Unprecedented evolution and transcriptional reprogramming of CYP81E subfamily in Carthamus tinctorius during flavonoid biosynthesis

Liu Jianyu
Jilin Agricultural University
Naveed Ahmad
Jilin Agricultural University
Hong Yingqi
Jilin Agricultural University
Wang Wenkai
Jilin Agricultural university
Liu Yueyang
Jilin Agricultural University
Ma Xintong
Jilin Agricultural University
Muhammad Noman
Jilin Agricultural University
Nguyen Quoc Viet Hoang
Jilin Agricultural University
Yao Na
Jilin Agricultural University
Liu Xiuming (xiuming1211@163.com)
Jilin Agricultural University https://orcid.org/0000-0002-1478-6994
Li Haiyan
Jilin Agricultural University

Research article

Keywords: CtCYP81E subfamily, Subcellular localization, Prokaryotic expression, expression analysis, transcriptional regulation, Safflower

Posted Date: November 6th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-100921/v1

Abstract

Cytochrome P450s are widely known as an important class of enzymes involved in multi-dimensional metabolic reactions which facilitate both primary and secondary metabolism in plants. Recent advances in genome sequencing of new plant species have greatly influenced our knowledge of the evolution of gene families. Herein, we present the extensive genome-wide identification study and early experimental groundwork of CtCYP81E subfamily extracted from safflower genome. The evolutionary divergence and several other molecular aspects of CtCYP81E enzymes were described with the help of phylogenetic reconstruction and robust in silico analysis. A total of 15 CtCYP81E candidate enzymes were identified and clustered together with A-type CYP71 clan of the model plant. The detail overview of their gene structures organization, conserved signatures motif, cis regulatory elements, Go functional categorization and protein-protein interaction network, respectively suggested novel insights for physiological and biosynthetic implications. Following multiple recombinant DNA approaches combined with the development of GPF fusion, heterologous expression, and transcriptional regulation network of CtCYP81E8 under normal and fluctuating environments, further functional validation was performed. The transient expression system using onion epidermal cells revealed the candidate protein's subcellular position to cell membrane. Similarly, the biochemical assay of recombinant CtCYP81E8 protein, effectively produced during heterologous expression, verified 2,4-dimethylphenol activity over different time periods. Moreover, the results of RNA-transcriptomic data and qRT-PCR analysis of 15 CtCYP81Es at different flowering stages indicated a differential expression levels defining their potential roles during safflower metabolite biosynthesis.Consequently, the transcriptional regulation of CtCYP81E exploited with various stress conditions indicated considerable susceptibility against these environmental drifts. Furthermore, the correlation analysis of CtCYP81E8 transcription and metabolite accumulation pattern in wild and mutant safflower lines also suggested positive outcomes during flower development. Although presumably, these results may be helpful in determining the fundamental idea of transcriptional regulation channels that strategically turn on the secondary metabolic pool of plant system in response cute environmental falls.

1. Introduction

Carthamus tinctorius L. also known as safflower is commonly known as one of the important economic plants globally. The wider use of dried petals of safflower as a rich source of Chinese traditional medicine against various diseases such as coronary heart diseases, hypertension, gynaecological diseases, and cerebral blood flow and cerebrovascular diseases [1, 2]. It also offers historic blueprints for enriched flavonoids content, fatty acids, various phenolic compounds, and lignin product [3]. Safflower contains a striking variety of secondary metabolites and in particular flavonoids which include carthamin chalcone glycoside, kaempferol glucosides, hydroxy safflor yellow A and B, and quercetin glucosides [4–8]. The economical value of safflower highlights massive genomic diversity for genome wide studies of gene families related to flavonoid biosynthesis [9]. However, the knowledge about P450 gene clusters and their associated subfamilies involved directly or otherwise during flavonoid biosynthesis in safflower is still scarce.

Cytochrome P450 corresponds to a diverse group of multigene family with hundreds of genes per genome identified in 50 differet plant species [10]. The molecular biology and biochemistry of cytochrome P450 have been crucial in understanding the complex enzyme structure and function interaction, gene expression and regulation, and other catalytic reactions during plant adaptation to various environmental constrains through secondary metabolism. They are mainly engaaged with essential biosynthetic reactions such as secondary metabolites, hormones, and fatty acid conjugates, as well as during oxidative detoxification pathways in plants [11, 12]. Recent studies have shown various biosynthetic pathways including flavonoid biosynthesis, lignans metabolic pathway, and alkaloids

metabolism, [13, 14]. In addition, cytochrome P450 genes were found crucial candidates during the metabolic and stress resistance to various allelochemicals in plants [15]. Despite various researches on the functional identification of multiple subfamily genes of cytochrome P450 super family have been successfully reported in Arabidopsis, wheat, ginseng and other plants [16–19], nontheless, the comprehensive genome-wide identification and functional characterization of CYP81E subfamily in safflower still remianed unexplained. As a matter of prime significance, it is essential to complex genome diversity of safflower with a particular focus on studies related to identification and characterization of candidate gene families of cytochrome P450 genes involved in flavonoid biosynthesis.

In this study, the draft genome sequence of safflower available online at (PRJNA399628; posted publicly to NCBI on August 23, 2017) was exploited to functionally characterize CtCYP81E subfamily genes by revealing their extensive molecular evolution, structural and functional diversity, conserved patterns of the molecular regulatory factors and diverse expression pattern under normal and stress conditions. In addition, a putative CtCYP81E8 was further characterized by carrying out multiple functional analysis following with molecular cloning, subcellular localization, prokaryotic expression and differential expression analysis during flowering stages of safflower under fluctuating climates. Our results not only provide practical basis for the understanding of the biosynthesis, regulation and metabolic network of flavonoid metabolism in safflower, but also present fundamental experimental groundwork for further studies related to secondary metabolism and regulation. On the other hand, these findings also facilitate new molecular breeding programmes for safflower varieties with high metabolites content.

2. Materials And Methods

2.1. Sequence retrieval and characterization of CtCYP81E subfamily in safflower

Genome assembly of Carthamus tinctorius from Jilin Agricultural University (PRJNA399628; submitted on august 23rd, 2017) were retrieved from the NCBI website (https://www.ncbi.nlm.nih.gov). The 246 full-length P450s in Arabidopsis along with 26 pseudogenes, available on the TAIR website (https://www.Arabidopsis. org/) were used as input for the identification of candidate CtCYP81Es using genome assembly of Carthamus tinctorius with the help of local BLAST tool through BioEdit software (Ibis Biosciences, Carlsbad, CA, USA). Further verification of the identified CtCYP81Es was performed with the Pfam database search (http://pfam.xfam.org). Various physicochemical properties including molecular weight (MW) and isoelectric point (pI) of each identified CtCYP81E protein were determined by the online webserver of ExPASy (http://www.expasy.org/). Furthermore, signal peptide analysis was performed using SignalP 4.1 program. The alignment of the amino acid sequences were carried out with the help of DNAMAN software (Vers. 7; Lynnon Corporation, Quebec, Canada), by using the preset parameters. The prediction of the subcellular localization of putative CtCYP81Es was determined with CELLO v2.5 (http://cello.life.nctu.edu.tw/).

2.2. Phylogeney analysis

The full-length 15 amino acid sequences of the CtCYP81E proteins obtained from C. tinctorius genome were subjected to multiple sequence alignment using Clustal W (2.0). To analyze the evolutionary divergence and sequence homology of CtCYP81Es in comparison with 246 full-length P450s in Arabidopsis and 26 identifiable pseudogenes, a neighbour-joining phylogenetic tree with 1000 bootstrap method was generated using MEGA 5 software version 4.1 (http://www.megasoftware.net/) [20]. The classification and divergence of the CtCYP81E

family in safflower with the Arabidopsis CYPome was demonstrated and clustered in several clans based on their structural and functional properties of their subfamilies.

2.3. Gene structure, protein motifs and promoter analysis

The gene structure organization including exons and introns along the length of CtCYP81E genomic sequences were monitored using the CDs and genomic sequences of CtCYP81E genes with the help of GSDS (Gene Structure Display Server) (http://gsds.cbi.pku.edu.cn/index.php) according the instructions given by [21]. The conserved protein motifs within 15 CtCYP81E protein sequences by adding them to MEME web server Version 4.8.1; available at http://meme.nbcr.net/meme/cgi-bin/meme.cgi) using the default values. The graphical representation of protein motifs within the ML Phylogenetic trees was edited in EvolView v.2 (http://www.evolgenius.info/). For the investigation of the cis-regulatory units of the promoter region in the selectedCtCYP81E genes of *C. tinctorius*, the 2 kb upstream 5` UTR flanking sequence of each putative gene was analyzed at PLACE (https://sogo.dna.affrc.go.jp/).

2.4. Gene term enrichment

Gene ontology annotation is basically a sequence-homology-based phylogentic tool commonly used to functionally classify a set of genes in silico. GO term analysis for *C. tinctorius* CtCYP81E subfamily was determined with the help of Blast2GO (https://www.blast2go.com/) [22]. For this purpose the full-length amino acid residues of CtCYP81E proteins were added to Blast2Go for initial blast search following by mapping and annotation. The Go term annotation was classified into three classes including biological processes, cellular component and molecular function.

2.5. Protein interaction network prediction

The functional protein interactive network of the putative CtCYP81E proteins, were manually predicted using the online web server of STRING database version 10 (https://string-db.org/). The online hierarchical network of interactor proteins showed the prediction of a variable group of experimental and hypothetical proteins that interact with CtCYP81E proteins during upstream and downstream regulation.

2.6. Experimental materials, vectors, and strains

The “Jihong No.1” cultivar seeds of *c. tinctorius* were purchased from Tacheng, Xinjiang province of China. The seeds were grown in the greenhouse station of the Engineering Research Center of Jilin Agricultural University Changchun, China, until harvesting. Agrobacterium tumefaciens strain EHA105, E. coli BL21, E. coli DH5α cells, prokaryotic expression vector (PET28a^{+}-CtCYP81E8), subcellular localization vector pCAMBIA1302-CtCYP81E8-GFP) were constructed and stored in a refrigerator at -80 ° C with 75% glycerol until the next use.

2.7. Transcriptomic profiling and expression analysis

The transcriptomic profiling and expression analysis of CtCYP81E subfamily genes in safflower was determined using RNA-seq data (whole Transcriptome Shotgun Sequencing) in five different tissues/organs (root, stem, seed, flower, and leaf tissue). A heatmap is created from the kilobase model of exon model per million mapped read (RPKM) method. For semi quantitative realtime PCR analysis, total RNA content was extracted from the aforesaid tissues of 4 months old Jihong No.1 cultivar of *C. tinctorius*. The first-strand cDNA templates were synthesised using the reverse transcription system. The quantitative real-time PCR assay was carried out to determine the transcription levels of 15 *CtCYP81E* genes using SYBR® Premix Ex Taq™ (TaKaRa). The system of Stratagene Mx3000P (Stratagene, CA, USA) was employed to determine the semi qRT-PCR analysis. The expression level was

normalised with 18 s ribosomal RNA gene from *C. tinctorius* using an internal reference gene. The relative expression level of *CtCYP81Es* in each tissue was calculated according to the 2-$\Delta\Delta^{CT}$ method [23]. Each experiment was repeated in three independent biological replicates. The primers are listed in (Table S1).

2.8. Cloning, expression analysis and subcellular localization of CtCYP81E8

The total RNA content was obtained from the flower petals of JH1 cultivar of safflower using an RNA Isoplus reagent (TIANGEN Biotech, China). The first-strand cDNA templates were prepared with the help PrimeScript™ RT Reagent Kit and gDNA Eraser (TaKaRa, China). The full length cDNA sequence of *CtCYP81E8* was amplified using the primers pair CtCYP81-F1 (5'CCCATGGGATGATGAGGATGATTAGTGG3').and.CtCYP81-R1. (5'TCAAAGATGCGATAATAGATTTGACTAGTC3'). These primers were designed according to the genomic sequences of *CtCYP81E8*. The correlation analysis of the expression level of CtCYP81E8 and the accumulation of safflower yellow pigment (SY) in different flowering stages of safflower was carried out using qRT-PCR assay. For this puropose, total RNA was extracted from four different flowering tissues of safflower according to the previous method. The primers pair CtCYP81RT-F (5'TGTATCGCCCACACGTTCACT3') and CtCYP81RT-R (5'TTTCCGGCAGGTCCTTTGTT3') was used for qRT-PCR analysis of CtCYP81E8. All qRT-PCR reactions were conducted on a Stratagene Mx3000P system (Stratagene, CA, USA) using SYBR Green kit (TaKaRa, China). The thermal cycle for our qRT-PCR reactions were as follow: one cycle of 5 min at 95 °C, followed by 35 cycles of 10 s at 95 °C, 10 s at 58 °C, and 30 s at 72 °C. The safflower 18SrRNA gene was kept as an internal standard. The expression level was normalised and calculated according to 2 − $\Delta\Delta^{Ct}$ method. All quantitative experiments were conducted in three biological and three technical replicates. The metabolite (SY) was quantified according to the aluminum chloride colorimetric method. After the effecient cloning of *CtCYP81E8*, we generated a CtCYP81E8-GFP-35S fusion construct simply by cloning the ORF of *CtCYP81E8* into the NcoI (CCATGG) and SpeI (ACTAGT) restriction sites of the binary vector pCAMBIA1302. The empty pCAMBIA1302 vector carrying GFP alone was used as a control. The investigation of the CtCYP81E8 subcellular localization was carried out in onion epidermal cells using the transient transformation sytem through agrobacterium mediated transformation prtocol according to the instructions of [24].

