589: 1778-1786.

Hong, B., Wang, Z., Xu, T., Li, C., Li, W., 2015. Matrix solid-phase dispersion extraction followed by high performance liquid chromatography-diode array detection and ultra performance liquid chromatography-quadrupole-time of flight-mass spectrometer method for the determination of the main compounds from *Carthamus tinctorius* L. (Hong-hua). J Pharm Biomed Anal, 107: 464-472.

Horvath, D.M., Chua, N.H., 1996. Identification of an immediate-early salicylic acid-inducible tobacco gene and characterization of induction by other compounds. Plant Mol Biol, 31: 1061-1072.

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., Lauvergeat, V., 2011. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. J Exp Bot, 62: 2465–2483.

Hsu, TM., Welner, D.H., Russ, Z.N., Cervantes, B., Prathuri, R.L., Adams, P.D., Dueber, J.E., 2018. Employing a biochemical protecting group for a sustainable indigo dyeing strategy. Nat Chem Biol, 14: 256-261

Ito, T., Fujimoto, S., Suito, F., Shimosaka, M., Taguchi, G., 2017. C-Glycosyltransferases catalyzing the formation of di-C-glucosyl flavonoids in citrus plants. Plant J, 91: 187-198.

Itkin, M., Davidovich-Rikanati, R., Cohen, S., Portnoy, V., Doron-Faigenboim, A., Oren, E., Freilich, S., Tzuri, G., Baranes, N., Shen, S., Petreikov, M., Sertchook, R., Ben-Dor, S., Gottlieb, H., Hernandez, A., Nelson, D.R., Paris, H.S., Tadmor, Y., Burger, Y., Lewinsohn, E., Katzir, N., Schaffer, A., 2018. The biosynthetic pathway of the nonsugar, high-intensity sweetener mogroside V from *Siraitia grosvenorii*. Proc. Natl. Acad. Sci. U.S.A., 115: E3862-E3862.

Jia, H.M., Jia, H.J., Cai, Q.L., Wang, Y., Zhao, H.B., Yang, W.F., Wang, G.Y., Li, Y.H., Zhan, D.L., Shen, Y.T., et al. 2019. The red bayberry genome and genetic basis of sex determination. Plant Biotechnol J, 17: 397-409.

Jian, W., Cao, H.H., Yuan, S., Liu, Y.D., Lu, J.F., Lu, W., Li, N., Wang, J.H., Zou, J., Tang, N., Xu, C., Cheng, Y.L., Gao, Y.Q., Xi, W.P., Bouzayen, M., Li, Z.G., 2019. SlMYB75, an MYB-type transcription factor, promotes anthocyanin accumulation and enhances volatile aroma production in tomato fruits. Hortic Res, 6: 22.

Jones, P., Vogt, T., 2001. Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. Planta, 213: 164-174.

Jung, S.C., Kim, W., Park, S.C., Jeong, J., Park, M.K., Lim, S., Lee, Y., Im, W.T., Lee, J.H., Choi, G., et al. 2014. Two Ginseng UDP-Glycosyltransferases synthesize ginsenoside Rg(3) and Rd. Plant Cell Physiol, 55: 2177-2188.

Kanehisa, M., Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res, 28: 27-30.

Katoh, K., Kuma, K., Toh, H., Miyata, T., 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res, 33: 511-518.

Kazuma, K., Takahashi, T., Sato, K., Takeuchi, H., Matsumoto, T., Okuno, T. 2000. Quinochalcones and flavonoids from fresh florets in different cultivars of *Carthamus tinctorius* L. Biosci Biotech Bioch, 64: 1588-1599.

Khan, H., Saeedi, M., Nabavi, S.M., Mubarak, M.S., Bishayee, A., 2019. Glycosides from medicinal plants as potential anticancer agents: emerging trends towards future drugs. Curr Med Chem, 26: 2389-2406.

Kim, D., Landmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. Nat Methods, 12: 357-U121.

Knoch, E., Sugawara, S., Mori, T., Nakabayashi, R., Saito, K., Yonekura-Sakakibara, K., 2018. UGT79B31 is responsible for the final modification step of pollen-specific flavonoid biosynthesis in *Petunia hybrida*. Planta, 247: 779-790.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res, 27: 722-736.

Korf, I. 2004. Gene finding in novel genomes. BMC Bioinform, 5: 59.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics, 35: 4453-4455.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K., 2018. Mega x: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol, 35: 1547-1549.

Lanot, A., Hodge, D., Jackson, R.G., George, G.L., Bowles, D.J., 2010. The glucosyltransferase UGT72E2 is responsible for monolignol 4-O-glucoside production in *Arabidopsis thaliana*. Plant J, 48: 286-295.

Lanot, A., Hodge, D., Lim, E.K., Vaistij, F.E., Bowles, D.J., 2008. Redirection of flux through the phenylpropanoid pathway by increased glucosylation of soluble intermediates. Planta, 228: 609-616.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics, 23: 2947-2948.

Lee, J.Y., Chang, E.J., Kim, H.J., Park, J.H., Choi, S.W., 2002. Antioxidative flavonoids from leaves of *Carthamus tinctorius*. Arch Pharm Res, 25: 313-319.

Letunic, I., Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res, 47: 256-259.

Li, J.H., Xu, G.L., Lin, S.M., and Qin, L., 2012. Systematic evaluation of effectiveness and safety on safflower injection in angina pectoris. J Emerg Tradit Chin Med, 21: 932–939. (in Chinese)

Li, P., Li, Y.J., Zhang, F.J., Zhang, G.Z., Jiang, X.Y., Yu, H.M., Hou, B.K., 2017. The Arabidopsis UDP-glycosyltransferases UGT79B2 and UGT79B3, contribute to cold, salt and drought stress tolerance via modulating anthocyanin accumulation. Plant J, 89: 85-103.

Li, Q., Yu, H.M., Meng, X.F., Lin, J.S., Li, Y.J., Hou, B.K., 2018. Ectopic expression of glycosyltransferase UGT76E11 increases flavonoid accumulation and enhances abiotic stress tolerance in Arabidopsis. Plant Biol (Stuttg), 20: 10-19.

Li, F., He, Z., Ye, Y., 2017. Isocartormin, a novel quinochalcone C-glycoside from *Carthamus tinctorius*. Acta Pharm Sin B, 7: 527-531.

Li, S., Zhao, H., and Bao, L., 2019a. Preliminary study on the mechanism of carvacrol regulating hepatocellular carcinoma based on network pharmacology. Lett Drug Des Discov, 16: 1286-1295.

Li, X.J., Kang, Y., Wang, R.R., Liao, X.L., Ou, X.F., Liu, J., Zuo, Y.X., 2019b. The effects of safflower yellow on acute exacerbation of chronic obstructive pulmonary disease: a randomized, controlled clinical trial. Evid-Based Compl Alt, 2019: 5952742.

Li, X., Xiang, F., Han, W., Qie, B., Zhai, R., Yang, C., Wang, Z., and Xu, L., 2021. The mir-domain of PbbHLH2 is involved in regulation of the anthocyanin biosynthetic pathway in "Red Zaosu" (*PyrusBretschneideri Rehd.*) Pear Fruit. Int J Mol Sci, 22: 3026.

