

Genome-wide identification and expression profiling of duplicated flavonoid 3'-hydroxylase gene family in *Carthamus tinctorius* L.

Nguyen Q.V. HOANG^{1a}, Kong JIE^{1b}, Naveed AHMAD^{1,3c},
Ma XINTONG¹, Zhang XINYUE¹, Wu YUHANG¹, Wang NAN¹,
Yao NA¹, Liu XIUMING^{1,2,4*}, Jin LIBO^{2,4*}

¹Jilin Agricultural University, College of Life Sciences, Engineering Research Center of the Chinese Ministry of Education for Bioreactor and Pharmaceutical Development, Changchun, 130118, China; viethoangjlau@gmail.com; 245976302@qq.com; naveedjlau@gmail.com; mxt7142020@163.com; 2977201634@qq.com; 2778935775@qq.com; 41459462@qq.com; nayao1103@jlau.edu.cn; xiuming1211@163.com (*corresponding author)

²Wenzhou University, Institute of Life Sciences, Wenzhou, Zhejiang, 325035, China; libo9518@126.com (*corresponding author)

³Institute of crop germplasm resources, Shandong Academy of Agricultural Sciences; Shandong Provincial Key Laboratory of Crop Genetic Improvement, Ecology and Physiology, Jinan 250100, China

⁴Collaborative Innovation Center of Wenzhou, Wenzhou, Zhejiang, China 325035

^{a,b,c}These authors contributed equally to this work.

Abstract

Flavonoid 3'-hydroxylase (F3'H) enzyme is essential in determining the flavonoids B-ring hydroxylation pattern. It is mainly implicated in the biosynthetic pathway of cyaniding-based anthocyanins, flavonols, and flavan-3-ols. However, the evolution and regulatory mechanism of these important flavonoid hydroxylases have not been systematically investigated in safflower (*Carthamus tinctorius* L.). In this study, we identified 22 duplicated *CtF3'H*-encoding genes from safflower through genome-wide prediction and conservation analysis. Phylogenetic analysis revealed the pattern of conservation and divergence of *CtF3'H*s encoding proteins and their homologs from different plant species. The distribution of conserved protein motifs and cis-regulatory units suggested several structural components that could be crucial in deciphering the final function of *CtF3'H* proteins. Furthermore, the results of RNA-seq and qRT-PCR assay in different flowering tissues suggested differential expression level of *CtF3'H* genes during flower development. Based on the unique homology of *CtF3'H5* with flavonoid 3' hydroxylases from other plant species, further validation of *CtF3'H5* was carried out. The transient expression of *CtF3'H5* in onion epidermal cells implied that the subcellular localization of the fusion construct containing *CtF3'H5* and GFP was predominantly detected in the plasma membrane. Similarly, the prokaryotic expression and western blot hybridization of *CtF3'H5* demonstrated the detection of a stable 50.3kD target protein. However, more efforts are needed to further extend the functional validation of *CtF3'H5* in safflower. This study provides a fundamental gateway for future functional studies and understanding the genetic evolution of *F3'H*s in plants.

Keywords: differential expression; flavonoid 3'-hydroxylase (F3'H); prokaryotic expression; safflower; transient expression

Received: 30 Sep 2021. Received in revised form: 07 Nov 2021. Accepted: 11 Nov 2021. Published online: 17 Nov 2021.

From Volume 49, Issue 1, 2021, Notulae Botanicae Horti Agrobotanici Cluj-Napoca journal uses article numbers in place of the traditional method of continuous pagination through the volume. The journal will continue to appear quarterly, as before, with four annual numbers.