2.9. Plant material and stress conditions

Safflower seeds were grown in the artificial cloimate chamber of College of Life Sciences, Engineering Research Center of the Chinese Ministry of Education for Bioreactor and Pharmaceutical Development, Jilin Agricultural University, Changchun 130118, China under natural photoperiodic conditions. The stress conditions were configured with methyl jasmonate 300 mmol/L by spraying it on the 3-week-old full flower petals of safflower, set the sampling time to 0 h, 2 h, 4 h, 6 h, 8 h, and 12 h respectively. Drought stress was induced with 20% PEG-6000 with sampling time ranging between 12 h-72 h, every 12 hours is a time gradient. In case of dark and light treatment groups, the plantlets were grown in a dark box to avoid light treatment, the light and dark stress sampling time was seleced between 12 h-60 h. The fresh samples were collected at their corresponding time and subjected to frozen quickly in liquid nitrogen, and stored at − 80 °C until further use. The total RNA content of flower tissue of safflower under four stress conditions at different times was extracted using RNAiso PLUS reagent. The first strand cDNA synthesis was performed using reverse-transcription reaction. The reverse-transcribed cDNA templates were used for subsequent qRT-PCR analysis under different stress consitions. The reactions from qRT-PCR have all been carried out using the Fast Real-Time PCR Method (Applied Biosystems, CA, USA) with a 20μL final reaction volume containing 10μL SYBR Premix Ex Taq (Tli RNaseH Plus) (TaKaRa).

2.10. Correlation analysis CtCYP81E8 transcription and total metabolite accumulation in wild and mutant safflower

A compartive analysis between wild and mutant safflower lines was used to unviel the relationship of CtCYP81E8 mRNA transcription with the accumulation of total metabolite content in red-typed and yellow-typed flowers. For this purpose, the flower petals of both red-typed and yellow-typed flowers were collected at four different flowering stages including bud, initial, full, and fade stage in wild safflower line using similar conditions. However, the mutant line was only tested for bud, initial, and full flowering stages as it lacks the phenotype of fading flower stage due to mutation. These petals were immediately transferred to liquid nitrogen after the collection of each replicate in already marked separate tubes. The experimental material of each flower type and safflower line was simultaneously subjected to RNA extraction and total metabolite extraction. After, the total RNA content extraction, cDNA templates were synthesized using reverse transcription PCR. The qRT-PCR assays were conducted in accordance to previously indicated system to determine the transcription regulation level of CtCYP81E8 at each flowering type and phase development of wild and mutant line. All experiments were conducted in three independent replicates at each growth stage, and the results were analyzed according to 2 − ΔΔCt. The 18 s ribosomal RNA gene (GenBank accession: AY703484.1) was used as a reference. At the same time and conditions, the remaining homogeneous mixture of the flower petals was immersed in 14 mL water-alcoholic solution for ultrasonication separation of the metabolites under the controlled conditions including 60 °C of extraction temperature, 30 min twice of extraction period, and 5000 rpm cycle of centrifugation for 10 min. A specimen of 0.5 mL (1 mg /mL) was subjected to mixing with 10% aluminium chloride and 1 M potassium acetate in addition 80% methanol solution. The absorbance measurments was performed spectrophotometrically at wavelength of 415 nm. The measurment of the percentage total flavonoid content (TFC) was calculated using the method of milligrams of TFC per 100 grams fresh weight or dry weight. Three biological replicates (n = 3) were used to minimize the risk of the possible error.

2.11. Prokaryotic expression and in vitro DMP activity of CtCYP81E8

The full-length CtCYP81E8 cDNA was amplified using Pfu DNA polymerase (Takara) using a different set of primers CtCYP81PE-F (5'CGGATCCGATGATGAGGATGATTAGTGG3') with an added EcoRI(GAATTC) site and CtCYP81PE-R (5'TCAAAGATGCGATAATAGATTTGGAATTCC3') with an added BamHI(GGATCC) site. The construction of the binary vector (pET28a+) was carried out using the double restriction digestion of CtCYP81E8 and pET28a + vector. Subsequently, the ligation of the CtCYP81E8 into the appropriate EcoRI, and BamHI restriction sites of the empty pET28a + vector was performed withn T4 ligase enzyme. The binary vector (Pet28a+-CtCYP81E8) was transformed into BL21 (E. coli cells). The CtCYP81E8 protein was effectively induced and then expressed. The BL21 cells transformed with pET-28a+-CtCYP81E8 were grown in LB media (500 mL) supplemented with 50 mg/L of kanamycin. The bacterial cells were harvested at an OD of 2 × 108 cells/mL (A600 = 1.0), followed by sonicaion with 0.4 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG) under controlled conditions. The bacterial cultures were extracted by eradicating the supernatants, and the only pellet was resuspended in a PBS buffer followed by centrifugation at 12,000 × g for 10 min at 4 °C. The Lysis buffer was used to collect the bacterial cells. The ultrasonication was folloed in three intervals for 15 s until soluble fractions. The soluble protein product of CtCYP81E8 was separated on 12% SDS-PAGE, and the expected bands were stained using Coomassie brilliant blue and the expected protein product was purified using western blot hybridization method. Three independent biological and two technical replicates were analyzed for all measurement. In addition, 2,4-dimethylphenol activity (DMP) test was emplyed to check the in vitro activity of CtCYP81E8 by measuring the dissolved oxygen concentration of the reaction mixture at various time periods. A mixture of hydrogen peroxide 5 ul, enzyme solution

30 ul, 100 Mm DMP 20 ul, citrate buffer solution 145 ul, mixed in the enzyme label strip, was fully reflected for 5 minutes, and finally measured the OD278 values. The otherwise conditions include a shaker incubator fixed at 28C, the time periods ranges from 0, 12 h, 24 h, 36 h, 48 h, 60 h, and 72 h and the centrifugation cycle pertains 10000 rpm at 4 °C for 5 minutes, with twice repeat.

3. Results

3.1. Identification and physicochemical properties CtCYP81E subfamily in safflower

In total, 15 full-length safflower CtCYP81E genes with *c. tinctorius* location markers were extracted from the draft genome database selected on the basis of P450 Pfam00067 domain. In order to further classify CtCYP81E genes into subfamily, we extensively carried out comparitive analysis with the overall P450 genes of *Arabidopsis* genome using Phytozome 4.0. A total of 246 P450s and 26 pseudogenes were retrieved for further characterization. In addition, the physicochemical properties were also determined with the help of ProtParam online tool (Table 1). The result of the protein size for all 15 CtCYP81E encoded amino acids was found between 115–516 amino acids. The expected molecular weight was recorded theoretically resulting within the range of 12.97 kDa (CtCYPE9) to 59.01 kDa (CtCYPE14), with an average value of 45.59 kDa. The values of isoelectric points (pI) ranged from 4.74 (CtCYPE9) to 9.35 (CtCYP81E10). Furthermore, the grand average of hydropathicity (GRAVY) index was also determined revealing that most of the CtCYP81E proteins were allocated to hydrophilic nature. Out of all, the most stable protein was CtCYPE4, comprising a stability index equals to 40.66.

Table 1
The physiochemical properties of safflower CYP81E subfamily proteins. The data was collected using the online tool of ExPASy (available online: http://web.expasy.org/protparam/).

Gene Name	Gene ID	Protein Length	PI	MW(Da)	Subcellular Localization	IS index	*Arabidopsis* Homology	GRAVY
CYP81E1	CCG011365.1	516	8.38	58237.60	Plasma membrane	49.02	AT4G37370.1	-0.111
CYP81E2	CCG011366.1	501	8.96	57297.86	Plasma membrane	46.06	AT4G37370.1	-0.125
CYP81E3	CCG011367.1	501	8.73	57298.61	Plasma membrane	43.78	AT4G37370.1	-0.162
CYP81E4	CCG011368.1	339	5.61	38282.11	Plasma membrane	40.66	AT1G66540.1	-0.332
CYP81E5	CCG007304.1	506	6.78	57579.54	Plasma membrane	53.61	AT4G37370.1	-0.173
CYP81E6	CCG007305.1	506	6.78	57579.54	Plasma membrane	53.61	AT4G37370.1	-0.173
CYP81E7	CCG007306.1	440	6.45	50001.45	Plasma membrane	54.20	AT4G37370.1	-0.255
CYP81E8	CCG007309.1	337	6.49	38489.90	Plasma membrane	44.14	AT1G66540.1	-0.320
CYP81E9	CCG007984.1	115	4.74	12976.39	Plasma membrane	70.48	AT4G37370.1	-0.640
CYP81E10	CCG008862.1	246	9.35	27858.38	Plasma membrane	47.65	AT4G37320.1	-0.147
CYP81E11	CCG008863.1	177	6.83	19946.29	Endoplasmic reticulum	47.57	AT1G66540.1	-0.194
CYP81E12	CCG013393.1	435	8.49	49493.63	Plasma membrane	44.98	AT4G37360.1	-0.143
CYP81E13	CCG013395.1	437	8.84	49746.09	Plasma membrane	42.55	AT4G37340.1	-0.127
CYP81E14	CCG021962.1	516	7.68	59005.54	Plasma membrane	45.29	AT4G37340.1	-0.156
CYP81E15	CCG004067.1	434	8.41	50124.07	Plasma membrane	46.92	AT4G37370.1	-0.286

3.2. Phylogenetic reconstruction of CtCYP81E and Arabidopsis P450 genes

A neighbor-joining (N-J) phylogenetic tree was created using 15 CtCYP81E candidate protein sequences obtained from safflower genome and 272 Arabidopsis P450s (Table S2). The division of cluster groups between these two plant species by means of P-distance method determined their phylogenetic origin using MEGA-X package [25]. The clustering of P450 genes were initially divided into two widely known classess known as A-type and non-A-type P450 sequences. These two clades were further subdivided into nine different clans including, clan71, clan51, clan710, clan85, clan711, clan86, clan97, clan72, and clan74 (Fig. 1). The CYP71 clan was found the largest A-type

class which contains 131 genes (48.51%) succeeded into 10 more subfamilies such as CYP71AH, CYP71AT, CYP71AU, CYP71AX,CYP71D, CYP71BE CYP71BG,, CYP71BL, CYP71BN and CYP71BP. Mostly, these subfamilies of CYP71 clan symbolize the presence of plant-specific enzymes that are involved during secondary metabolic reactions mainly in flavonoid biosynthesis. Our phylogenetic analysis indicated that CtCYP81E gene family of safflower is consistently clustered with the CYP71 clan of Arabidopsis. Hence, it is possible that the CtCYP81E subfamily in safflower is most likely involved in flavonoid biosynthesis as found in the CYP71 clan of the model plant. [26, 27].

3.3. Analysis of CtCYP81E gene structure, motifs and promoter

The details analysis of CtCYP81E subfamily including gene organization, signature motifs and cis regulatory elements were carried out. As described in (Fig. 2), most of the CtCYP81E genes (Table S3) shared a common organization of exon/intron makeup. Nonetheless, a small number of CtCYP81E genes did not contain the basic gene structures. For instance, the gene structure organizations of CtCYP81E6 and CtCYP81E1 showed that both of these genes consist of a short exon and long intron when compare with the other candidate genes present in the same family. Our findings were found consistent with CYP450 organization in *A. thaliana* [28] which confirrns that the number of exons in CYP71 clans of *A. thaliana* was ranged between 2–5. In addition, the shortest exon (27 bp) in the CYP71 clan (CYP71B32) of *Arabidopsis* was found longer than the shortest exon of CtCYP81E6 (16 bp) in safflower. Safflower CYP81E genes most likely contain 1–3 exons as demonstrated in (Fig. 2B). Mostly, these genes shared two (60%, 9/15), three (26.6%, 4/15), and one (13.3%, 2/15) of exon arrangments. Whereas the length of the introns in safflower P450 genes were estimated from 126 to 5946 bp indicating parallel results with the A. thaliana [28] and C. elegans [29].