Lilan, C., Shengbo, Y., Xinlong, D., Qinggang, Y., Yajun, L., Xiaolan, J., Yahui, W., Yumei, Q., Yongzhen, P., Liping, G., 2016, Identification of UDP-glycosyltransferases involved in the biosynthesis of astringent taste compounds in tea (*Camellia sinensis*). J Exp Bot, 2285-2297.

Lim, C.E., Ahn, J.H., Lim, J., 2006. Molecular genetic analysis of tandemly located glycosyltransferase genes,*UGT73B1,UGT73B2*, and*UGT73B3*, in *Arabidopsis thaliana*. J Plant Biol, 49: 309-314.

Lim, E.K., Jackson, R.G., Bowles, D.J., 2005 Identification and characterisation of *Arabidopsis* glycosyltransferases capable of glucosylating coniferyl aldehyde and sinapyl aldehyde. Febs Letters, 579: 2802-2806.

Lin, J.S., Huang, X.X., Li, Q., Cao, Y., Bao, Y., Meng, X.F., Li, Y.J., Fu, C., Hou, B.K., 2016. UDP-glycosyltransferase 72B1 catalyzes the glucose conjugation of monolignols and is essential for the normal cell wall lignification in *Arabidopsis thaliana*. Plant J, 88: 26-42.

Liu, C.J., 2010. Biosynthesis of hydroxycinnamate conjugates: implications for sustainable biomass and biofuel production. Biofuels, 1: 745-761.

Liu, M.Y., Zhang, C.J., Duan, L.X., Luan, Q.Q., Li, J.L., Yang, A.G., Qi, X.Q., Ren, Z.H., 2019. CsMYB60 is a key regulator of flavonols and proanthocyanidans that determine the colour of fruit spines in cucumber. J Exp Bot, 70: 69–84.

Liu, S., Wang, Y., Shi, M., Maoz, I., Gao, X., Sun, M., Yuan, T., Li, K., Zhou, W., Guo, X., Kai, G., 2022. SmbHLH60 and SmMYC2 antagonistically regulate phenolic acids and anthocyanins biosynthesis in Salvia miltiorrhiza. J Adv Res, 42: 205–219.

Liu, S., Gao, X., Shi, M., Sun, M., Li, K., Cai, Y., Chen, C., Wang, C., Maoz, I., Guo, X., Kai, G., 2023. Jasmonic acid regulates the biosynthesis of medicinal metabolites via the JAZ9-MYB76 complex in Salvia miltiorrhiza. Hortic Res, 10: 3.

Liu, X.N., Cheng, J., Zhang, G.H., Ding, W.T., Duan, L.J., Yang, J., Kui, L., Cheng, X.Z., Ruan, J.X., Fan, W., Chen, J.W., Long, G.Q., Zhao, Y., Cai, J., Wang, W., Ma, Y.H., Dong, Y., Yang, S.C., Jiang, H.F., 2018. Engineering yeast for the production of breviscapine by genomic analysis and synthetic biology approaches. Nat Commun, 9: 448.

Liu, Y., Zhao, S.Z., Wang, J.S., Zhao, C.Z., Guan, H.S., Hou, L., Li, C.S., Xia, H., Wang, X.J., 2015. Molecular cloning, expression, and evolution analysis of type II CHI gene from peanut (*Arachis hypogaea* L.). Dev Genes Evol, 225: 1-10.

Lotkowska, M.E., Tohge, T., Fernie, A.R., Xue, G.P., Balazadeh, S., Mueller-Roeber, B., 2015. The Arabidopsis transcription factor MYB112 promotes anthocyanin formation during salinity and under high light stress. Plant Physiol, 169: 1862-1880

Lowe, T,. Eddy, S., 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res, 25: 955-964.

Lu, S.N., Wang, J.Y., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M.Z., Zhang, D.C., Zheng, C.J., Lanczycki, C.J., Marchler-Bauer, A., 2020. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res, 48: 265-268.

Lunkenbein, S., Bellido, M., Aharoni, A., Salentijn, E.M.J., Kaldenhoff, R., Coiner, H.A., Munoz-Blanco, J., Schwab, W., 2006. Cinnamate metabolism in ripening

fruit. Characterization of a UDP-glucose: Cinnamate glucosyltransferase from strawberry. Plant Physiol, 140: 1047-1058.

Luzhetskyy, A., Mendez, C., Salas, J.A., Bechthold, A., 2008. Glycosyltransferases, important tools for drug design. Curr Top Med Chem, 8: 680-709.

Majoros, W.H., Pertea, M., Salzberg, S.L, 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics, 20: 2878-2879.

Martin, C., Li, J., 2017. Medicine is not health care, food is health care: plant metabolic engineering, diet and human health. New Phytol, 216: 699–719.

Mashima, K., Hatano, M., Suzuki, H., Shimosaka, M., Taguchi, G., 2019. Identification and characterization of apigenin 6-c-glucosyltransferase involved in biosynthesis of isosaponarin in wasabi (*Eutrema japonicum*). Plant Cell Physiol, 60: 2733-2743.

Mathews, H., Clendennen, S.K., Caldwell, C.G., Liu, X.L., Connors, K., Matheis, N., Schuster, D.K., Menasco, D.J., Wagoner, W., Lightner, J., et al. 2003. Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell, 15: 1689-1703.

McConkey, B.J., Sobolev, V., Edelman, M., 2002. Quantification of protein surfaces, volumes and atom-atom contacts using a constrained Voronoi procedure. Bioinformatics, 18: 1365-1373.

Miller, K.D., Guyon, V., Evans, J.N.S., Shuttleworth, W.A., Taylor, L.P., 1999. Purification, cloning, and heterologous expression of a catalytically efficient flavonol 3-O-galactosyltransferase expressed in the male gametophyte of *Petunia hybrida*. J Biol Chem, 274: 34011-34019.

Modolo, L.V., Blount, J.W., Achnine, L., Naoumkina, M.A., Wang, X.Q., Dixon, R.A., 2007. A functional genomics approach to (iso)flavonoid glycosylation in the model legume. Plant Mol Biol, 64: 499-518.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J., 2009. Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. J. Computational Chemistry 2009, 16: 2785-2791.

Nagatomo, Y., Usui, S., Ito, T., Kato, A., Shimosaka, M., Taguchi, G., 2014. Purification, molecular cloning and functional characterization of flavonoid C-glucosyltransferases from *Fagopyrum esculentum* M. (buckwheat) cotyledon. Plant J, 80: 437-448.

Nakatsuka, T., Haruta, K.S., Pitaksutheepong, C., Abe, Y., Kakizaki, Y., Yamamoto, K., Shimada, N., Yamamura, S., Nishihara, M., 2008. Identification and characterization of R2R3-MYB and bHLH transcription factors regulating anthocyanin biosynthesis in gentian flowers. Plant Cell Physiol, 49: 1818-1829.

Nawrocki, E.P., Eddy, S.R., 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics, 29: 2933-2935.

Noda, K.-i., Glover, B.J., Linstead, P., Martin, C., 1994. Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. Nature, 369: 661-664.

Ou, S., Jiang, N., 2019. LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. Mob DNA, 10: 48.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., Salzberg, S.L., 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol, 33: 290.

Qu, C., Yue, S., Lin, H., Kai, J., Shang, G., Tang, Y., Tao, W., and Duan, J., 2015. Chemical constituents of *Carthamus tinctorius*. Chin Tradit Herbal Drugs46: 1872-1877. (in Chinese)

Quattrocchio, F., Wing, J., Woude, K.v.d., Souer, E., Vetten, N.d., Mol, J., Koes, R., 1999. Molecular analysis of the anthocyanin2 gene of Petunia and its role in the evolution of flower color. Plant Cell, 11: 1433-1444.