Introduction

Safflower (*Carthamus tinctorius* L.) belongs to the Asteraceae family and is widely known as one of the essential self-pollinated species comprising a diploid genome ($2n = 24$). To date, researchers identified seven 'centres of similarity' in safflower based on morphological variability including the Middle East, Far East, Europe, Egypt, Sudan, Ethiopia, and India-Pakistan and) that contain prevalent morphotypes of safflower at each centre (Knowles, 1969). In contrast to other states, China has been employing its flowers as a therapeutic herb for years and has already established a particular growing centre. In addition, safflower contains various active ingredients, including flavonoids, quinolones, alkaloids, and safflower polysaccharides (Ambreen *et al.*, 2018), with biological effects, including antioxidant, anti-inflammatory (Alaiye *et al.*, 2020), and antibacterial activities (Cho *et al.*, 2017). It has been reported that safflower improves acute cerebral infarction and ischemic stroke (Wang *et al.*, 2019). Among flavonoids, Safflower Yellow is the main active ingredient of safflower, and it has a variety of pharmacological effects such as dilating blood vessels and protecting myocardial ischemia. Flavonoids are an important class of natural products (Cerqueira *et al.*, 2021); In particular, they belong to the polyphenolic group of plant metabolites, which are widely found in fruits, vegetables, and some beverages. The biochemical effects and antioxidant abilities of flavonoids are directly related to various diseases such as cancer, atherosclerosis, and Alzheimer's diseases (Du *et al.*, 2021). They also sometimes act as potent inhibitors of different groups of enzymes, such as xanthine oxyase (Lin *et al.*, 2021), cyclo-oxyase, lipoxygenase, and 3-kinase phosphoinositide (Panche *et al.*, 2016). However, the identification of essential genes that regulate the biosynthetic pathway of safflower yellow and anthocyanin in safflower is still underway. Hence, the discovery of new genes will facilitate the underlying principles of the accumulation of plant specialized metabolites such as anthocyanins, luteolins, cyanadins to unravel the mechanism of pigmentation in flower petals, seed coats, or hypocotyls.

The flavonoid 3'-hydroxylase (F3'H) is the core enzyme of the flavonoid metabolic pathway, which belongs to the superfamily of cytochrome P450 (Choudhary *et al.*, 2016; Ren *et al.*, 2021). F3'H hydroxylation of position 3 'of the B ring of flavonoids (Guo *et al.*, 2019) directs the production of cyanide red pigments. Several *F3'H* genes have been cloned and well investigated in various plant species such as hybrid *Petunia* (De Palma *et al.*, 2014) and *Arabidopsis thaliana* (Gao *et al.*, 2020). Furthermore, it was also suggested that the precursor's molecules of flavonols are usually hydroxylated using 3'-hydroxylation via the overexpression of *SbF3'H* in *sorghum*. Similarly, the incredible loss of *Glycine max* fuzz was achieved by viral-mediated silence of the *F3'H* gene (Nagamatsu *et al.*, 2009). The results of the ectopic expression level of the *MdF3'H* identified from *Malus × domestica* stimulated the accumulation pattern of anthocyanin (Han *et al.*, 2010). Recently, it is reported that the late branches of the flavonoid biosynthetic pathway produce flavanols or anthocyanins. The difference in the content of flavonoids found at the pubescence stages was attributed to F3'H enzyme during the onset of pigmentation (Iwashina *et al.*, 2006). The roles of *F3'H* and *F3'5'H* have been shown to regulate the formation of the two important classes of flavonoids i.e., dihydroxylated and trihydroxylated. The transcription of these genes has been demonstrated to have a direct effect on the deposition of cyanidin/delphinidin during anthocyanin accumulation in grapevine, which causes colour differences (Castellari *et al.*, 2006). However, to date, no study has been established on the genome-wide identification and systematic characterization of *F3'H* genes in safflower.

The integration of metabolomics and transcriptomics studies facilitates the discovery of gene-metabolite crossroads, resulting in the identification of essential genes underlying their regulatory pathways. Recently, a combination of HPLC and RNA-seq approaches, several catechin biosynthetic genes tightly linked with their respective catechins in *C. sinensis* were recently discovered (Wu *et al.*, 2014). Another study revealed the systematic identification of *F3'H* and *F3'5'H* genes using the RNA-seq analysis in shrubs or small trees (Wei *et al.*, 2015). We also previously identified different classes of transcription factors involved in the downstream regulatory pathway of flavonoid biosynthesis (Hong *et al.*, 2019; Li *et al.*, 2020). Therefore, it is essential to fully understand and investigate the identification, evolution and regulation of *flavonoid 3'-hydroxylase* genes,

which are of great significance for the downstream regulation of flavonoid in safflower. In this study, we present the genome-wide identification and the evolutionary relationship of *CtF3'H*-encoding genes in safflower. We uncovered several regulatory networks underlying the control mechanism of *CtF3'H* using comprehensive bioinformatic analyses including physico-chemical characterization, phylogenetic classification, conserved protein motifs, and promoter analysis. The differential expression of *CtF3'H* in different flowering stages of safflower was also investigated from RNA-seq data and qRT-PCR assay. We also confirmed the experimental validation of a candidate *CtF3'H5* gene using routine molecular biology techniques such as molecular cloning, subcellular localization, prokaryotic expression, and western blot hybridization. This study not only consolidated the preliminary work in understanding the regulation scheme of the 3' B-ring containing flavonoids but also presented a complete model for future studies on F3'H-encoding genes in plants.