In the next level, all 15 CtCYP81E genes from saflower were analyzed for the identification of naturally conseved protein motifs. The multiple sequence alignemt of these proteins demonstrated that almost all 15 candidate safflower P450 genes contained the basic signature motifs of P450 family such as heme binding region, PERF region, K-helix region, and I-helix motif (Fig. S1). A total of 10 conserved motifs of CtCYP81E proteins have been found consistent with the Arabidopsis P450s MEME online investigation. The output results implied that nearly all candidate CtCYP81E proteins contained these conserved motifs exluding CtCYP81E4, CtCYP81E8, CtCYP81E9, CtCYP81E11, which described a different conservation pattern. In addition, the position of few conserved motifs in CtCYP81E was not inlined with the *Arabidopsis* P450 s. The sequences and organization of these signatory motifs of CtCYP81E proteins were demonstrated in (Fig. 2C). These results suggested that CtCYP81Es in safflower inherit basic structural and functional domians during the process of evolution. Moreover, the cis- regulatory elements organizations of these CtCYP81Es specifically in the 2 kb 5'flanking region upstream to the start codon (Table S4) was thoroghly analyzed. Altogether, six major types of cis-elements were found in the 2 kb 5'flanking region of the promoter (Fig. 2D). Among these genes promoter, the result of a few CtCYP81Es members contain endosperm expression elements (AACA_motif; GCN4_motif), some of these CtCYP81Es contained element essential for the anaerobic induction (ARE), another group of CtCYP81Es showed the presence of abscisic acid (ABA) responsive element (ABRE), and low-temperature responsiveness elements (P-box; TATC-box), while the remaining CtCYP81Es consist hormonal responsive elements (methyl jasmonate (MeJA), and light responsiveness element (G-box; AAAC-motif). Conclusively, the occurrence of these cis-elements identified in safflower CtCYP81Es genes indicated that they are most likely to be involved during plant growth, development and plant adaptation to various stress responses and hormone signal pathways.

3.4. Functional classification of the safflower CtCYP81E subfamily

The functional categorization of the 15 CtCYP81E transcripts in safflower was carried out using GO analysis. The results of the in silico classfication revealed that all 15 CtCYP81E transcripts were allocated into one or additional GO terms. These CtCYP81E transcripts were found in all three fundamental functional categories including biological process indicated as (BP), molecular function indicated as (MF), and cellular component indicated as (CC). Moreover, eight functional subcategories were also demonstarted in the next level wherein two CC subcategories: integral compoent of membrane and membrane were detected. Five MF subcategories: oxioreductase activity, iron ion binding, heme binding, mrtal binding and isoflavone 2'-hydroxylase activity; and one BP subcategories: oxidation-reduction process (Fig. 3). Despite the fact that CtCYP81E gene belongs to type A, as previously reported, there is no significant difference found in the functional annotation of the type A and non-type A P450 sequences [30].

3.5. Protein Clustering Networks

The monooxygenases (cytochrome P450) play important roles in xenobiotic metabolism and biosynthesis of internal nutrients such as flavonoids, vitamins, steroids, hormones, and fatty acids. The capability of the P450 encoded enzymes to catalyze important substrates that involve interaction with its redox protein counterparts. This biochemical catalysis can be altered in association with the membrane-bind heme protein cytochrome b5 [31]. With the help of AtCYP81E orthologous, we systematically predicted the PPI interaction network of CtCYP81Es subfamily in safflower. We confirmed 10 widely spread proteins co-associated with these CtCYP81Es, which include translocation (2), membopane lipoprotein (1), aquaporin-like (1), ABC transporter (1) and protein kinase (1) (Fig. 4). The independent interactor protein networks indicated that CtCYP81E protein (1, 2, 3, 5, 6, 7, 9 and 15) interacts with the UGT74E2 protein, which is mainly involved in the biosynthesis of IBA (indole-3-butyric acid) and directly influence the homeostasis of auxin. Additionally, CtCYP81E proteins (1, 2, 3, 5, 6, 7, 9 and 15) work together with AT5G25930 proteins which are largly associated with the protein amino acid phosphorylation. The CtCYP81E proteins (4, 8 and 11) interact with the AT5G48605 protein and can enhance plant defense mechanism. Besides this, CtCYP81E protein (4, 8 and 11) interacts with the AT1G59660 protein which acts as key regulator in the water channel. Notably, we found that other CtCYP81E orthologous except CtCYP81E (4, 8 and 11) proteins interact with ABCD1 proteins and may be involved in the transportation mechanism. The PPI network of CtCYP81E orthologous highlights its potential role in several physiological and biosynthetic process occurred simultaniously in plants system.

3.6. Expression analysis and functional annotation of CtCYP81E subfamily genes

Expression levels of P450 variants in safflower were initially determined with the help of RNA-seq data (whole Transcriptome Shotgun Sequencing) in different tissue specifying five selected tissues/organs (root, stem, seed, flower and leaf tissues). The expression level was calculated according to kilobase model of exon model per million mapped read (RPKM) method according to the instruction given by (Mortazavi et al. 2008). The RNA-seq data was obtained from the safflower genome database (PRJNA399628; posted to NCBI on August 23, 2017). In general, the expression signals of almost all selected safflower P450 genes were detected in all organs but with different patterns. As revealed in (Fig. 5A), the expressed P450 genes in safflower were clustered into five groups including, G1 (6.6%, 1/15), G2 (6.6%, 1/15), G3 (33.3%, 5/15), G4 (33.3%, 5/15), and G5 (13.3%, 2/15) that were more preferably expressed in the leaves, stems, seeds, flowers, and roots, respectively. Furthermore, to validate the transcript abundance of CtCYP81E genes and their correlation in biosynthetic processes, we extensively carried out qRT-PCR analysis of these 15 genes at different flowerering stages such as bud, initial, flower and fading stage.

Across CtCYP81E subfamily, the expression level of CtCYP81E2, CtCYP81E8, and CtCYP81E15 were abundantly detected at flower stage indicating that there might be a strong link between the regulation of transcription of CtCYP81E genes and cellular metabolism in safflower.In addition, the transcripts of CtCYP81E1, CtCYP81E2, CtCYP81E5, and CtCYP81E7 were identified in high expression level at the fading stage of flowering suggesting the transcription regulation of these genes at a later flower developmental period. The transcripts of CtCYP81E14, and CtCYP81E15 showed high expression level at initial flowering of safflower (Fig. 5B). Altogether, the qRT-PCR assay suggested a differential expression pattern and fold-change values of the selected CtCYP81E subfamily genes highlight their decisive roles in plant defense systems and developmental processes.

3.7. Subcellular localization and transcriptional regulation of CtCYP81E8

Based on our previous study on a CtCYP82G24, [32], we aimed to investigate the correlation between the quantitative expression trend of CtCYP81E8 gene and metabolite accumulation at different flowering stages of safflower. As described in (Fig. 6A&B), the expression level of CtCYP81E8 was detected consistent with the accumulation rate of total metabolites content in safflower petals. These findings provide a practical basis for the funtional characterization of CtCYP81E8, which could be a crucial modulator in the biosythetic pathway of flavonoid biosynthesis in safflower. Taking into consideration the functional importance and differential expression pattern of CtCYP81E8 gene, we therefore, cloned the full length sequence of CtCYP81E8 from safflower (Fig. 6C) and then constructed a fusion vector of CtCYP81E8 and GFP gene under the control of the 35S promoter (pCAMBIA1302-CtCYP81E8-GFP-35S) in order to determine the experimental subcellular localization. After the efficient construction of the plant overexpression vector fusion, the recombinant vector was then transiently transformed intothe onion epidermal cells through agrobacterium mediated transformation system. Fluorescence imaging of infected epidermal cells of onion bombarded with CtCYP81E8-GFP showed cell membrane localization (Fig. 6D). These findings revealed important evidence to support the assumption that CtCYP81E is able to catalyze cellular based biological reactions occurred in *Carthamus tinctorius*.

3.8. The induction of CtCYP81E8 transcription under variable stress conditions

The transcriptional regulatory network of CtCYP81E8 mRNA under variable stress environments has been demonstrated to confirm the underline notion of Cytochrome P450s involvement during a variety of plant secondary metabolites biosynthesis (Mizutani) Ohata 2010; Nelson, Werck-Reichhart 2011). By exploiting the temporal transcriptional regulatory channels of CtCYP81E8 under artificial environmental switches, we demonstrated a multi-regulation control system using qRT-PCR assays. The treatment group with methyl jasmonate at 0–12 hours, compared with the control group, the expression level of CtCYP81E8 gene showed an upward trend, among which the expression of CtCYP81E8 showed a unique increase at 8 h where, the transcription level was reached to its maximum. In contrast, at 12 h timepoint, the expression decreased significantly (Fig. 7A). Under drought stress conditions, the CtCYP81E8 gene expression was significantly induced at 4–8 h than the control plants. The expression level was reached to its maximum at 8 h timepoint under PEG induced stress however, the transcription of CtCYP81E8 was down-regulated at 12 h treatment times (Fig. 7B). Under strong light irradiation, the gene expression level of CtCYP81E8 at different treatment times 12–60 h was surprisingly down-regulated compared with control plants. The down-regulation was most significant at 36 h treatment time indicating intense susceptibility towards light stress (Fig. 7C). The transcription level of this gene after dark treatment was expectedly upregulated in

all treatment times reaching to the maximum at 36 h. In general, the expression level was consistently rising from 12–36 h, and suddenly drops sharply after 48 h, but the overall expression level is up-regulated compared to the control group (Fig. 7D). These findings unanimously represented the multi-dimensional periodic regulatory network of CtCYP81E8 transcriptional system upon different stress conditions, highlighting cruicial blueprints in the molecular regulation system of plants adaptation to biotic and abiotic stress responses.

3.9. Transcriptional regulation system CtCYP81E8 overlapping with flavonoid accumulation in wild and mutant safflower

The correlation between the transcription level of CtCYP81E8 and accumulation pattern of total metabolite content through multiple flower developmental stages of wild and mutatnt safflower varieties was extensively investigated using qRT-PCR assay. Simultaneously, the accumulation content of total metabolites was purposely investigated using the same phases of the two naturally occuring flowers types in safflower including red and yellow flower. Interestingly, the expression profile of CtCYP81E8 showed a programmed expression system correspondantly during flowering developmental phases both in wild and mutant type of safflower (Fig. 8A&B). The transcription control level of CtCYP81E8 during the red flowering development stages except at the bud flowering phase (R1), confirmed that the increased trend of CtCYP81E8 transcript simultaneously affect the accumulation level of total metabolite content in red-typed wild safflower. In the same way, the yellow-typed wild safflower showed a consistent network of increased trend in the accumulation of metabolite content with the increase in the expression level of CtCYP81E8 excluding the bud flowering stage (Y1). The estimated theme was further confirmed by conducting simillar analysis in the mutant safflower line, suggested almost a similar type of correlation, however, the opposite trend was also found as in the wild type safflower but through a different flower developmental stage. It was suggested that the accumulation content of safflower metabolites was significantly increased with the increase in the transcription level of CtCYP81E8 during all three flowering stages of the red-typed mutant safflower. Nonetheless, a discontinued scheme of the metabolite accumulation in the white-typed mutant safflower was observed at bud flowering (M5) and full flowering (M7) stages with an axception to intitial flowering phase (M6), indicating a reverse order in comprison to their corresponding flower stages. As mentioned earlier, the expression level of CtCYP81E8 was significantly exploited under different stress conditions, indicating the concept of secondary metabolic activation in plants under variable abiotic stress conditions. Conclusively, these results insistingly suggested that the transcription regulation of CtCYP81E8 has a certain relationship with the accumulation profiling of metabolite content of different safflower varieties. Though assumuingly, but these findings could be cruicial in understanding the core concept of molecular regulatory signals that strategically switch on the secondary metabolic flux by intervening through a bulk of genetical and particularly, transcriptional events, to ensure plant's survival under acute environmental drifts.

3.10. Heterologous Expression and in vitro enzymatic assay of CtCYP81E8

In order to validate the potential function of CtCYP81E8 in vitro, the full length cDNA of CtCYP81E8 was cloned into the prokaryotic expression vector (pET28a+), and then transformed into E. coli BL21DE3 cells by thermal and electric shock transformation, and then induced by adding different concentrations of IPTG. The bacterial solution without IPTG induction and no load were used as control. The heterologous expression of CtCYP81E8 recombinant protein was mainly detected by Coomassie blue staining SDS-PAGE and Western Blot hybridization. The analysis of SDS-PAGE showed that the recombinant CtCYP81E protein was expressed at the 36.4 KDa site, but it seemed that the

concentration of IPTG did not affect the protein expression (Fig. 9A). Then we purified the recombinant CtCYP81E8 protein and further identified the expression of the protein at different IPTG concentrations by western blot hybridization. As shown in (Fig. 9B), we found a single purpose band and a change in the protein expression with the concentration of IPTG induction. From our findings, we deduced that target protein of CtCYP81E8 was stably detected on SDS-PAGE and western blot hybridization on nylon membrane, moreover, the product size was also consistent with theoretical molecular weight of CtCYP81E8 protein (36.4 kDa), suggesting that the target protein was effetiently expressed in prokaryotic system, however, the different concentrations of IPTG could potentially affect the expression level of the target protein [16].