Ravaglia, D., Espley, R.V., Henry-Kirk, R.A., Andreotti, C., Ziosi, V., Hellens, R.P., Costa, G., Allan, A.C., 2013. Transcriptional regulation of flavonoid biosynthesis in nectarine (*Prunus persica*) by a set of R2R3 MYB transcription factors. BMC Plant Biol, 13: 68.

Ren, C., Chen, C., Dong, S., Wang, R., Xian, B., Liu, T., Xi, Z., Pei, J., Chen, J., 2022. Integrated metabolomics and transcriptome analysis on flavonoid biosynthesis in flowers of safflower (*Carthamus tinctorius* L.) during colour-transition. *PeerJ*, *10*: e13591.

Rodas, F.R., Rodriguez, T.O., Murai, Y., Iwashina, T., Sugawara, S., Suzuki, M., Nakabayashi, R., Yonekura-Sakakibara, K., Saito, K., Kitajima, J., et al. 2014. Linkage mapping, molecular cloning and functional analysis of soybean gene Fg2 encoding flavonol 3-O-glucoside (1 -> 6) rhamnosyltransferase. Plant Mol Biol, 84: 287-300.

Sato, S., Akiya, T., Nishizawa, H., Suzuki, T., 2006. Total synthesis of three naturally occurring 6,8-di-C-glycosylflavonoids: phloretin, naringenin, and apigenin bis-C-beta-D-glucosides. Carbohydr Res, 341: 964-970.

Schrödinger, L., DeLano, W., 2020. PyMOL. Retrieved from http: //www.pymol.org/pymol

Schnable, P.S., Ware, D., Fulton, R.S., et al 2009. The B73 maize genome: complexity, diversity, and dynamics. Science, 326: 1112-1115.

Schnable, P.S., 2012. The B73 maize genome: complexity, diversity, dynamics. Science, 337: 1040-1040.

Schwinn, K., Venail, J., Shang, Y.J., Mackay, S., Alm, V., Butelli, E., Oyama, R., Bailey, P., Davies, K., Martin, C., 2006. A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus Antirrhinum. Plant Cell, 18: 831-851.

Shannon, P., 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res, 13: 2498-2504.

Shen, Q., Zhang, L.D., Liao, Z.H., Wang, S.Y., Yan, T.X., Shi, P., Liu, M., Fu, X.Q., Pan, Q.F., Wang, Y.L., Lv, Z.Y., Lu, X., Zhang, F.Y., Jiang, W.M., Ma, Y.N., Chen, M.H., Hao, X.L., Li, L., Tang, Y.L., Lv, G., Zhou, Y., Sun, X.F., Brodelius, P.E., Rose, J.K.C., Tang, K.X., 2018. The genome of *Artemisia annua* provides insight into the evolution of Asteraceae family and artemisinin biosynthesis. Mol Plant, 11: 776-788.

Spitzer-Rimon, B., Marhevka, E., Barkai, O., Marton, I., Edelbaum, O., Masci, T., Prathapani, N.K., Shklarman, E., Ovadis, M., Vainstein, A., 2010. EOBII, a Gene Encoding a flower-specific regulator of phenylpropanoid volatiles biosynthesis in Petunia. Plant Cell, 22: 1961-1976.

Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31: 3210-3212.

Song, C., Liu, Y.F., Song, A.P., Dong, G.Q., Zhao, H.B., Sun, W., Ramakrishnan, S., Wang, Y., Wang, S.B., Li, T.Z., Niu, Y., Jiang, J.F., Dong, B., Xia, Y., Chen, S.M., Hu, Z.G., Chen, F.D., Chen, S.L., 2018. The *Chrysanthemum nankingense* genome provides insights into the evolution and diversification of Chrysanthemum flowers and medicinal traits. Mol Plant, 11: 1482-1491.

Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B., 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res, 32: W309-W312.

Stevenson, P. C., Nicolson, S.W. and Wright, G.A., 2017. Plant secondary metabolites in nectar: impacts on pollinators and ecological functions. Funct Ecol, 31: 65-75.

Sun, C., Deng, L., Du, M., Zhao, J., Chen, Q., Huang, T., Jiang, H., Li, C.B., and Li, C., 2020. A transcriptional network promotes anthocyanin biosynthesis in tomato flesh. Mol Plant, 13: 42-58.

Sun, Y., Xu, D.P., Qin, Z., Wang, P.Y., Hu, B.H., Yu, J.G., Zhao, Y., Cai, B., Chen, Y.L., Lu, M., Liu, J.G., Liu, X., 2018. Protective cerebrovascular effects of hydroxysafflor yellow A (HSYA) on ischemic stroke. Eur J Pharmacol, 818: 604-609.

Sun, Y., Ji, K., Liang, B., Du, Y., Leng, P., 2017. Suppressing ABA uridine diphosphate glucosyltransferase (SlUGT75C1) alters fruit ripening and the stress response in tomato. Plant J, 91: 574-589.

Tabata, S., and Kaneko, T., and Nakamura, Y., Kotani, H., and Kato, T., and Asamizu, E., and Miyajima, N., and Sasamoto, S., and Kimura, T., and Hosouchi, T., et al. 2000. Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. Nature, 408: 820-822.

Takos, A.M., Jaffe, F.W., Jacob, S.R., Bogs, J., Robinson, S.P., Walker, A.R., 2006. Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. Plant Physiol, 142: 1216-1232.

Teng, S., Thomson, P.A., McCarthy, S., Kramer, M., Muller, S., Lihm, J., Morris, S., Soares, D.C., Hennah, W., Harris, S., Camargo, L.M., Malkov, V., McIntosh, A.M., Millar, J.K., Blackwood, D.H., Evans, K.L., Deary, I.J., Porteous, D.J., McCombie, W.R., 2018. Rare disruptive variants in the DISC1 Interactome and Regulome: association with cognitive ability and schizophrenia. Mol Psychiatry, 23: 1270-1277.

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O., and Borodovsky, M., 2008. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res, 18: 1979-1990.

Tohge, T., de Souza, L.P., Fernie, A.R., 2017. Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. J Exp Bot, 68: 4013-4028.

Vogt, T., Jones, P., 2000. Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. Trends Plant Sci, 5: 380-386.

Verdonk, J.C., Haring, M.A., van Tunen, A.J., Schuurink, R.C., 2005. ODORANT1 regulates fragrance biosynthesis in petunia flowers. Plant Cell, 17: 1612-1624.

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J., 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics, 8: 77-80.

Wang, X., 2009. Structure, mechanism and engineering of plant natural product glycosyltransferases. Febs Letters, 583: 3303-3309.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., Kissinger, J.C., Paterson, A.H., 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res, 40: 49.

Wang, Z.L., Gao, H.M., Wang, S., Zhang, M., Chen, K., Zhang, Y.Q., Wang, H.D., Han, B.Y., Xu, L.L., Song, T.Q., Yun, C.H., Qiao, X., Ye, M., 2020. Dissection of the general two-step di-C-glycosylation pathway for the biosynthesis of (iso)schaftosides in higher plants. Proc Natl Acad Sci U S A, 117: 30816-30823.