Materials and Methods

Plant materials, vectors, and strains

The safflower Jihong early-maturing variety was used as the experimental source in this study. The seeds were purchased from Xinjiang Honghuayuan Technology Co., Ltd, China and planted in the experimental station of the Jilin Agricultural University. The petals of safflower bud, initial, flowering, and fade stages were used for expression analysis. *Agrobacterium tumefaciens* strain EHA105, *E. coli* BL21, *E. coli* DH5 α cells, the prokaryotic expression vector (PET28a+*CtF3'H5*), subcellular localization vector pCAMBIA1302-*CtF3'H5*-GFP) were constructed and stored in 75% glycerol at -80 °C refrigerator until further use.

Identification and characterization of CtF3'H in safflower

The database of Hidden Markov Model (HMMsearch) was utilized to screen all putative *CtF3'H* genes in the safflower genome by providing the family domain identifiers PF00067 and PR00385 in the Pfam database available online at (<http://pfam.xfam.org/>). Further classification was performed in the BioEdit software to BLAST the putative *CtF3'H* genes in safflower genome. The genomic and protein sequences of *CtF3'Hs* were retrieved for further analysis. In addition, we also investigated the presence of two highly conserved regions including “GGEK” and “LPPGP” which are specific to F3'H proteins with the help of the online webserver of MARCOIL accessible available at (<http://toolkit.tuebingen.mpg.de/marcoil>). The amino acid sequence of each protein was aligned using DNAMAN software (Version 7; Lynnon Corporation, Quebec, Canada), with default parameters. The non-redundant *CtF3'H* protein sequences lacking the conserved entities specified to this family was removed and the redundant *CtF3'H* accessions were assembled for further analysis. Various physico-chemical properties of the selected *CtF3'H* including protein size, molecular weight (MW) and isoelectric point (pI), and GRAVY were investigated using the online tools of ExPASy software available at (<http://www.baoasy.org/>). The prediction of theoretical subcellular localization was carried out with cello web server (<http://cello.life.nctu.edu.tw/>) and WoLF PSORT (<https://wolfsort.hgc.jp/>).

Phylogenetic analysis

The phylogenetic reconstruction of the candidate *CtF3'H* proteins was performed following multiple sequence alignments by Clustal W (2.0) software. A total of 22 *CtF3'H* from safflower and one each F3'H protein from *Vitis vinifera*, *Allium cepa*, *Arabidopsis thaliana*, *Petunia* \times *hybrida*, *Sorghum bicolor*, *Glycine max*, *Ipomoea purpurea*, and *Oryza sativa* in combination with six *CtF3'H* proteins from *Salvia miltiorrhiza* were aligned for phylogenetic tree construction (Supplementary File 1). The phylogenetic relationship of *CtF3'H* proteins with other plants F3'H protein was demonstrated with a neighbour-joining phylogeny tree (1000 bootstrap method) with the help of MEGA 5 software version 4.1 (<http://www.megasoftware.net/>)

(Schrago *et al.*, 2018). The relationship and evolutionary pattern were then analysed following the assembly of different members of F3'H proteins from different plant species.

Motif elicitation and promoter analysis

The conserved protein motifs of the CtF3'H proteins were determined by uploading 22 amino acid sequences into the MEME web server's online tool (<http://meme.nbcr.net/meme/cgi-bin/version/meme.cgi>). The parameters are optimized as follows: zero or one, the rate of occurrence of a pattern on each sequence; 10 BP, module width range; and three other broad ranges of patterns were identified, and finally, ten motifs were selected. All different parameters follow default values. In the same way, the distribution of the all conserved cis-elements of the promoter sequences of safflower *CtF3'H* genes was extensively investigated by exploiting the PlantCARE software available at (<https://sogo.dna.affrc.go.jp/>).