The primary objective of the DPM assay was designed to explore the complete consumption of oxygen by CtCYP81E8 recombinant enzyme for the oxidation of 100 mM DMP by direct absorbance method under different time periods. For the present study, by adding hydrogen peroxide in the reaction, the highest removal of 100 mM DMP concentrations (60%) in CtCYP81E8 batch was variably detected including the optimum removel at 24 h followed by 48 h respectively as compare to the control group (Fig. 10). For 12 h and 72 h, the removal of 100 mM DMP was less found than 15–40%, but it was also observed that after 36 h reaction the removal of DMP was reached to almost 40%. These findings depicts that insufficient absorbance of oxygen in a uniform reaction batch may be due to the poor efficiency rate. It was also suggested that the rate of reaction was independent of dissolved oxygen at the start of the reaction, until enough oxygen was present. But, after the sufficient utilization of the dissolved oxygen in the aforesaid reaction, the dependency of the reaction becomes essential for exogenous oxygen addition. Hence, more efforts are still needed to provide further insights in obtaining more efficient removal of DMP during in vitro activity assay.

4. Discussion

4.1. Evolutionary classification of CtCYP81E genes in safflower

The complete genome sequencing of the Arabidopsis model plant (Arabidopsis Genome Initiative, 2000) [33] has broadened the genome wide identification studies of functionally important gene families. Among which, a stricking on already discovered in Arabidopsis is largest enzymne-encoding Cytochrome P450 monooxygenases gene family. In this study, we also conducted a comparitive genome wide study of the putative CtCYP81E subfamily in safflower with Arabidopsis P450s to investigate their ancestral relationship by studying the annotation result of similar clans of related P450 families. The clustering of 15 CtCYP81E enzyme encoding sequences were found on the largest A-type clan CYP71 clan supporting our hypothetical notion as most of CYP71 subfamilies symbolize a group of enzymes involved during secondary metabolic biosynthesis. The CYP71 clan was found the largest A-type class which contains 131 genes (48.51%) succeeded into 9 and with the inclusion of the new safflower CtCYP81E sequences, in total 10 clans such as CYP71AH, CYP71AT, CYP71AU, CYP71AX,CYP71D, CYP71BE CYP71BG,, CYP71BL, CYP71BN and CYP71BP (Fig. 1). Our Phylogenetic reconstructions revealed that four clans namely clan51, clan710, clan711 and clan74 belong to the same family clans, however, the rest of the five clans covers various other families of P450 genes [34, 35]. The CYP72 clan contains eight subfamilies suggested that CYP72 clan is the largest non-A family comprising 20 genes (13.70%). Further classification of CYP72 clan was catogorised into two subgroups including CYP72A and CYP72D. The sequences encoded by the non-A-type clan surround 139 of the 270 sequences and also a group of other enzymes paticipating during the biosynthesis of primary and secondary metabolic compounds in plants for example sterols, fatty acids, hormones and other signaling molecules [36]. Lastly, CYP74 clan contains four different subgroups which is designated as outgroup in our phylogenetic tree becuase it is an atypical plant P450 clan which doeos not contains monooxygenase activity.

4.2. Hierarchy of gene structure, conserved motifs and cis-regulatory classification of CtCYP81E genes in safflower

Intron-exon organization and their pattern of gain and loss mutations greatly highlight the evolutionary mechanism of cetain gene families falls within the same phylogenetic clade. The understanding of the conserved introns organization likely offer ancient elements to understand similar group of genes involved in multple physiological processes of plants [37]. During our analysis, the safflower CtCYP81E subfamily genes mostly contain 1–3 exons and two introns that were found comparable with clan71 gene families of Arabidopsis and mullbery genome [28, 38] (Fig. 2B). In addition, it was also found that the two conserved introns were not detected in the non-A type Arabidopsis P450 gene families as compare to A-type clan71 gene families indicating the significant course of intron evolution during gene families organizations [39, 40].

Cytochrome P450s comprise comparitively a different pattern of amino acid conservation [41], however, the universal topology of the secondary and tertiary structures and other basic signatory regions/motifs remains consistent throughout plant kingdom. In this study, safflower CtCYP81E subfamily also revealed the occurrence of the five well known P450 motifs containing heme-binding region (PFxxGxRxCxG/A), C-helix (WxxxR), PERF motif (PxxFxPE/DR), K-helix (ExLR), and I-helix motifs (GxE/DTT/S) [36, 42]. Furthermore, the overspread conservation of the specific amino acid groups including tryptophan (W) and arginine (R) in the conserved C-helix motif, glycine (G) and threonine (T) in the conserved structure of I-helix motif, phenylalanine (F), glycine (G), arginine (R) and cysteine (C) repeatition inside the conserved signature of the heme-binding motif, glutamic acid (E) and arginine (R) of the widely spread K-helix region, and proline (P) residues inside the conserved PERF band were consistently found throughout CtCYP81E subfamily in safflower (Fig. 2C).

Moreover, the conservation and distribution of cis-regulatory elements in CtCYP81E promoters suggested the presence of widely known stress-responsive regulatory units containing low-temperature responsiveness elements (P-box; TATC-box), abscisic acid-responsive element (ABRE) [43], dehydration/drought-responsive element (DRE) [44], hormonal responsive elements (methyl jasmonate (MeJA) and C-repeat [45] (Fig. 2D). The conservation of such important cis regulatory units within the promoter regions of several gene clusters such as NAM, ATAF, and CUC (NAC) genes [46] have suggested their stress tolerance potential under extreme climatic changes. The RNA-seq results of CtCYP81E expression showed a slightly up-regulated pattern in flower petals and the presence of overrepresented endosperm expression elements (AACA_motif; GCN4_motif) and TCP transcription factor elements in the promoter region, which is well studied in flower developmental growth and different hormonal biosynthesis reactions [47–49]. Therefore, CtCYP81E subfamily genes could be crucial candidate genes in safflower underlining the core concept of floral development and secondary metabolism. Conclusively, the overall survey of gene structure compositions combined with widely identified conserved motifs and commonly spread cis-regulatory units of CtCYP81E candidate subfamily of safflower demonstrated unique evolutionary pattern which further put emphasis on the functional dynamics of these genes during plant adaptation to various stress responsesand other crucial secondary biosynthetic pathways.

4.3. Transcriptional regulation of CtCYP81E genes during floral development in safflower

The gene expression at transcription level deeply relies on phases of plant growth and development, age, environmental influences, degree of expression, tissue variability, and various biotic and abiotic stress responses. In the present study, we also tend to investigate the undelying molecular regulatory network of CtCYP81E genes at the

transcriptional level. The RPKM data obtained from RNA-sequencing suggested a distinct expression profiling of CtCYP450 genes in safflower by clustering into five main groups including, G1 (6.6%, 1/15), G2 (6.6%, 1/15), G3 (33.3%, 5/15), G4 (33.3%, 5/15), and G5 (13.3%, 2/15) detected in various tissues/organs such as leaves, stem, seed, flowers, and root respectively (Fig. 5A). These results were found compatible with the results of vasu et al 2019 who described that 31.33% of Solanum lycopersicum P450 genes demonstrate differential expression profiling through different tissues/organs development. Our findings also provides close proximity with soybean (31.92%) [19], mulberry (23.6%) [38], and rice (49.81%) [50] P450 genes exhibiting five major groups of expression pattern were found. Furthermore, the transcriptional regulation of CtCYP81E8 in different flowering stages of safflower under normal and challenging climate conditions was expoloited to unleash their molecular regulatory channels during plant adaption system by activating the secondary metabolic signals. Our results significantly depicts the regulation of CtCYP81E8 mRNA abundance through multiple flower developmental phases in safflower during temporal exposure to hormonal (MeJA) and various abiotic stress conditions including drought, light and dark environments (Fig. 7). The specific outcome from this study represented the multi-facet and periodic regulation system of CtCYP81E8 transcriptional channels during developmental phases of the flower tissues encountering a variety of climate shifts, emphisizing essential elements in the quest of plant adaptations against biotic and abiotic stimuli. These findings were effectively supported by our recent studies on CtCYP82G24, [32], which outlined the core concept of the correlation analysis between quantitative gene expression trend and metabolite accumulation pattern under abiotic stress encounter in transgenic Arabidopsis. In addition to this, we also reported the upregulation of CtCYP82C1 transcription under artificial hormonal and abiotic stress exposure [51]. Even so, these findings facilitate practical basis underlining the funtional characterization of CtCYP81E8 with a particular focus on flavonoid biosynthesis in safflower, more robust approaches are however, required to screen out appropriate candidate genes clustered together from a large repertoire of Cytochrome P450 supergene family in safflower for extended functional characterization studies.

4.4. Functional dynamics of CtCYP81E8 and heterologous expression

Cytochrome P450 is the largest monooxygenase superfamily found in all forms of life, however, the genral tendency of this multigene familly in plants revolves around the distinct metabolic pathways through different developmental stages. So far, the primary functional characterization of the CYP71 clan subfamilies has been suggested to coexist during shikimate biosynthetic pathway [52–54]. Particularly, the investigation of P450 genes clusters involved in flavonoids and alkaloids biosynthetic pathways has also been demonstrated in other plants for instance, CYP73A and CYP93 subfamily genes [55]. The catalytic activity of P450s shares a common oxidative and reductive mechanism. However, to fully understand the wide range of substrate specificity of novel P450 enzyme encoding sequences, the development of a stable heterologous expression system is still a challenging mode. The adavatage of bacterial P450 expression systems over others have been comprehensively summarized by [56] providing a fast and robust expression system. This study also mimics the establishment of an efficient and stable expression system for the in vitro amplificatiom of the candidate CtCYP81E8 protein using prokaryotic machinery of bacterial system (Fig. 9). Our findings suggested an efficient system of the heterologous protein expression of putative CtCYP81E8 protein resulting in a product of 36.4 kDa size protein. The induction of the putative CtCYP81E8 protein with IPTG suggested a stable induction however, using variable IPTG concentrations could potentially influence the expression level of the target protein [16]. In addition, the DMP in vitro activity assay was also emplyed to tie the biochemical aspects of CtCYP81E8 reactions to their actual physiological and biosynthetic pathways which is not yet revealed (Fig. 10).

5. Conclusion

After the identification of cytochrome P450 superfamily, considerable knowledge has been reported about the unique biology of this special category of hemoprotein, however, it is essential to know further, the latest P450s molecular properties, particularly structural and functional discrepancies. Such unique properties can alter the effectiveness of the functional dynamics in specific P450s, which can explain new footprints that P450s may play besides their typical functions. In our own study, we suggested a comprehnsive structural and functional model of a putative member of CtCYP81E subfamily in safflower, providing crucial insights during the regulation of the adpative mechanism of plant secondary metabolites through a number of genetical, transcriptional and biochemical events. Though assumingly, these findings indicate considerable amount of information towards the partial functional identification of putative CtCYP81E8 member in safflower. Many efforts are, however, still required to understand the overall biological nature of CtCYP81E subfamily concerning its natural and functional diversity.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Competing Interest

The authors declare no conflict of interest.

Funding

This work was funded by grants from the National Natural Science Foundation of China (No. 31771868, 31501366), the Science and Technology Development Project of Jilin Province (No. 20190201172JC, 20190201175JC, 20150623024TC-11).

Authors' contributions

LX and HL conceived and designed this research. LJ and HY conducted experiments. WW, LY and ZX, MN analyzed data. MX, NH and YN perfomed formal analysis. NA wrote and edited the manuscript. All authors read and approved the manuscript.