Wilson, A.E., Wu, S., Tian, L., 2019. PgUGT95B2 preferentially metabolizes flavones/flavonols and has evolved independently from flavone/flavonol UGTs identified in *Arabidopsis thaliana*. Phytochemistry, 157: 184-193.

Witte, S., Moco, S., Vervoort, J., Matern, U., Martens, S., 2009. Recombinant expression and functional characterisation of regiospecific flavonoid glucosyltransferases from *Hieracium pilosella* L. Planta, 229: 1135-1146.

Wu, Z., Liu, H., Zhan, W., Yu, Z., Qin, E., Liu, S., Yang, T., Xiang, N., Kudrna, D., Chen, Y., et al. 2021. The chromosome-scale reference genome of safflower (*Carthamus tinctorius*) provides insights into linoleic acid and flavonoid biosynthesis. Plant Biotechnol J, 19: 1725-1742.

Xian, B., Wang, R., Jiang, H., Zhou, Y., Yan, J., Huang, X., Chen, J., Wu, Q., Chen, C., Xi, Z., Ren, C., Pei, J., 2022. Comprehensive review of two groups of flavonoids in *Carthamus tinctorius* L. Biomed Pharmacother, 153: 113462.

Xiaojia, S., Guoan, S., Shaokang, D., Dixon, R.A., Yongzhen, P., 2017. Characterization of UGT716A1 as a multi-substrate UDP: flavonoid glucosyltransferase gene in *Ginkgo biloba*. Front Plant Sci, 8: 2085.

Xie, F., Xiao, P., Chen, D., Xu, L., Zhang, B., 2012. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. Plant Mol Biol, 80: 75-84.

Xie, K., Chen, R., Li, J., Wang, R., Chen, D., Dou, X., Dai, J., 2014. Exploring the catalytic promiscuity of a new glycosyltransferase from *Carthamus tinctorius*. Org Lett, 16: 4874-4877.

Xu, Z., Wang, H., 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res, 35: W265-W268.

Xu, F., Ning, Y., Zhang, W., Liao, Y., Li, L., Cheng, H., Cheng, S., 2014. An R2R3-MYB transcription factor as a negative regulator of the flavonoid biosynthesis pathway in *Ginkgo biloba*. Funct Integr Genomics, 14: 177-89.

Yin, H.B., He, Z.S., 2000, A novel semi-quinone chalcone sharing a pyrrole ring C-glycoside from *Carthamus tinctorius*. Tetrahedron Letters, 41: 1955-1958.

Yu, G., Wang, L.G., Han, Y., He, Q.Y., 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS, 16: 284-287.

Yuan, S., Yang, Y., Kong, J.Q., 2018. Biosynthesis of 7,8-dihydroxyflavone glycosides via OcUGT1-catalyzed glycosylation and transglycosylation. J Asian Nat Prod Res, 20: 1-13.

Yue, S., Tang, Y., Xu, C., Li, S., Zhu, Y., Duan, J.A., 2010. Two new quinochalcone c-glycosides from the florets of *Carthamus tinctorius*.Int J Mol Sci, 15: 16760-16771.

Yue, S.J., Tang, Y.P., Li, S.J., Duan, J.A., 2013. Chemical and biological properties of quinochalcone c-glycosides from the florets of *Carthamus tinctorius*. Molecules, 18: 15220-15254.

Zhang, H., Duan, C.P., Luo, X., Feng, Z.M., Yang, Y., Zhang, X., Jiang, J.S., Zhang, P.C., 2020. Two new quinochalcone glycosides from the safflower yellow pigments. J Asian Nat Prod Res, 22: 1130-1137.

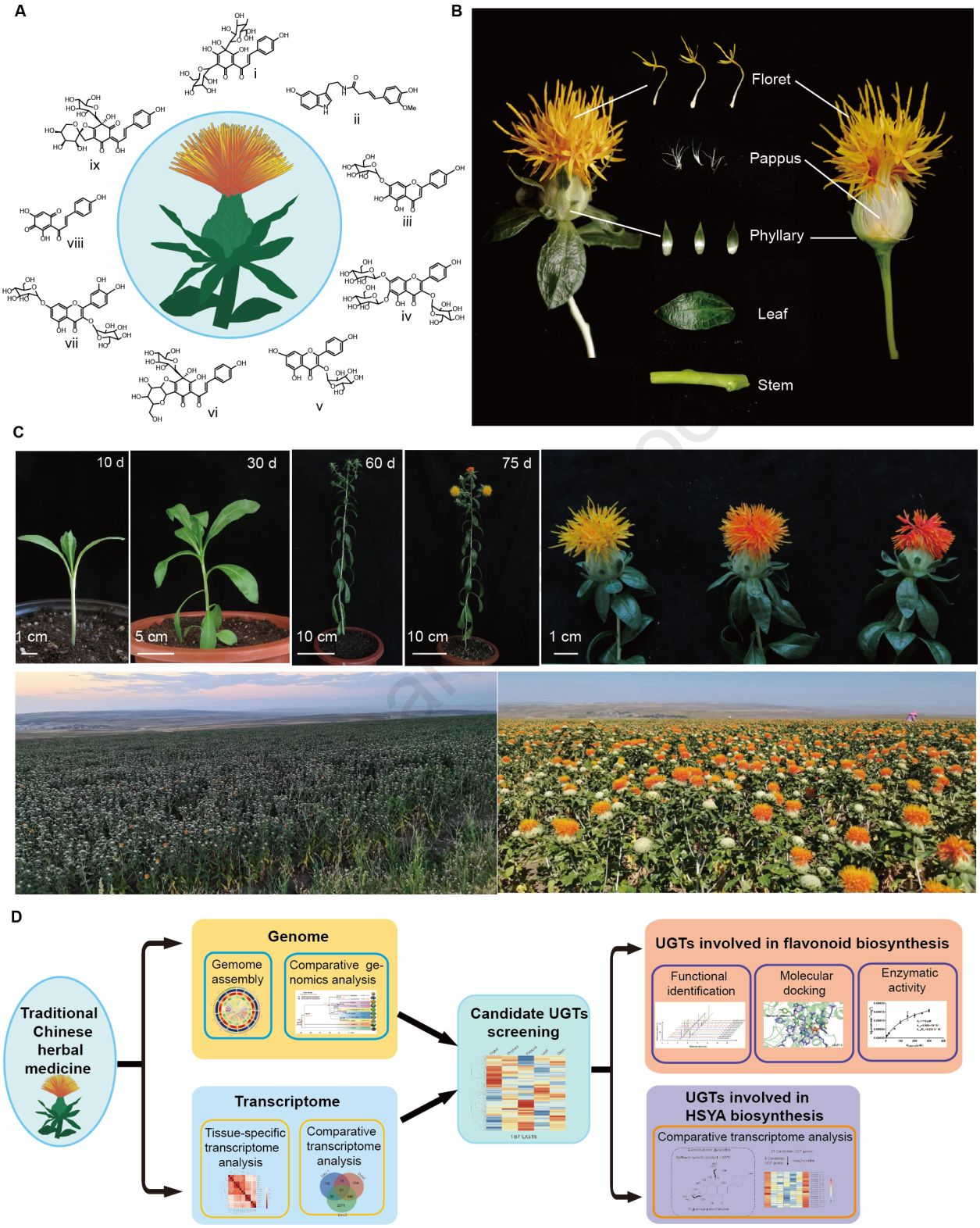
Zhang, L.L., Tian, K., Tang, Z.H., Chen, X.J., Bian, Z.X., Wang, Y.T., Lu, J.J., 2016a. Phytochemistry and pharmacology of *Carthamus tinctorius* L. Am J Chin Med, 44: 197-226.