Differential expression analysis

The differential expression level of the identified *CtF3'H* genes were initially investigated from RNA-seq data obtained from different tissues/organs of safflower including root, seed stem, and four different flowering stages (bud, initial, full and fade). A combined heatmap was generated from the kilobase exon model per million mapped read (RPKM) statistics calculated from the expression level of each tissue/organ/stage. Furthermore, the expression level of candidate *CtF3'H* transcripts in four different flowering stages was validated by qRT-PCR analysis. For this purpose, the primer pairs for each *CtF3'H* gene were synthesized. The total RNA extraction was carried out from each flower tissue with the help of TRIzol reagent. The synthesis of cDNA templates was produced using the reverse transcriptase enzyme. All qRT-PCR reactions were carried out with SYBR® Premix Ex Taq™ (TliRNaseH Plus) from TaKaRa Biotechnology Co., Ltd. (Dalian, China) using a real-time PCR device (Biosystems 7500). Following the manufacturer's protocol, a 20 µL of PCR mixture was prepared, including 1.0 µL cDNA, 0.4 µM each primer (F/R), 0.4 ROX dye, 10 µl main mixture, and RNase free water 7.8 µL. The internal reference gene 18s ribosomal RNA of safflower was used as a control. The relative expression level for each transcript was measured according to the $2^{-\Delta\Delta Ct}$ method (Min *et al.*, 2020). The primer details are listed in (Supplementary Table S1).

Gene cloning and subcellular localization of CtF3'H5

The total RNA content was extracted from the preserved flower petals of the safflower with TRIzol reagent. RNA concentrations were determined by NanoDrop 2000 based on OD260/280 values and electrophoresis on 1% agarose gel respectively. Following the manufacturer's instructions, reverse transcription for synthesizing the first-strand cDNA templates for PCR amplification was carried out. The gene-specific primers (Supplementary Table S2) were designed according to the genomic sequence of *CtF3'H5* for cloning. The PCR reaction conditions were as follow 98 °C for 10 s, 98 °C for 10 s, 65 °C for 15 s, 72 °C for 2 min, 35 cycles; 72 °C for 10 min. The amplified PCR product of the candidate *CtF3'H5* gene was detected on 1% agarose gel electrophoresis. The resultant bands were recovered from the gel using a gel recovery kit and then ligated into the cloning vector (pEASY-T1). The recombinant vector was then transformed to the competent cells of bacteria using heat and shock method. The positive colonies of pEASY-T1- *CtF3'H5* were selected on LB agar plates containing kanamycin, and the results were confirmed with a half colony PCR method using gene-specific primers. After PCR confirmation, the plasmids were extracted from the positive colonies, confirmed with double restriction enzyme digestion, and then sent to Shanghai Biotech Engineering Services Co., Ltd for sequencing. Similarly, using the *CtF3'H5* T vector plasmid, the full-length cDNA sequence of *CtF3'H5* was amplified using new primer pairs (Supplementary Table S2) containing *BglII* and *NcoI* dual restriction sites. The target band was subcloned into a new plant expression vector pCAMBIA1302-GFP-35S for the investigation of subcellular localization. The recombinant plasmid was confirmed initially with double restriction digestion using *BglII* and *NcoI* enzymes and then verified by sequencing. The recombinant plasmid (pCAMBIA1302-GFP-*CtF3'H5*) was then transformed into *Agrobacterium EHA105* competent cells using

the heat and shock method, and positive bacterial strains were selected on YEP agar medium. The transient transformation system was established using the *Agrobacterium*-mediated infection of the onion epidermal cell under controlled conditions. The *Agrobacterium*-infected onion epidermal cells containing the recombinant plasmid (pCAMBIA1302-GFP-*CtF3'H5*) were plated on the solid medium of 1/2 MS, placed in a dark culture medium at 28 °C for 18 h. Then the results of the GFP expression were analysed by scanning confocal laser microscopy.