References

1. China PC: Pharmacopoeia of the People's Republic of China: Chemical Industry Press; 2000.

2. Lou Z, Liu M: A pharmacognostical study of the Chinese drug, Honghua, Carthami tinctor II. *Acta Pharm Sin* 1956, 4:233-239.
3. Dai Y, Witkamp G-J, Verpoorte R, Choi YH: Natural deep eutectic solvents as a new extraction media for phenolic metabolites in Carthamus tinctorius L. *Analytical chemistry* 2013, 85(13):6272-6278.
4. Qu C, Wang L-Y, Jin W-T, Tang Y-P, Jin Y, Shi X-Q, Shang L-L, Shang E-X, Duan J-A: Comparative analysis of the effects of Hydroxysafflor yellow a and Anhydrosafflor yellow B in safflower series of herb pairs using prep-HPLC and a selective knock-out approach. *Molecules* 2016, 21(11):1480.
5. Wei X, Liu H, Sun X, Fu F, Zhang X, Wang J, An J, Ding H: Hydroxysafflor yellow A protects rat brains against ischemia-reperfusion injury by antioxidant action. *Neuroscience Letters* 2005, 386(1):58-62.
6. Yue S, Tang Y, Li S, Duan J-A: Chemical and biological properties of quinochalcone C-glycosides from the florets of Carthamus tinctorius. *Molecules* 2013, 18(12):15220-15254.
7. Zhu H, Wang Z, Ma C, Tian J, Fu F, Li C, Guo D, Roeder E, Liu K: Neuroprotective effects of hydroxysafflor yellow A: in vivo and in vitro studies. *Planta medica* 2003, 69(05):429-433.
8. Zhang Y, Guo J, Dong H, Zhao X, Zhou L, Li X, Liu J, Niu Y: Hydroxysafflor yellow A protects against chronic carbon tetrachloride-induced liver fibrosis. *European Journal of Pharmacology* 2011, 660(2-3):438-444.
9. Liu X, Dong Y, Yao N, Zhang Y, Wang N, Cui X, Li X, Wang Y, Wang F, Yang J: De Novo Sequencing and Analysis of the Safflower Transcriptome to Discover Putative Genes Associated with Safflor Yellow in Carthamus tinctorius L. *International journal of molecular sciences* 2015, 16(10):25657-25677.
10. De Montellano PRO: Cytochrome P450: structure, mechanism, and biochemistry: Springer Science & Business Media; 2005.
11. Nelson DR: Cytochrome P450 and the individuality of species. *Archives of biochemistry and biophysics* 1999, 369(1):1-10.
12. Fischer M, Knoll M, Sirim D, Wagner F, Funke S, Pleiss J: The Cytochrome P450 Engineering Database: a navigation and prediction tool for the cytochrome P450 protein family. *Bioinformatics* 2007, 23(15):2015-2017.
13. Morant M, Bak S, Møller BL, Werck-Reichhart D: Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. *Current opinion in biotechnology* 2003, 14(2):151-162.
14. Schuler MA: Plant cytochrome P450 monooxygenases. *Critical reviews in plant sciences* 1996, 15(3):235-284.
15. Baldwin WS, Marko PB, Nelson DR: The cytochrome P450 (CYP) gene superfamily in Daphnia pulex. *BMC genomics* 2009, 10(1):1-12.
16. Wei K, Chen H: Global identification, structural analysis and expression characterization of cytochrome P450 monooxygenase superfamily in rice. *BMC genomics* 2018, 19(1):35.
17. Li Y, Wei K: Comparative functional genomics analysis of cytochrome P450 gene superfamily in wheat and maize. *BMC plant biology* 2020, 20(1):1-22.
18. Wang Y, Li X, Lin Y, Wang Y, Wang K, Sun C, Lu T, Zhang M: Structural variation, functional differentiation, and activity correlation of the cytochrome P450 gene superfamily revealed in ginseng. *The plant genome* 2018, 11(3).
19. Guttikonda SK, Trupti J, Bisht NC, Chen H, An Y-QC, Pandey S, Xu D, Yu O: Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases. *BMC plant biology* 2010, 10(1):243.
20. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and*

evolution 2011, 28(10):2731-2739.

21. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G: GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 2014, 31(8):1296-1297.
22. Conesa A, Götz S: Blast2GO: A comprehensive suite for functional analysis in plant genomics. *International journal of plant genomics* 2008, 2008.
23. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. *methods* 2001, 25(4):402-408.
24. Yang J, Ji L, Wang X, Zhang Y, Wu L, Yang Y, Ma Z: Overexpression of 3-deoxy-7-phosphoheptulonate synthase gene from Gossypium hirsutum enhances Arabidopsis resistance to Verticillium wilt. *Plant cell reports* 2015, 34(8):1429-1441.
25. Kumar S, Stecher G, Tamura K: MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution* 2016, 33(7):1870-1874.
26. Otani K, Takahashi T, Furuya T, Ayabe S: Licodione synthase, a cytochrome P450 monooxygenase catalyzing 2-hydroxylation of 5-deoxyflavanone, in cultured Glycyrrhiza echinata L. cells. *Plant Physiology* 1994, 105(4):1427-1432.
27. Akashi T, Aoki T, Ayabe S-i: Cloning and functional expression of a cytochrome P450 cDNA encoding 2-hydroxyisoflavanone synthase involved in biosynthesis of the isoflavonoid skeleton in licorice. *Plant Physiology* 1999, 121(3):821-828.
28. Paquette SM, Jensen K, Bak S: A web-based resource for the Arabidopsis P450, cytochromes b5, NADPH-cytochrome P450 reductases, and family 1 glycosyltransferases (http://www. P450. kvl. dk). *Phytochemistry* 2009, 70(17-18):1940-1947.
29. Gotoh O: Divergent structures of Caenorhabditis elegans cytochrome P450 genes suggest the frequent loss and gain of introns during the evolution of nematodes. *Molecular biology and evolution* 1998, 15(11):1447-1459.
30. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: Gene ontology: tool for the unification of biology. *Nature genetics* 2000, 25(1):25.
31. Tan J, Tu L, Deng F, Hu H, Nie Y, Zhang X: A genetic and metabolic analysis revealed that cotton fiber cell development was retarded by flavonoid naringenin. *Plant Physiology* 2013, 162(1):86-95.
32. Ahmad N, Jianyu L, Xu T, Noman M, Jameel A, Na Y, Yuanyuan D, Nan W, Xiaowei L, Fawei W: Overexpression of a Novel Cytochrome P450 Promotes Flavonoid Biosynthesis and Osmotic Stress Tolerance in Transgenic Arabidopsis. *Genes* 2019, 10(10):756.
33. Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Lin X: Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *nature* 2000, 408(6814):796-815.
34. Nelson DR: Plant cytochrome P450s from moss to poplar. *Phytochemistry Reviews* 2006, 5(2-3):193-204.
35. Cramer CL, Edwards K, Dron M, Liang X, Dildine SL, Bolwell GP, Dixon RA, Lamb CJ, Schuch W: Phenylalanine ammonia-lyase gene organization and structure. *Plant molecular biology* 1989, 12(4):367-383.
36. Bak S, Beisson F, Bishop G, Hamberger B, Höfer R, Paquette S, Werck-Reichhart D: Cytochromes P450. *The Arabidopsis Book/American Society of Plant Biologists* 2011, 9.
37. Roy SW, Gilbert W: Rates of intron loss and gain: implications for early eukaryotic evolution. *Proceedings of the National Academy of Sciences* 2005, 102(16):5773-5778.
38. Ma B, Luo Y, Jia L, Qi X, Zeng Q, Xiang Z, He N: Genome-wide identification and expression analyses of cytochrome P450 genes in mulberry (Morus notabilis). *Journal of integrative plant biology* 2014, 56(9):887-901.

39. Barvkar VT, Pardeshi VC, Kale SM, Kadoo NY, Gupta VS: Phylogenomic analysis of UDP glycosyltransferase 1 multigene family in Linum usitatissimum identified genes with varied expression patterns. *BMC genomics* 2012, 13(1):1-13.
40. Paquette SM, Bak S, Feyereisen R: Intron–exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of Arabidopsis thaliana. *DNA and cell biology* 2000, 19(5):307-317.
41. Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, Waterman MR, Gotoh O, Coon MJ, Estabrook RW: P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* 1996, 6(1):1-42.
42. Du Y, Chu H, Chu IK, Lo C: CYP93G2 is a flavanone 2-hydroxylase required for C-glycosylflavone biosynthesis in rice. *Plant Physiology* 2010, 154(1):324-333.
43. Narusaka Y, Nakashima K, Shinwari ZK, Sakuma Y, Furihata T, Abe H, Narusaka M, Shinozaki K, Yamaguchi-Shinozaki K: Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of Arabidopsis rd29A gene in response to dehydration and high-salinity stresses. *The Plant Journal* 2003, 34(2):137-148.
44. Zhou M-L, Ma J-T, Pang J-F, Zhang Z-L, Tang Y-X, Wu Y-M: Regulation of plant stress response by dehydration responsive element binding (DREB) transcription factors. *African Journal of Biotechnology* 2010, 9(54):9255-9269.
45. An D, Ma Q, Wang H, Yang J, Zhou W, Zhang P: Cassava C-repeat binding factor 1 gene responds to low temperature and enhances cold tolerance when overexpressed in Arabidopsis and cassava. *Plant molecular biology* 2017, 94(1-2):109-124.
46. Sun H, Hu M, Li J, Chen L, Li M, Zhang S, Zhang X, Yang X: Comprehensive analysis of NAC transcription factors uncovers their roles during fiber development and stress response in cotton. *BMC plant biology* 2018, 18(1):150.
47. Kieffer M, Master V, Waites R, Davies B: TCP14 and TCP15 affect internode length and leaf shape in Arabidopsis. *The Plant Journal* 2011, 68(1):147-158.
48. Aguilar-Martínez JA, Poza-Carrión C, Cubas P: Arabidopsis BRANCHED1 acts as an integrator of branching signals within axillary buds. *The Plant Cell* 2007, 19(2):458-472.
49. Martín-Trillo M, Cubas P: TCP genes: a family snapshot ten years later. *Trends in plant science* 2010, 15(1):31-39.
50. Zhong L, Wang K, Tan J, Li W, Li S: Putative cytochrome P450 genes in rice genome (Oryza sativa L. ssp. indica) and their EST evidence. *Science in China Series C: Life Sciences* 2002, 45(5):512-517.
51. Ahmad N, Li T, Liu Y, Hoang NQV, Ma X, Zhang X, Liu J, Yao N, Liu X, Li H: Molecular and biochemical rhythms in dihydroflavonol 4-reductase-mediated regulation of leucoanthocyanidin biosynthesis in Carthamus tinctorius L. *Industrial Crops and Products* 2020, 156:112838.
52. Hamberger B, Bak S: Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2013, 368(1612):20120426.
53. Katsumata T, Fukazawa J, Magome H, Jikumaru Y, Kamiya Y, Natsume M, Kawaide H, Yamaguchi S: Involvement of the CYP78A subfamily of cytochrome P450 monooxygenases in protonema growth and gametophore formation in the moss Physcomitrella patens. *Bioscience, biotechnology, and biochemistry* 2011, 75(2):331-336.
54. Wang J, Liu Y, Cai Y, Zhang F, Xia G, Xiang F: Cloning and functional analysis of geraniol 10-hydroxylase, a cytochrome P450 from Swertia mussotii Franch. *Bioscience, biotechnology, and biochemistry* 2010, 74(8):1583-

1590.

55. Pandey RP, Parajuli P, Koffas MA, Sohng JK: Microbial production of natural and non-natural flavonoids: pathway engineering, directed evolution and systems/synthetic biology. *Biotechnology advances* 2016, 34(5):634-662.
56. Zelasko S, Palaria A, Das A: Optimizations to achieve high-level expression of cytochrome P450 proteins using Escherichia coli expression systems. *Protein expression and purification* 2013, 92(1):77-87.

Figures

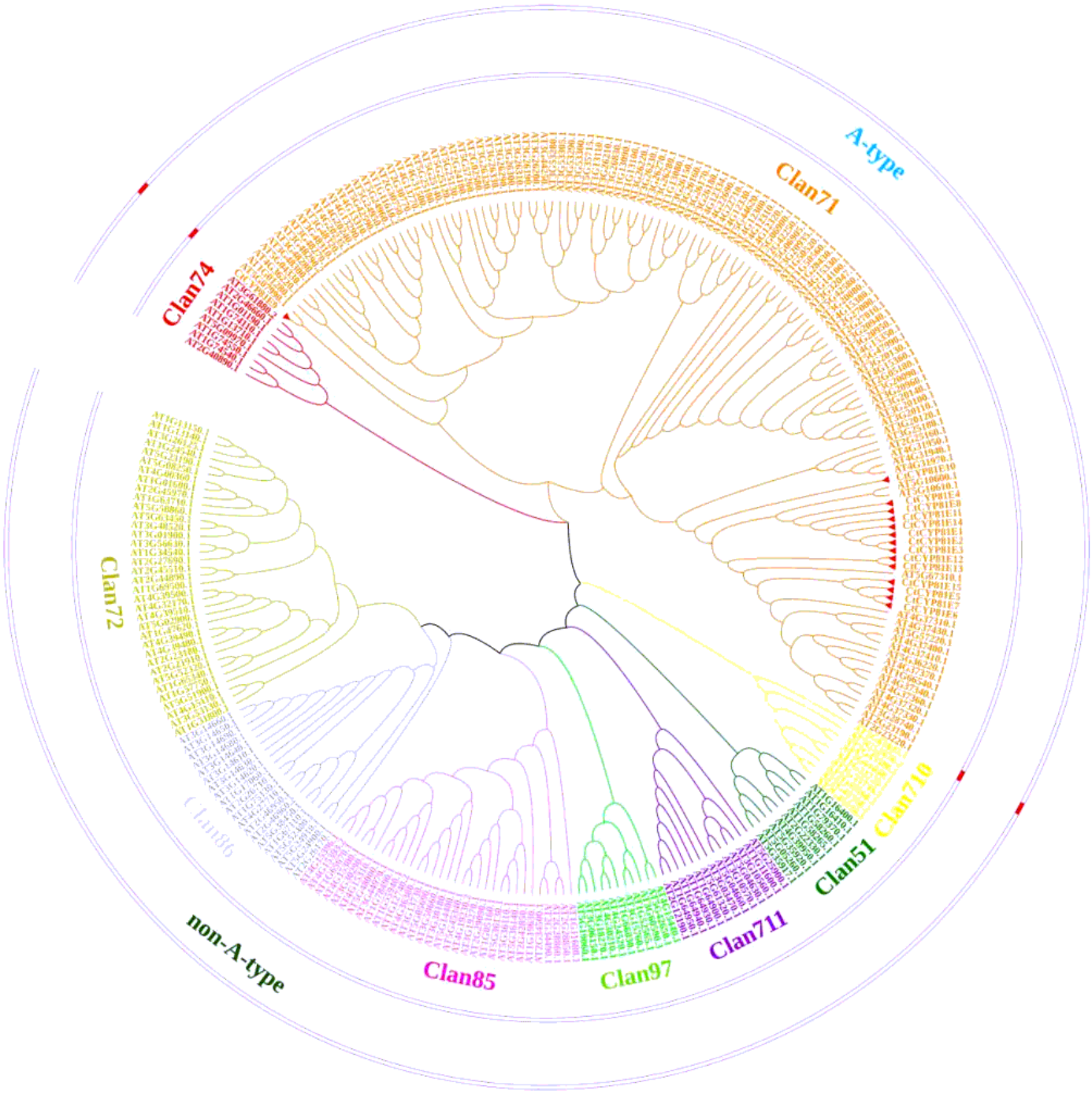

Figure 1

The construction of neighbor-joining (N-J) phylogenetic tree containg 260 CYP450 proteins from safflower and Arabidopsis using MEGA-X software. The pattern of different background colors represents different clans of CYP450 subfamilies. CtCYP81Es are represented by red triangle.