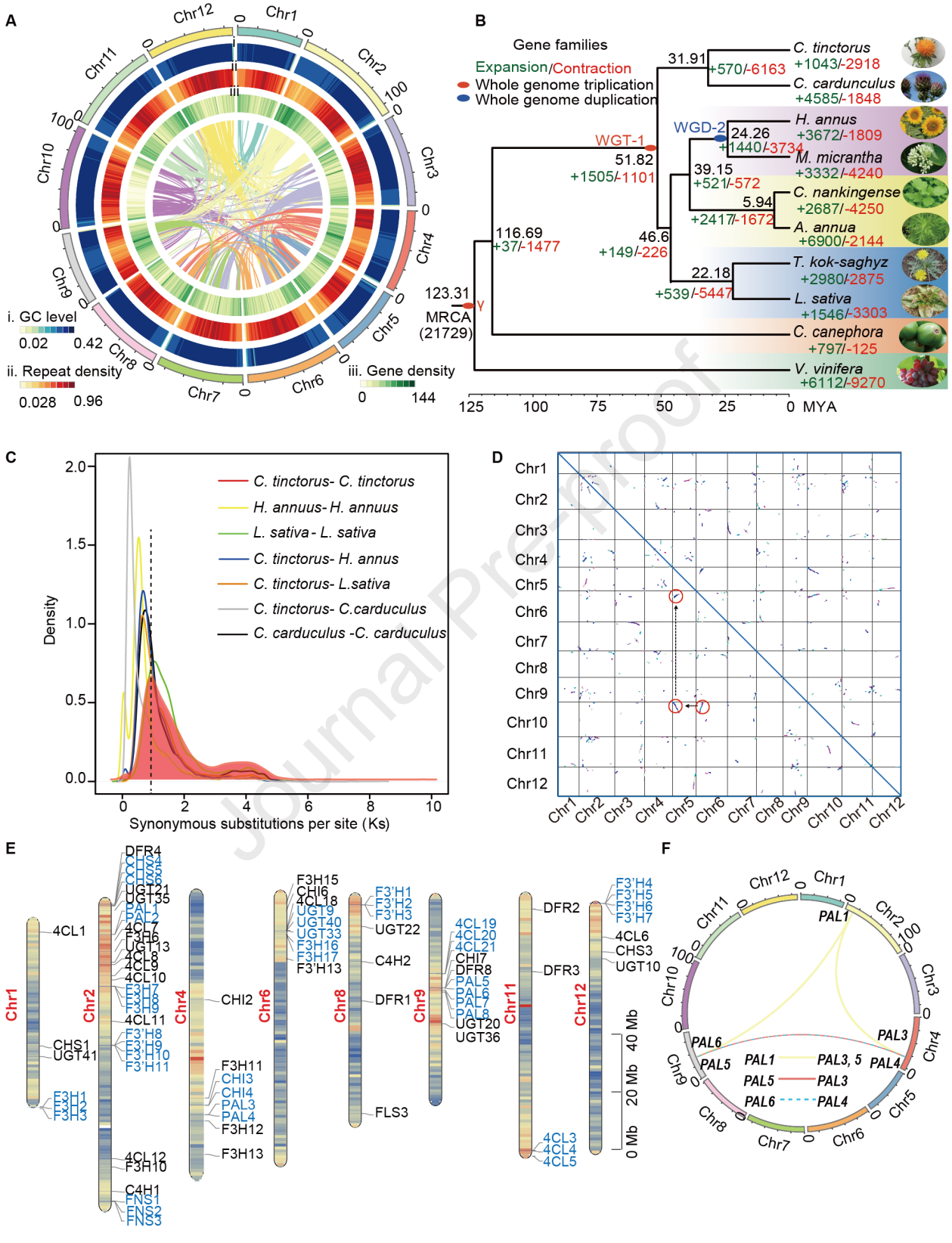
Zhang, Y.J., Xie, K.B., Liu, A.J., Chen, R.D., Chen, D.W., Yang, L., and Dai, J.G., 2019. Enzymatic biosynthesis of benzylisoquinoline alkaloid glycosides via promiscuous glycosyltransferases from *Carthamus tinctorius*. Chin Chem Lett, 30: 443-446.

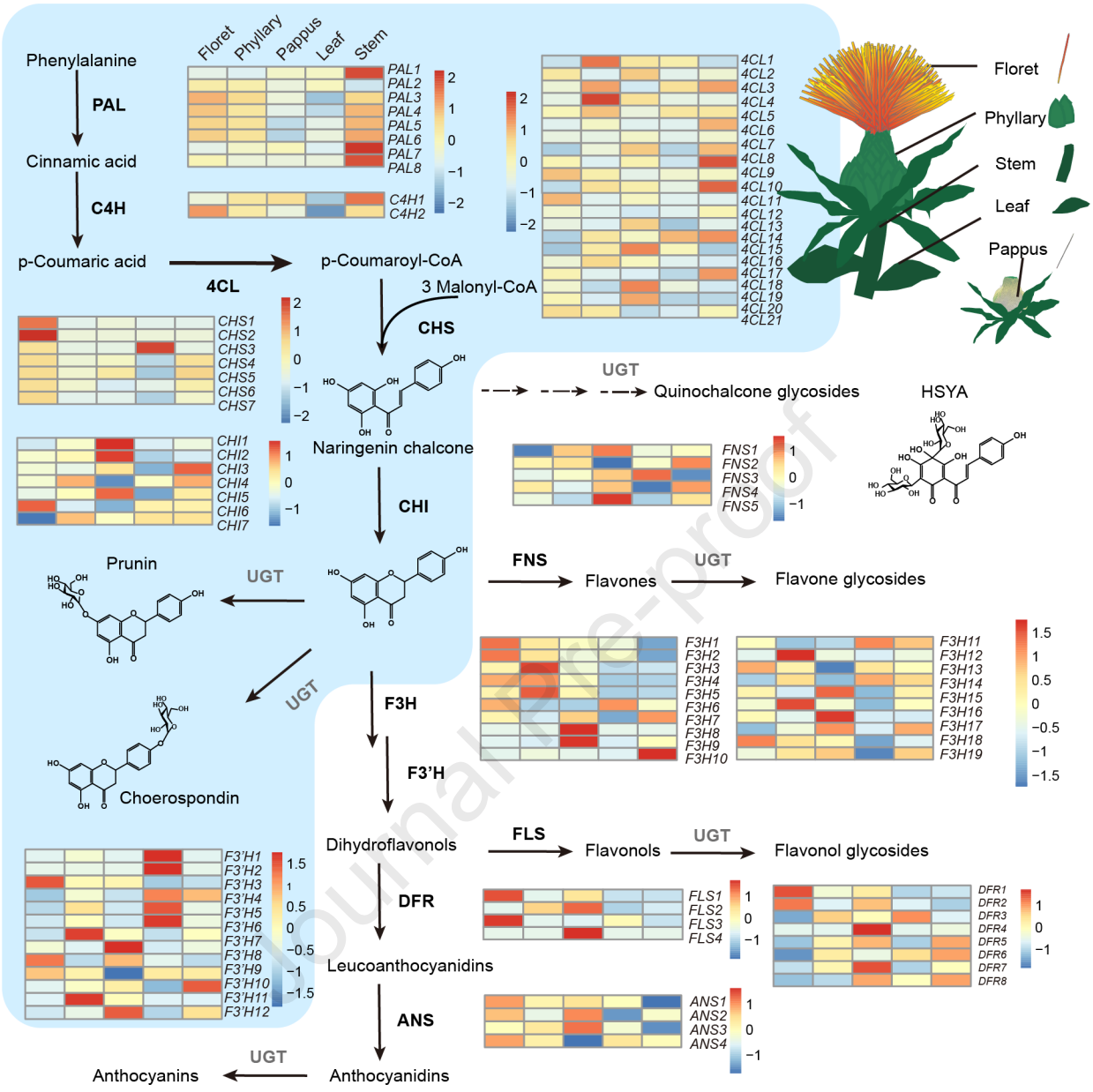
Zhou, H., Lin-Wang, K., Liao, L., Gu, C., Lu, Z., Allan, A.C., Han, Y., 2015. Peach MYB7 activates transcription of the proanthocyanidin pathway gene encoding leucoanthocyanidin reductase, but not anthocyanidin reductase. Front Plant Sci, 6: 908.