Prokaryotic expression and Western Blot analysis

The full-length cDNA sequence of *CtF3'H5* was amplified using Pfu DNA polymerase (Takara, Beijing, China) with the help of a new pair of primers (Table S2) with an added *HindIII* and *XhoI* restriction sites. The amplified *CtF3'H5* gene was retrieved from the gel, and ligated into the prokaryotic expression vector pET-28a+ using T4 ligase enzyme with a 4:1 ratio. Then ligation was performed into the same digested empty pET28a+ vector using T4 DNA ligase enzyme overnight at 16 °C for 16 h. The ligated product was transferred to the DH5 α strain, and recombinant strains were selected on a kanamycin-resistant LB-solid medium. PCR confirmed the positive clones following by validation with double restriction digestion system and sequencing. The successfully constructed expression plasmid pET-28a+*-CtF3'H5* was transformed into the competent BL21 cell. The positive clones were identified and cultured at a speed of 180 rpm/min at 37 °C and 28 °C respectively in 10ml of LB liquid medium containing 50 μ g/mL of ampicillin. When the absorbance at 600nm (A600) reaches 0.8, the induction with IPTG was carried out using different concentrations (0.2,0.4,0.5,0.6,0.7,0.8) at 2 h, 4 h, 5 h, 6 h, and 8 h, respectively. The soluble protein extract of *CtF3'H5* was detected on 12% SDS-PAGE, and then the target bands were stained with coomassie brilliant blue dye. The proteins obtained were further subjected to Western blot hybridization using a PVDF membrane for 2 hours. The membrane was washed for 5 min using Tris buffer saline Tween (TBST) and then blocked with the blocking buffer (TBST buffer containing 5% skim milk powder) for 1 h. After performing the incubation with the primary antibody at room temperature, we discarded the blocking buffer. The incubation was allowed at 4 °C overnight followed by thorough washing with TBST 5 times for 5 min each time. The second antibody onto the membrane was added for 2 h incubation following by washing five times with TBST for 5 min each time. The antibody binding complex was analysed with the help of ECL chemiluminescence, and this experiment was repeated three times.

Results

Identification and physicochemical properties of the safflower CtF3'H

To identify all putative *CtF3'H* genes in safflower, we performed extensive analysis using Pfam ID scan in the hidden Markov model (HMMsearch) was carried out to identify all genuine *CtF3'HS* genes in safflower. A total of 37 candidate *CtF3'H* genes were identified in the safflower genome. Of which, 15 *CtF3'H* sequences were determined to be redundant due to their incomplete information, and therefore excluded from the analysis. The remaining 22 *CtF3'H* sequences were selected as non-redundant. The different physico-chemical properties of *CtF3'H* proteins were further investigated with the help of various in silico analysis including the protein length ranging from 128aa to 518aa, the molecular weight varies from 14.84 kDa (*CtF3'H22*) to 58.6 kDa (*CtF3'H23*), with an average is 40.75 kDa. The isoelectric points measurements ranged from 4.94 (*CtF3'H7*) to 9.49 (*CtF3'H37*), with an average of 7.14. Similarly, the grand average of hydropathicity (GRAVY) analysis showed that most of the *CtF3'H* proteins showed an average value of hydropathicity between -0.008 (*CtF3'H26*) and 0.088 (*CtF3'H37*), suggesting the hydrophilic and hydrophobic nature of these proteins (Table 1).

Table 1. Physicochemical characteristics of CtF3'H-encoding proteins in safflower