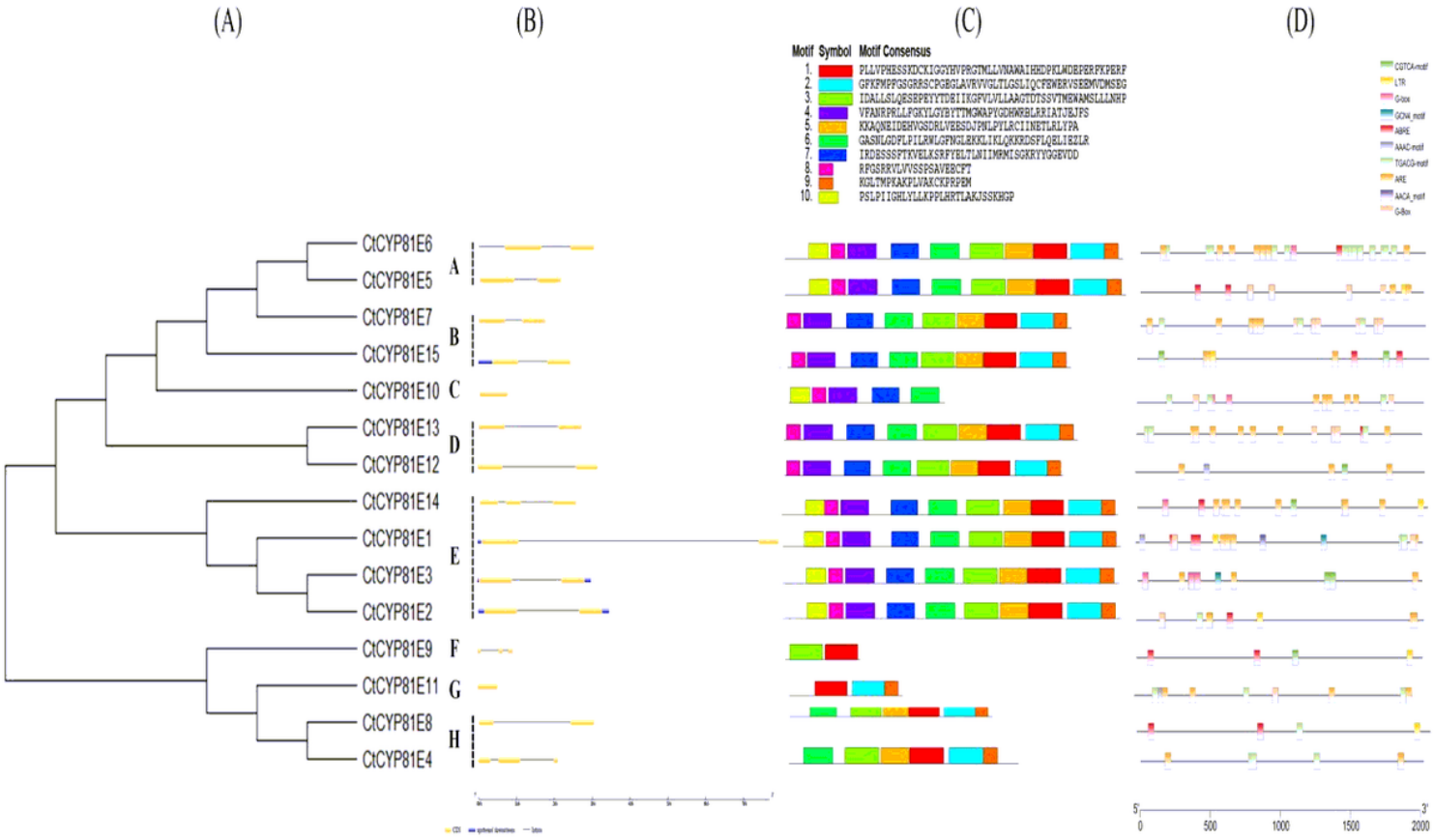

Figure 2

(A) Maximum liklihood (ML) phylogenetic tree was constructed from the selected sequences of safflower CtCYP81E subfamily. (B) The exon/intron organization and complete gene structure of CtCYP81E subfamily in safflower where yellow color boxes represent the presence of exons. The representation of grey lines indicates the number of introns, and untranslated regions (UTRs) were demonstrated with blue color boxes. (C) The distribution of the conserved protein motifs in safflower proteins encoding by CtCYP81E subfamily was predicted with the help of MEME online webtool (D) In silico promoter analyses of 15 CtCYP81E genes where various regulatory elements have been shown with multiple colors at specific position within the promoter regions.

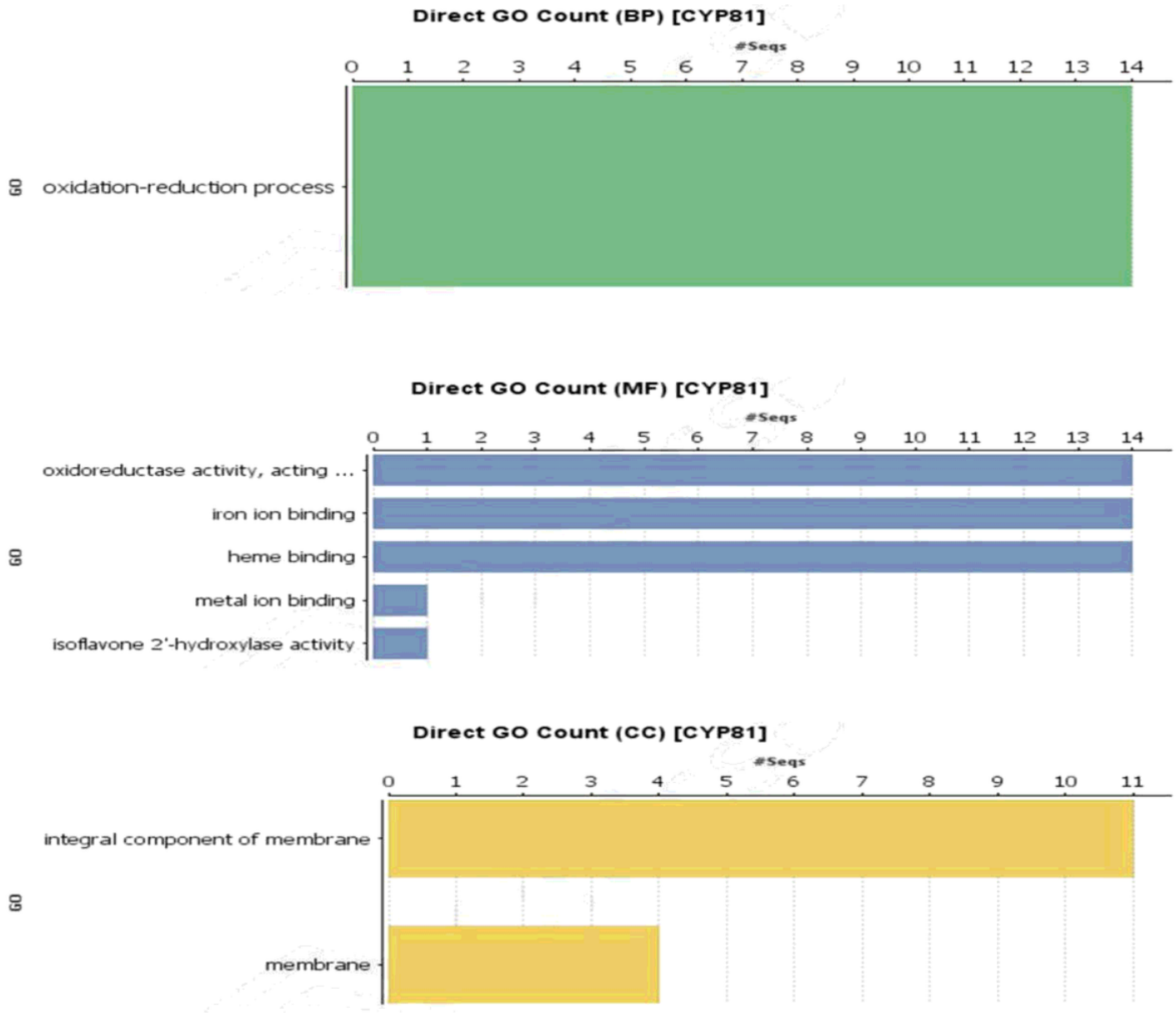

Figure 3

Gene ontology annotation ofCtCYP81E genes in safflower. The GO term annotation of functional categorization for CtCYP81E genes into three classes such as biological processes (BP), molecular function (MF), and cellular components (CC).

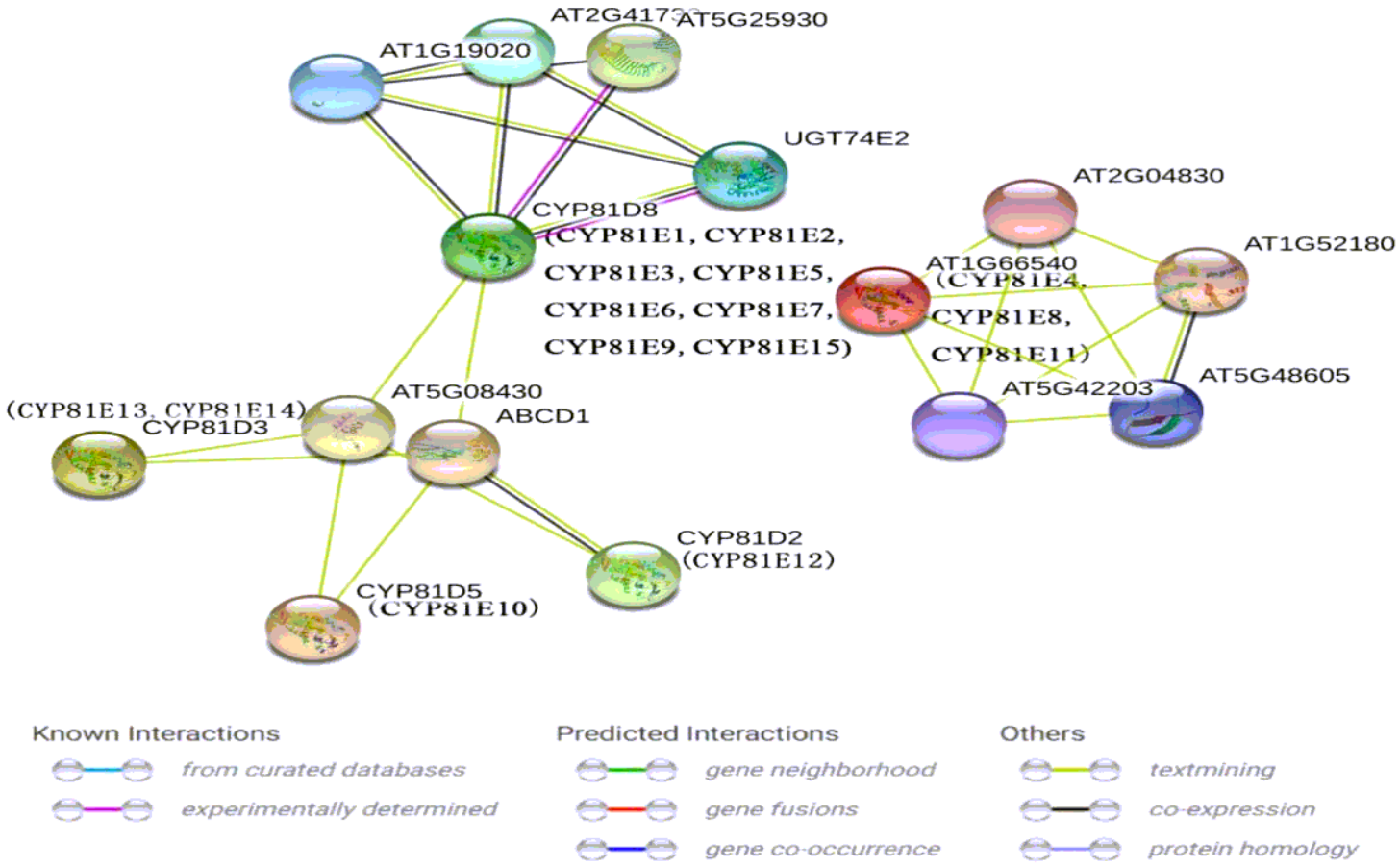

Figure 4

The PPI network of CtCYP81Es subfamily in safflower based on data obtained from CtCYP81E orthologs in Arabidopsis. The online webserver of STRING was exploited to annotate the entire poential network. The indication of red lines showed the groups of proteins coexisted with more than four other P450 proteins members. The CtCYP81E proteins were labbelled in the parentheses right benath the Arabidopsis orthologs.