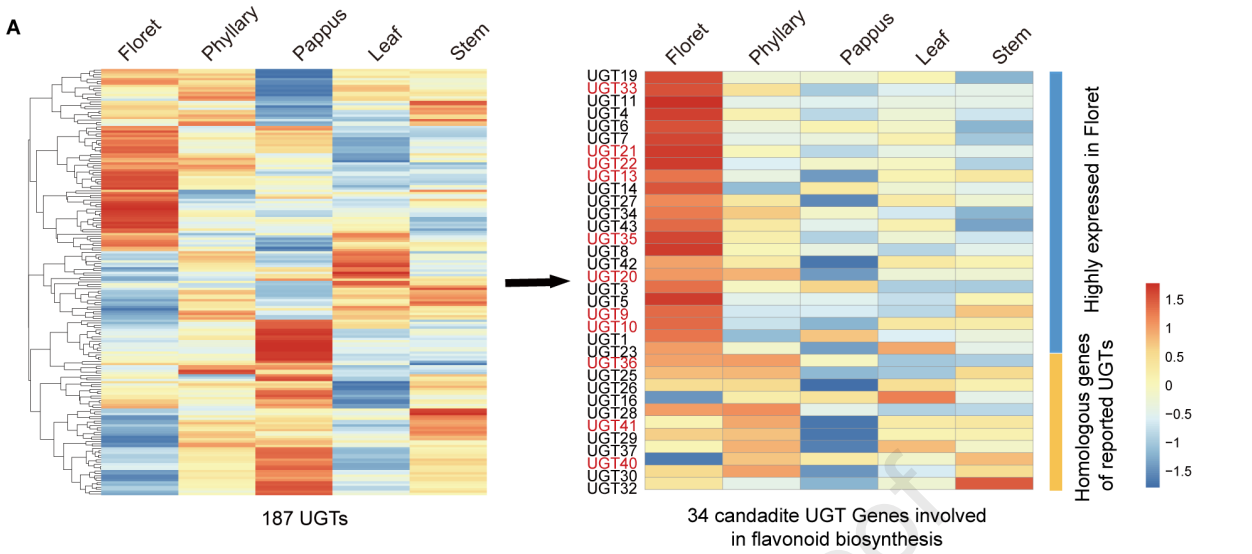
Zhou, W., Shi, M., Deng, C., Lu, S., Huang, F., Wang, Y., Kai, G., 2021. The methyl jasmonate-responsive transcription factor SmMYB1 promotes phenolic acid biosynthesis in Salvia miltiorrhiza. Hort Res, 8: 10.

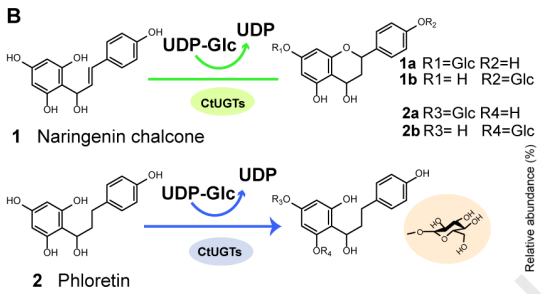
Zhou, X.D., Tang, L.Y., Xu, Y.L., Zhou, G.H., and Wang, Z.J., 2014. Towards a better understanding of medicinal uses of *Carthamus tinctorius* L. in traditional Chinese medicine: A phytochemical and pharmacological review. J Ethnopharmacol, 151: 27-43.

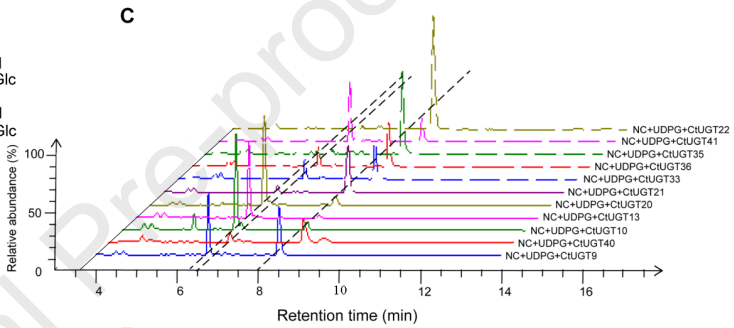

A
i
ii
iii
iv
v
vi
vii
viii
ix
B
Floret
Pappus
Phyllary
Leaf
Stem
C
10 d
30 d
60 d
75 d
1 cm
5 cm
10 cm
10 cm
1 cm
D
Traditional Chinese herbal medicine
Genome
Gemome assembly
Comparative genomics analysis
Transcriptome
Tissue-specific transcriptome analysis
Comparative transcriptome analysis
Candidate UGTs screening
UGTs involved in flavonoid biosynthesis
Functional identification
Molecular docking
Enzymatic activity
UGTs involved in HSYA biosynthesis
Comparative transcriptome analysis