Name	Gene ID	Size	PI	MW (kDa)	Subcellular Localization	Instability Index	Gravy
CtF3'H1	CCG001824.1	79	7.63	20.19	Endoplasmic reticulum	33.20	-0.116
CtF3'H2	CCG001825.2	179	6.74	20.25	Endoplasmic reticulum	32.78	-0.133
CtF3'H5	CCG001872.1	455	6.20	50.03	Plasma membrane	35.29	-0.035
CtF3'H7	CCG002178.1	249	4.94	28.59	Plasma membrane	36.21	-0.167
CtF3'H8	CCG002598.1	258	6.54	29.63	Plasma membrane	38.59	-0.335
CtF3'H9	CCG004791.1	308	9.35	35.98	Plasma membrane	39.38	-0.396
CtF3'H10	CCG004792.1	195	7.73	22.70	Endoplasmic reticulum	37.62	-0.385
CtF3'H13	CCG006736.1	507	6.77	57.39	Plasma membrane	37.02	-0.107
CtF3'H14	CCG006903.1	507	7.26	58.26	Plasma membrane	38.97	-0.267
CtF3'H15	CCG006904.1	507	7.30	58.51	Plasma membrane	37.87	-0.266
CtF3'H20	CCG017086.1	496	7.55	56.48	Plasma membrane	35.67	-0.142
CtF3'H22	CCG019544.1	128	6.16	14.84	Endoplasmic reticulum	31.53	-0.183
CtF3'H23	CCG020436.1	518	8.76	58.61	Plasma membrane	43.66	-0.207
CtF3'H24	CCG022595.1	160	6.90	18.86	Endoplasmic reticulum	42.80	-0.179
CtF3'H26	CCG023726.1	745	6.89	82.58	Endoplasmic reticulum	34.45	-0.008
CtF3'H27	CCG024108.1	311	8.23	36.27	Plasma membrane	34.14	-0.532
CtF3'H28	CCG024245.1	516	7.25	58.80	Plasma membrane	47.76	-0.214
CtF3'H31	CCG026487.1	252	5.81	28.87	Plasma membrane	40.31	-0.302
CtF3'H32	CCG027037.1	505	8.19	57.42	Plasma membrane	39.51	-0.146
CtF3'H33	CCG028222.1	449	5.82	49.25	Endoplasmic reticulum	36.29	-0.015
CtF3'H34	CCG031317.1	207	5.57	23.54	Endoplasmic reticulum	44.48	-0.194
CtF3'H37	CCG031319.1	265	9.49	29.46	Plasma membrane	20.55	0.088

Phylogenetic analysis of CtF3'Hs

To examine the evolutionary history of CtF3'Hs gene family in safflower, a neighbour-joining phylogeny tree was constructed containing 22 candidates CtF3'H protein sequences and one F3'H sequence from *V. vinifera*, *A. cepa*, *A. thaliana*, *P. hybrida*, *S. bicolor*, *G. max*, *I. purpurea*, *O. sativa* and six homologs

from *S. miltiorrhiza*. The *Arabidopsis* F3'H was selected as the out-group. Importantly, the phylogenetic relationship exhibited remarkable topological consensus, indicating that the derived phylogeny was rather accurate. As described in Figure 1, the candidate CtF3'Hs were clustered together into different groups, corresponding to other F3'H from different plant species. A closer examination of the constructed phylogeny suggested that most of the CtF3'H protein shared an identical pattern of evolutionary pattern with *Salvia miltiorrhiza*. Particularly, CtF3'H9, CtF3'H10, CtF3'H14, CtF3'H15, and CtF3'H23 showed the first divergence followed by CtF3'H1, CtF3'H2, CtF3'H7, CtF3'H8, CtF3'H20, CtF3'H27, CtF3'H31 and CtF3'H32. Interestingly, CtF3'H34 indicated the divergence before *S. miltiorrhiza*. On the other hand, the intermediate divergence between CtF3'H4, CtF3'H13, CtF3'H28, CtF3'H34 and CtF3'H37 was found between safflower and *S. miltiorrhiza*. In case of other plants, *A. cepa*, *S. bicolor* and *O. sativa* diverged before *V. vinifera*, *A. thaliana*, *P. hybrida*, *G. max*, and *I. purpurea*. Importantly, the candidate members of CtF3'H5, CtF3'H26, CtF3'H33 from safflower were clustered together with *V. vinifera*, *P. hybrida* and *I. purpurea*.