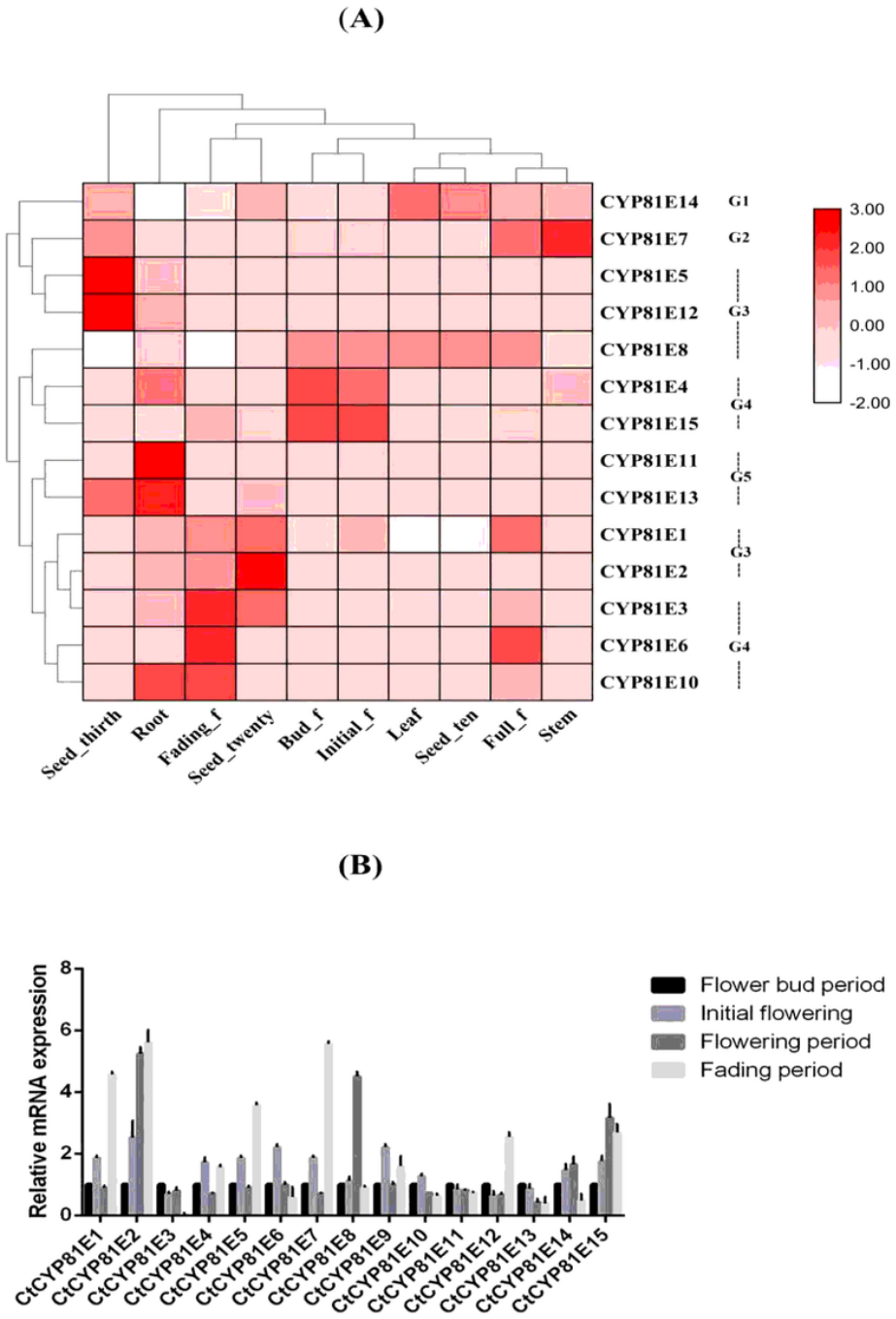

Figure 5

The expression level analysis of the CtCYP81E subfamily genes in safflower. (A) The digital expression analysis of CtCYP81E subfamily genes based on RNA sequencing data. The expression data were illustrated as heatmap created from the RPKM profile of five different tissues of safflower. The expression level of CtCYP81E genes was normalized using Mev software. All selected genes were divided into 5 groups with the help of KMC method. The color scale at the right represents the extent of expression level where green colorrepresents low expression and red color represents high expression (B) The quantitative realtime expression profiling of 15 safflower CYP81E genes in four different flowering stages using qRT-PCR assays. The relative fold expression level was normalized according to 18S rRNA gene (internal reference). The error bars denote the differences found in the expression level of three independent biological replicates.

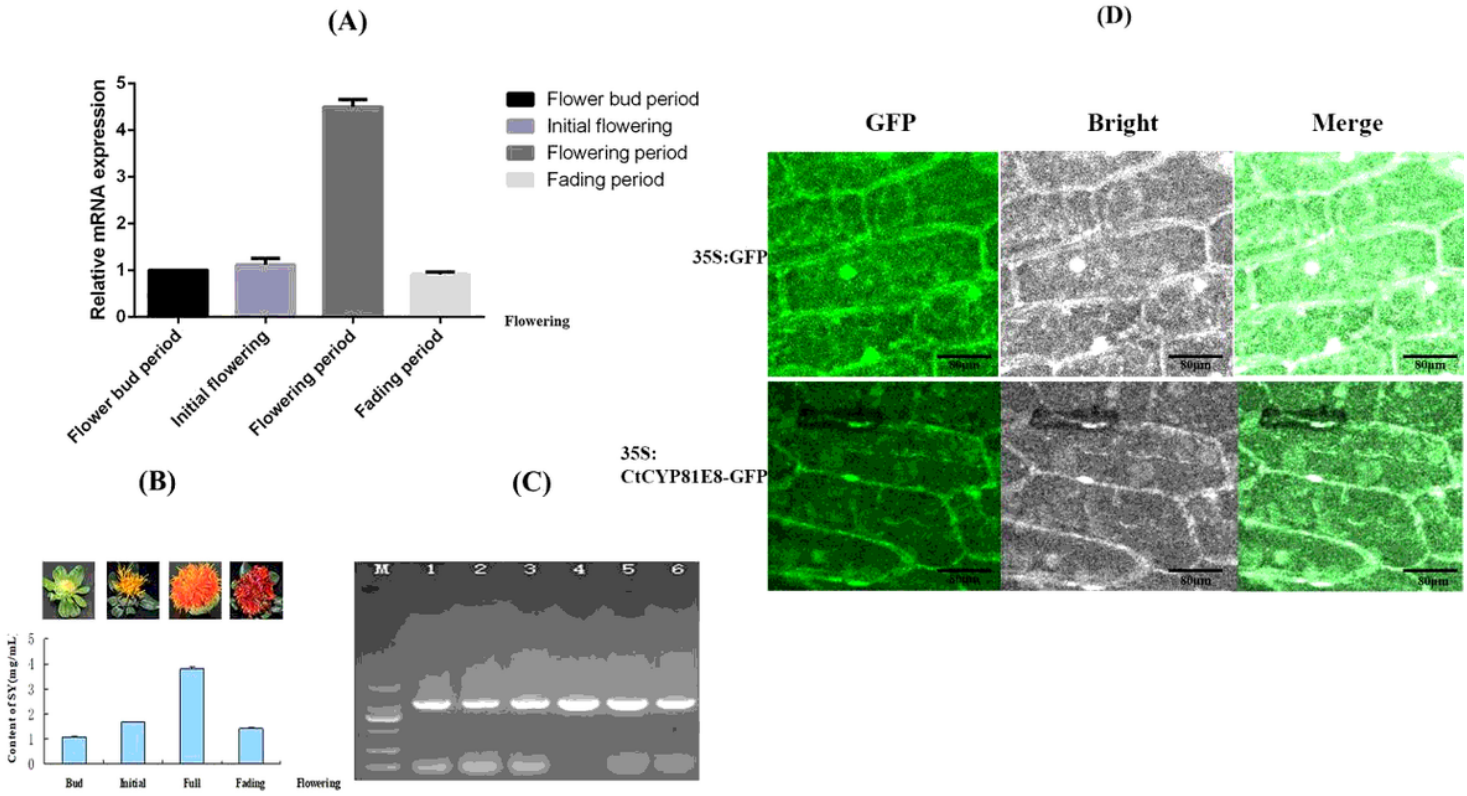

Figure 6

(A) Quantitative expression of CtCYP81E8 gene at different stages of safflower petals (B) metabolite accumulation in different flowering stages of safflower (C) Amplification of the full length CtCYP81E8 products detected on 1.5% gel electrophoresis (D) Subcellular localization of CtCYP81E8 through GFP fluorescence imaging after the transient expression in onion.

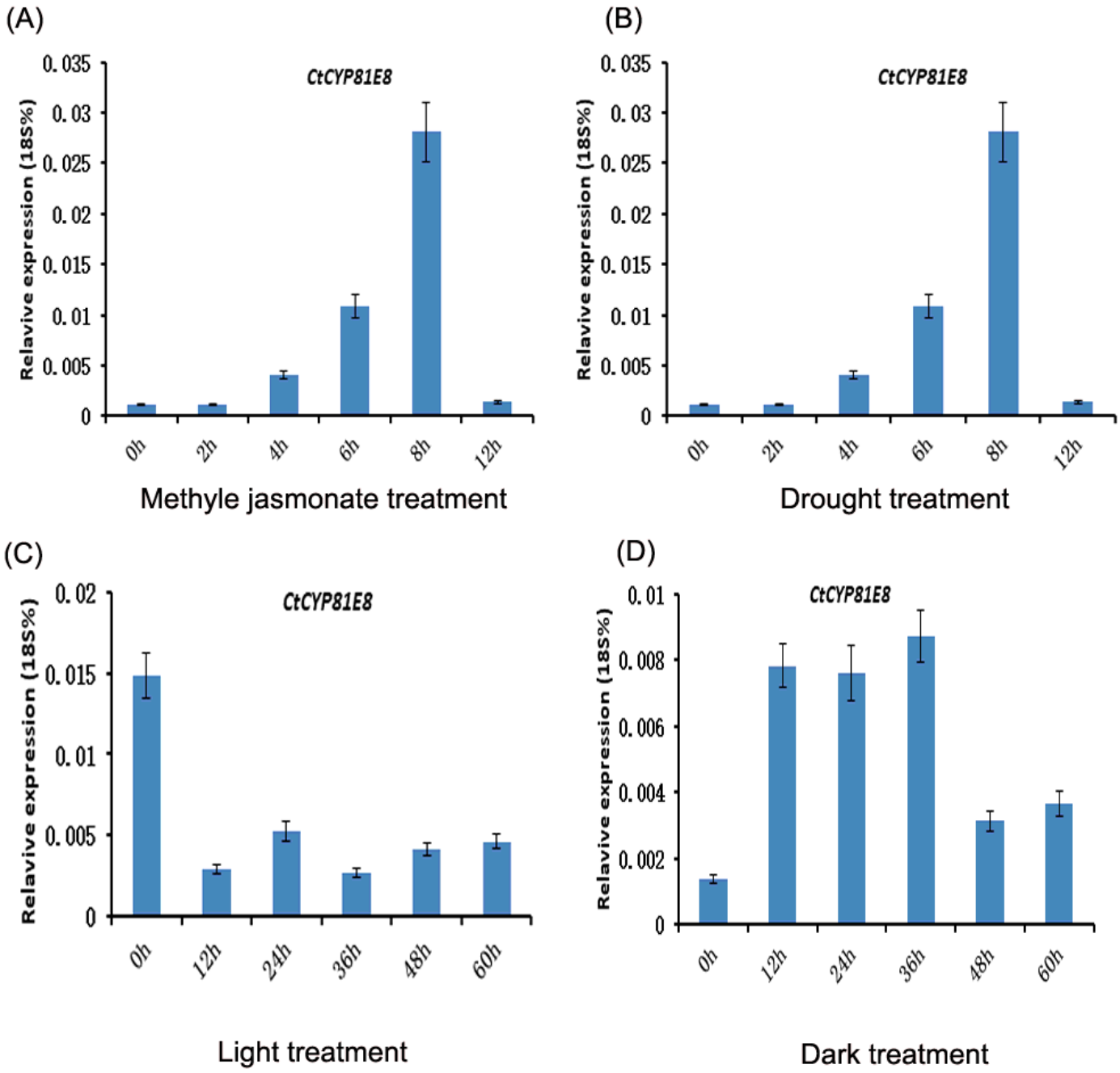

Figure 7

The transcriptional regulation network of CtCYP81E8 gene in the flower tissue of safflower under methyl jasmonate, drought, light and dark stress conditions. The relative fold expression data of six different time points (0, 12, 24, 36, 48 and 60 h) was demonstrated with graphs where the bars indicate mean values of three independant biological samples ± standard deviation (SD).