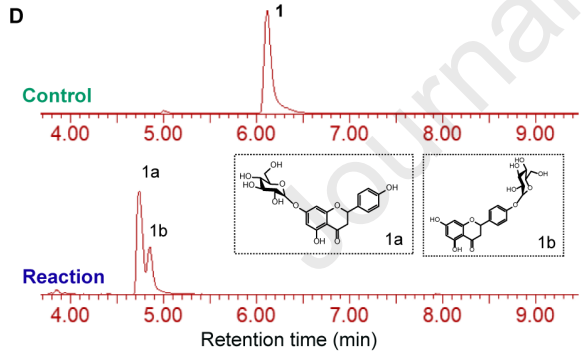


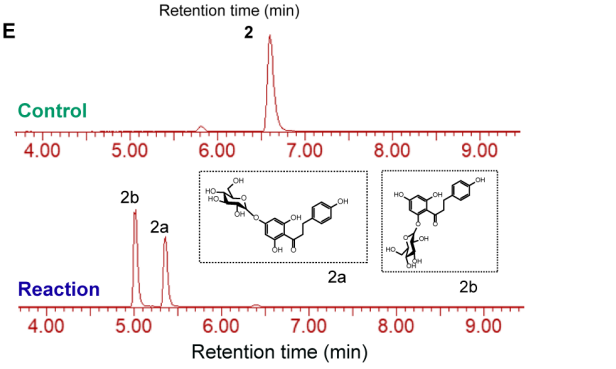


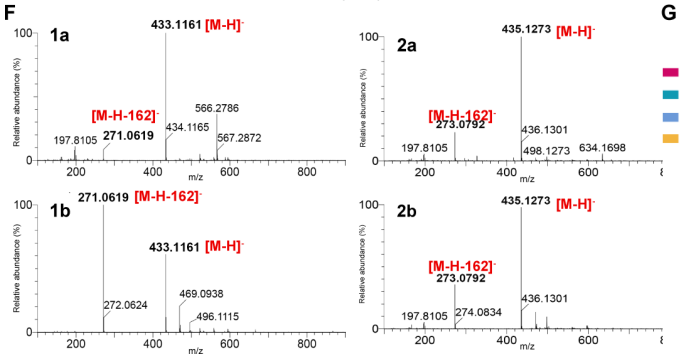


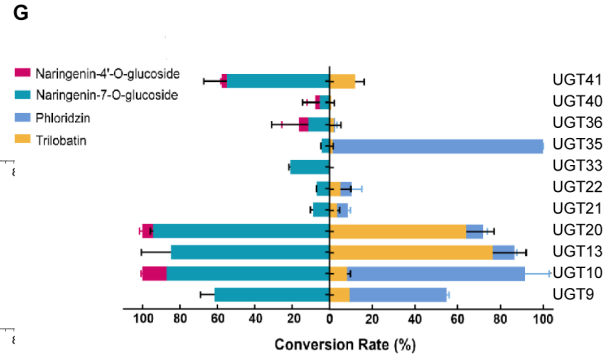


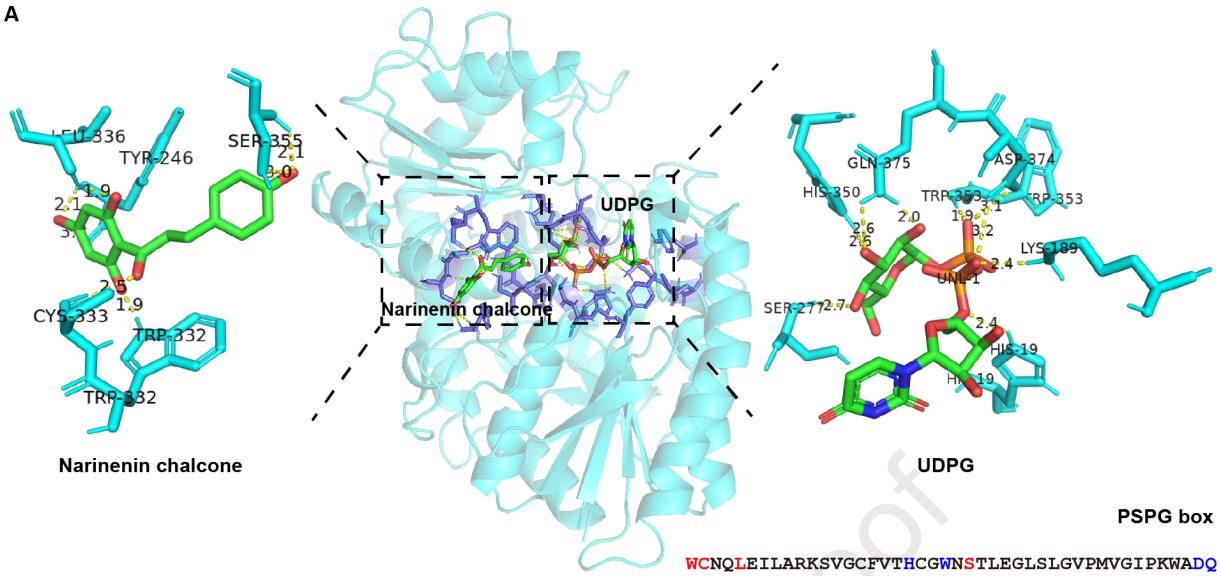


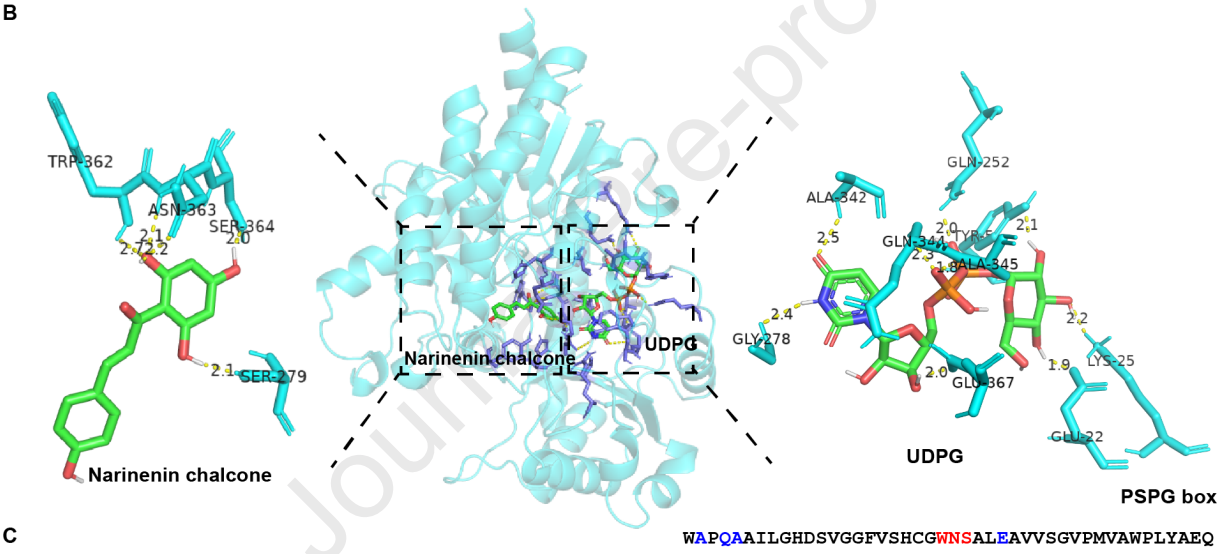


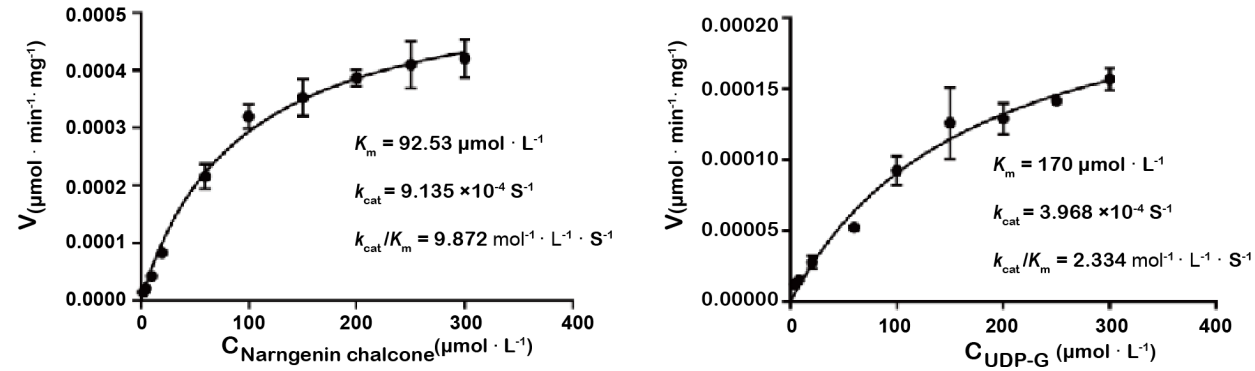


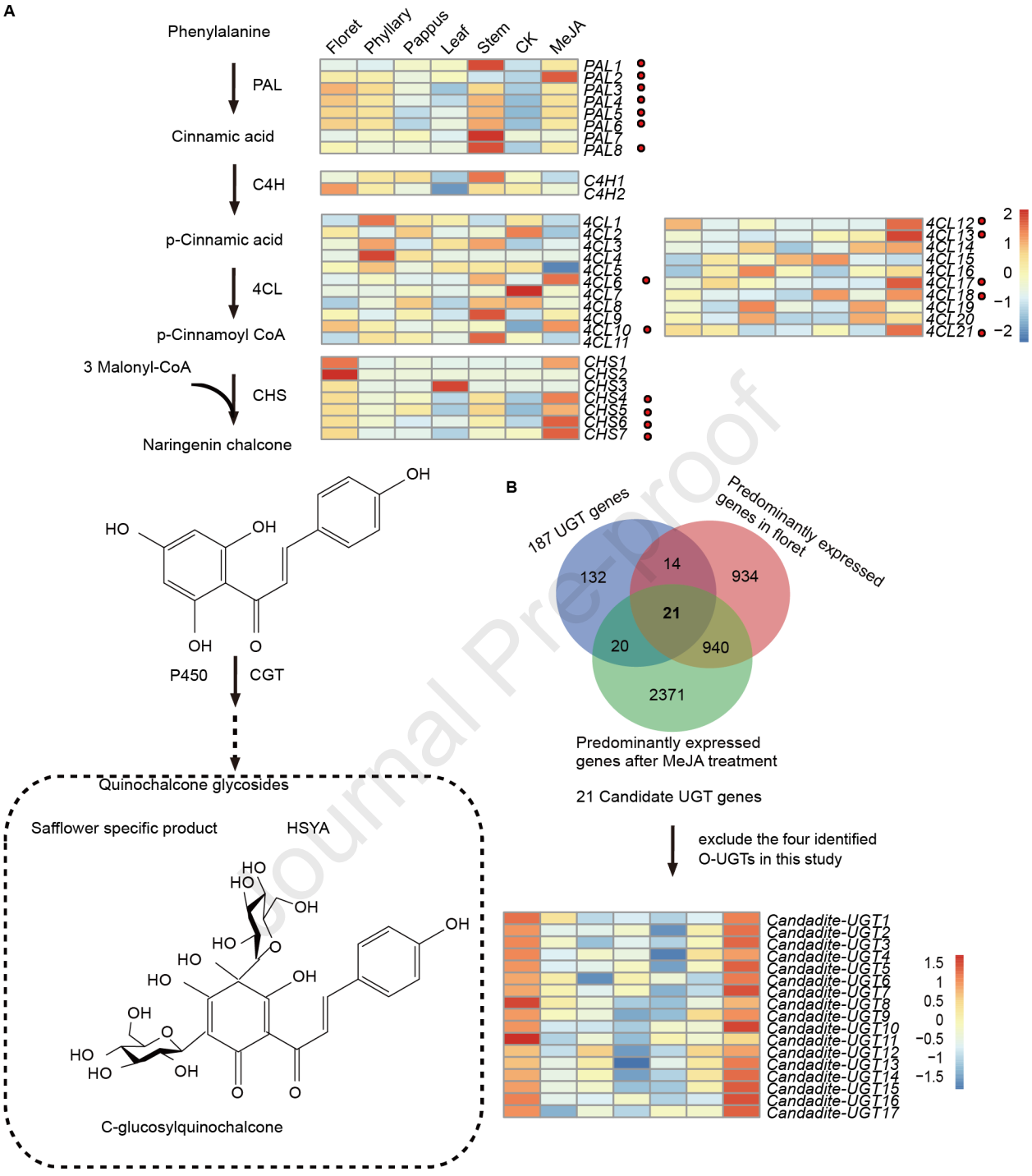





A
LEU-336
TYR-246
SER-355
CYS-333
TRP-332
TRP-332
Narinenin chalcone
UDPG
Narinenin chalcone
GLN-375
ASP-374
HIS-350
TRP-353
TRP-353
LYS-189
UNL-1
SER-277
HIS-19
HIS-19
UDPG
PSPG box
WCNQLEILARKSVGCFVTHCGWNSTLEGLSLGVPMVGIPKWADQ


B
TRP-362
ASN-363
SER-364
SER-279
Narinenin chalcone
Narinenin chalcone
UDPG
GLN-252
ALA-342
GLN-344
ALA-345
GLY-278
GLU-367
LYS-25
GLU-22
UDPG
PSPG box
WAPQAAILGHDSVGGFVSHCGWNSALEAVVSGVPMVAWPLYAEQ


C
V(μmol · min-1 · mg-1)
0.0005
0.0004
0.0003
0.0002
0.0001
0.0000
0
100
200
300
400
CNarngenin chalcone(μmol · L-1)
Km = 92.53 μmol · L-1
kcat = 9.135 ×10-4 S-1
kcat/Km = 9.872 mol-1 · L-1 · S-1
V(μmol · min-1 · mg-1)
0.00020
0.00015
0.00010
0.00005
0.00000
0
100
200
300
400