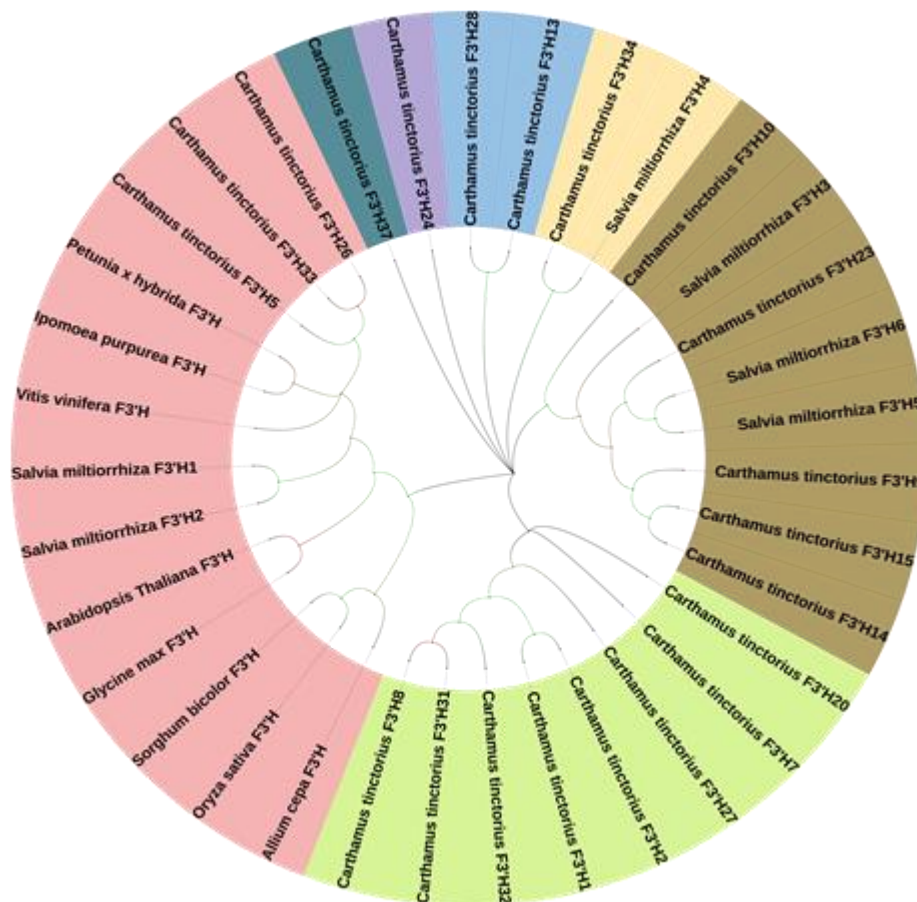


Figure 1. Phylogenetic analysis of CtF3'Hs family in safflower. The evolutionary relationship was inferred with cluster analysis using neighbour-joining method. The percentage of replicate trees in which the identical taxa clustered together in 1000 bootstrap method was shown next to the branches. The tree is drawn to scale with branch lengths in the same units as those of the evolutionary distances used to infer the phylogeny tree

The distribution of conserved protein motifs and alignment of CtF3'H proteins

The organization and distribution of the conserved protein motifs of CtF3'H protein were extensively investigated and screened with MEME online tool version 5.1.1. The results of the conserved motifs of CtF3'H proteins were combined into a phylogenetic tree. We found two highly conserved protein motifs expressed as red and blue masses, correspondingly (Figure 2A). The location and distribution pattern revealed that most of

the conserved motifs were localized to the C-terminus of CtF3'H proteins. The two members of CtF3'H family namely CtF3'H9 and CtF3'H37 could not detect the conservation of the protein motifs. Similarly, the composition of motifs conservation in CtF3'H27 described an incomplete pattern. Notably, in most cases, the putative proteins of safflower CtF3'H demonstrated the two conserved regions, which are usually linked next to each other; but, in CtF3'H26, these two conserved regions were found separated from each other. In addition, we also performed multiple sequence alignment of all putative 22 CtF3'H proteins to determine their sequence homology (Supplementary Figure 1). The highly conserved homology within the same family members was shown with red and blue colours. The similarity of amino acid conservation confirms the previous results of MEME scores obtained from CtF3'H proteins. The corresponding locations of each CtF3'H protein pattern were found conserved. However, motif 1 and motif 2 were found scattered among all CtF3'H proteins. Importantly, the occurrence of motif 3 is present in all protein sequences except CtF3'H1 and CtF3'H2. Similarly, motif 4 is not present in CtF3'H7 and part of CtF3'H8, whereas motif 5 was found in some proteins such as CtF3'H1, 3, 4, 5, and 10. Motif 6 is located in CtF3'H1,3,4,6, while motif 8 was only screened in CtF3'H8, 9, and 10. Importantly, the detection and analysis of protein motifs network revealed significant evolutionary and conservational features of the CtF3'Hs in safflower.