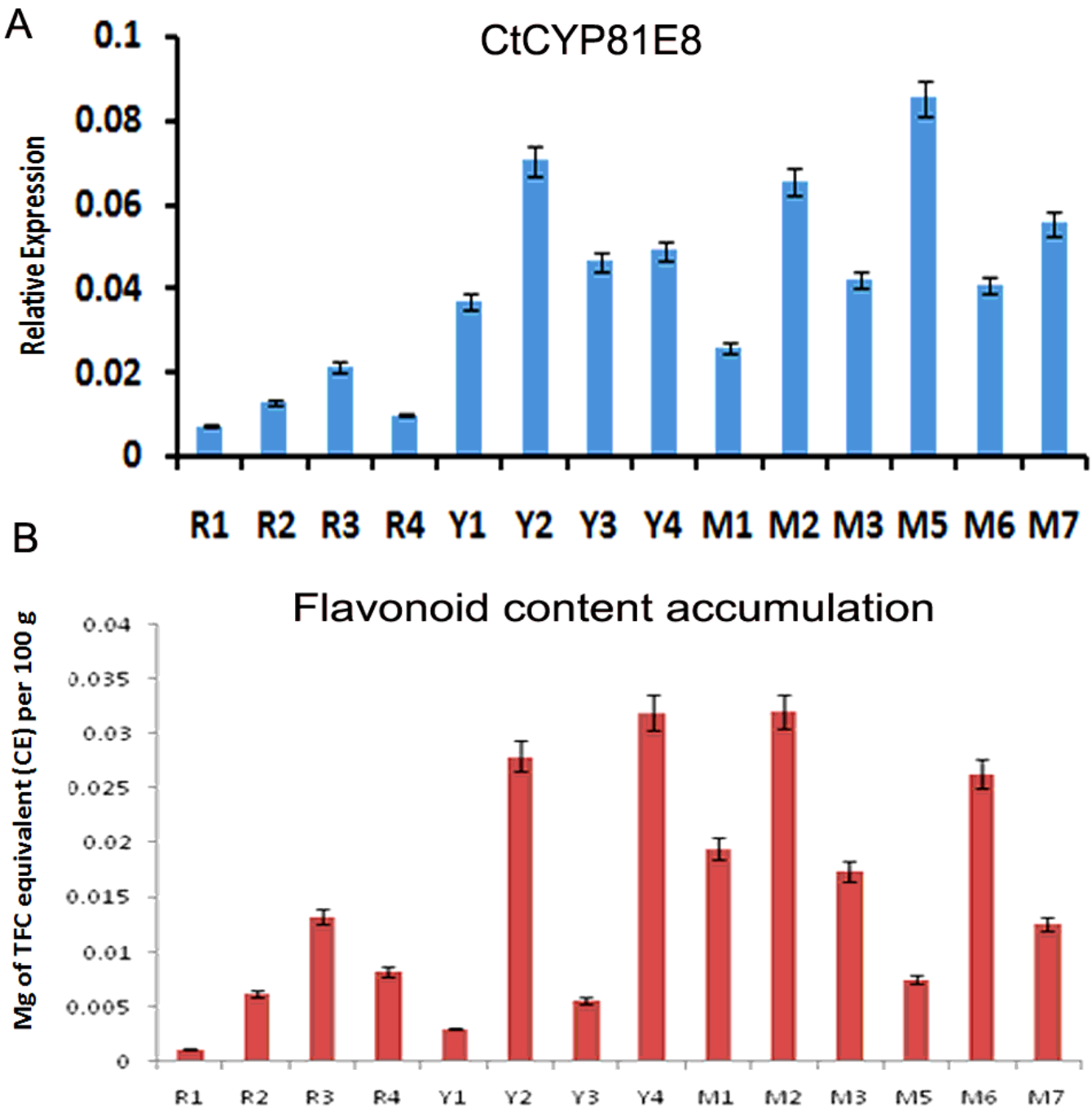

Figure 8

The correlation analysis of CtCYP81E8 transcript abundance and metabolite accumulation profiling in various flower developmental stages of wild and mutatnt safflower varieties using qRT-PCR assay and total metabolite content measurments. . Error bars indicate the SE (n = 3) ± standard deviation (SD). The alpha-numeric codes denote; R: represents wild safflower line red flower, R1: bud flowering, R2: initial blooming flowering, R3: full flowering, R4: Fade flowering. Y: Yellow flower, Y1: bud flowering, Y2: initial bloomimg flowering, Y3: full bloom flowering, Y4: fade flowering. M1-3 represents different stages of the mutant safflower line red flower (bud, initial, full) and M5-7 indicate stages of the mutat safflower line yellow flower (bud, initial, full).

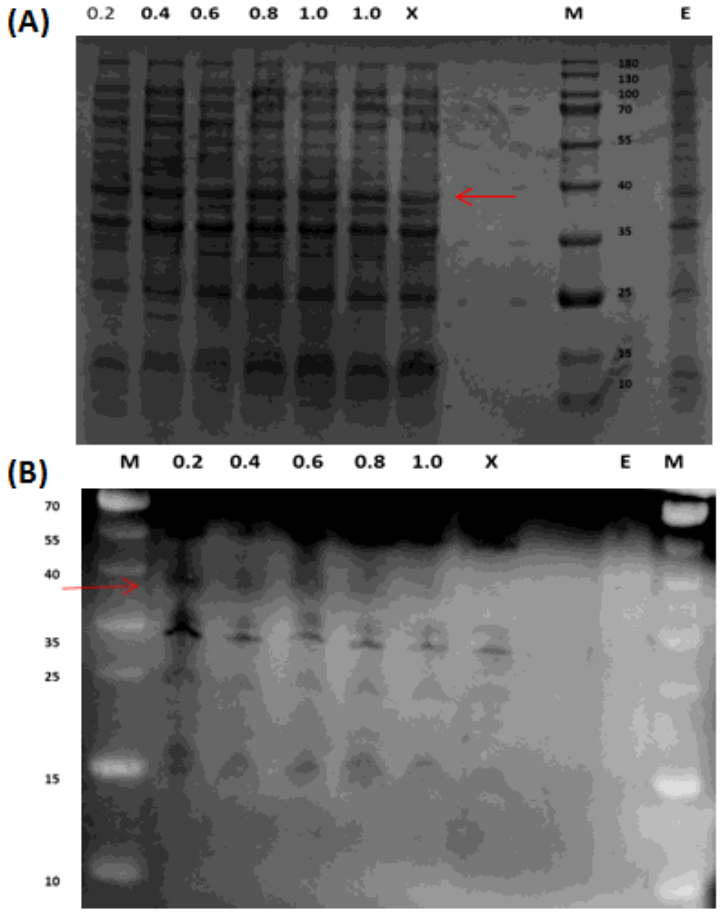

Figure 9

The detection of the recombinant CtCYP81E8 protein usingSDS-PAGE and western blot hybridization of fusion protein purification (A) SDS-PAGE analysis where M1: Protein marker; 1: CtCYP81E8 expressed protein with 0.1mM IPTG induction at 28◦C; 2: CtCYP81E8 expressed protein with 0.4mM IPTG induction at 28◦C; 3: CtCYP81E8 expressed protein with 0.6mM IPTG induction at 28◦C; 4: CtCYP81E8 expressed protein with 0.8mM IPTG induction at 28◦C; 5: Expressed product of pET-28a(+) with-out CtCYP81E8 protein with induction with IPTG at 28◦C. (B) Western blot hybridization for CtCYP81E8 fusion protein purification where M1: Protein marker; 1: Purified sample with induction with 0.8mM IPTG at 28◦C; 2: Expressed product of pET-28a(+) with-out CtCYP81E8 protein with induction with IPTG at 28◦C; 3: Puri-fied sample with induction with 0.6mM IPTG at 28◦C; 4: Puri-fied sample with induction with 0.4mM IPTG at 28◦C.

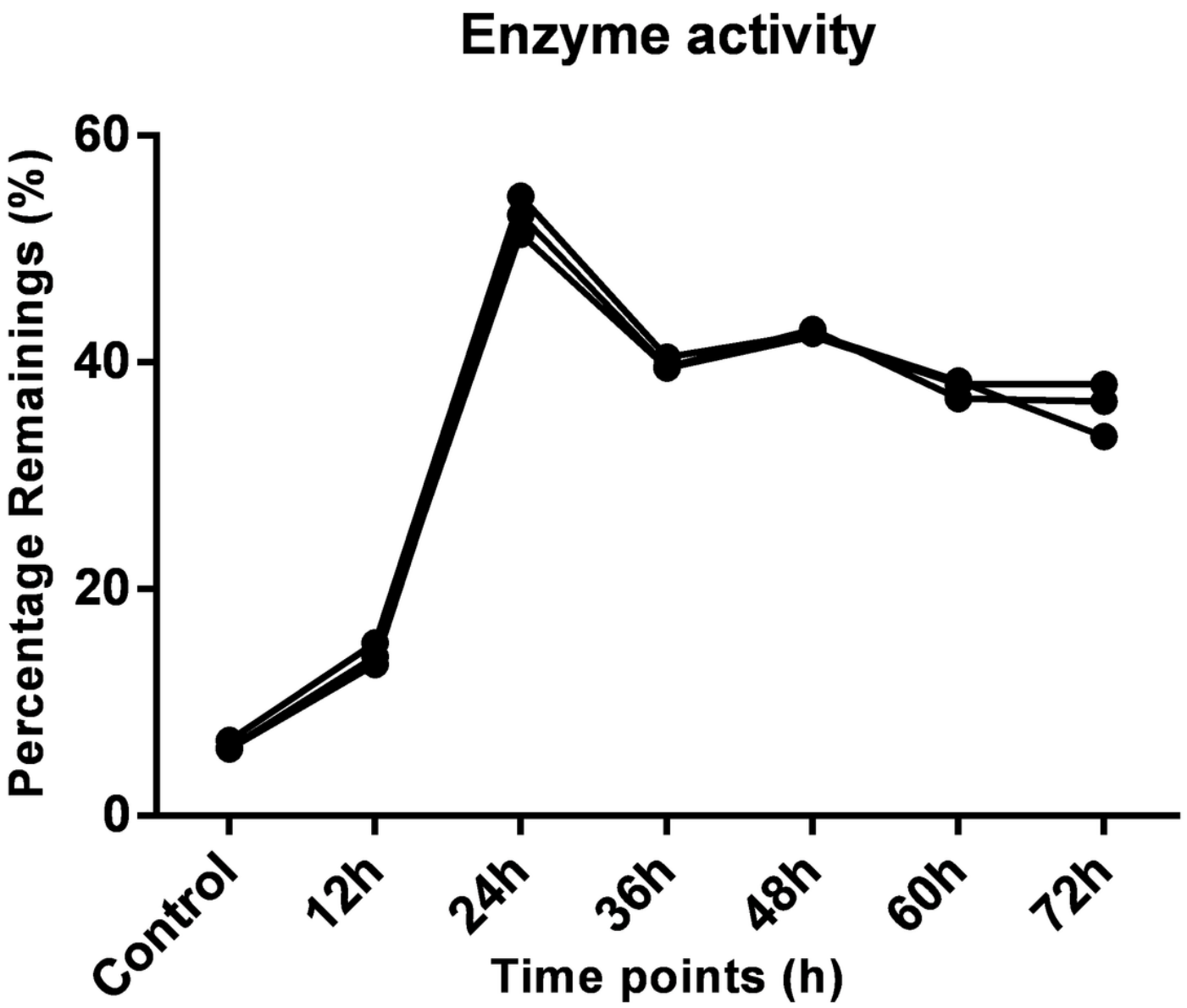

Figure 10

The in vitro activity of the candidate CtCYP81E8 protein using DMP method over multiple time ranges (12h to 72h). Optimum reaction conditions includes (pH=5.6, analytical method-DA).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.rar