CUDP-G (μmol · L-1)
Km = 170 μmol · L-1
kcat = 3.968 ×10-4 S-1
kcat/Km = 2.334 mol-1 · L-1 · S-1


A
Phenylalanine
PAL
Cinnamic acid
C4H
p-Cinnamic acid
4CL
p-Cinnamoyl CoA
3 Malonyl-CoA
CHS
Naringenin chalcone
P450
CGT
Quinochalcone glycosides
Safflower specific product
HSYA
C-glucosylquinochalcone
Floret
Phyllary
Pappus
Leaf
Stem
CK
MeJA
PAL1
PAL2
PAL3
PAL4
PAL5
PAL6
PAL7
PAL8
C4H1
C4H2
4CL1
4CL2
4CL3
4CL4
4CL5
4CL6
4CL7
4CL8
4CL9
4CL10
4CL11
4CL12
4CL13
4CL14
4CL15
4CL16
4CL17
4CL18
4CL19
4CL20
4CL21
CHS1
CHS2
CHS3
CHS4
CHS5
CHS6
CHS7
B
187 UGT genes
Predominantly expressed genes in floret
132
14
934
21
20
940
2371
Predominantly expressed genes after MeJA treatment
21 Candidate UGT genes
exclude the four identified O-UGTs in this study
Candadite-UGT1
Candadite-UGT2
Candadite-UGT3
Candadite-UGT4
Candadite-UGT5
Candadite-UGT6
Candadite-UGT7
Candadite-UGT8
Candadite-UGT9
Candadite-UGT10
Candadite-UGT11
Candadite-UGT12
Candadite-UGT13
Candadite-UGT14
Candadite-UGT15
Candadite-UGT16
Candadite-UGT17

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.