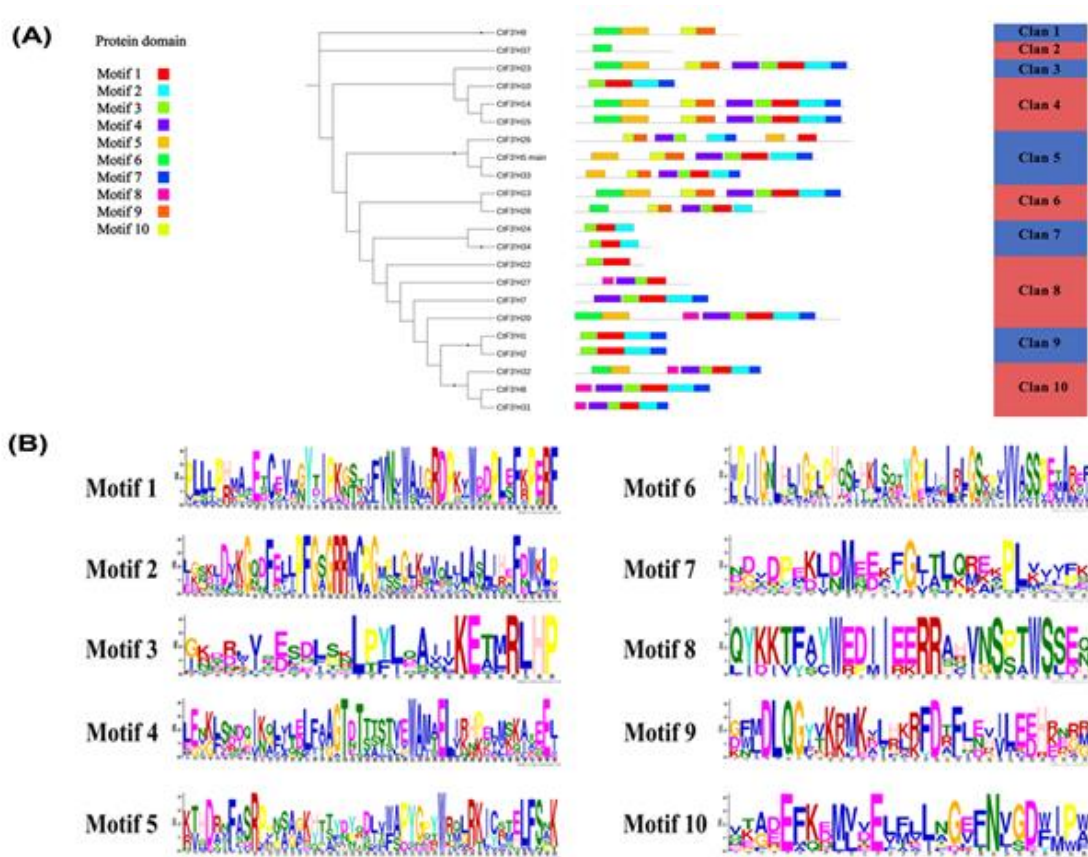


Figure 2. The analysis of the conserved protein motifs of CtF3'H proteins identified from Safflower. (A) The illustration and positioning of most common conserved motifs combined into a phylogenetic tree. The different colour represents the type of motifs whereas the grey line indicates the length of each protein. The red and blue masses represent the highly conserved motifs in CtF3'H proteins (B) The complete logos of the conserved protein motifs identified in CtF3'H proteins

Organization of cis-regulatory units of CtF3'H promoters

To identify different gene promoter cis-regulatory elements, we extensively analysed the 2kb upstream to the ATG starting site of CtF3'H genes using PlantCARE's software. The abundantly annotated cis-elements

within the promoter sequence of most of the *CtF3'H* genes contained factors related to light reactions. For example, a total of 19/22 genes *CtF3'H* include the G-Box cis-element, suggested that signals induction might play crucial roles during the transcription activation and regulation of *CtF3'H* genes in safflower (Figure 3). Moreover, our analysis revealed MYB and MYC responsive elements, which are actively implicated during flavonoid biosynthesis. In addition, the occurrence of cis-regulatory units conserved in other significant plant reactions has also been studied within the promoters of different *CtF3'H* genes, such as drought (MBS) and abscisic acid reaction (ABRE). Convincingly, the presence of such types of cis-acting elements in safflower *CtF3'H* genes suggested positive hallmarks during secondary plant metabolism and various stress-related and hormonal induced pathways.

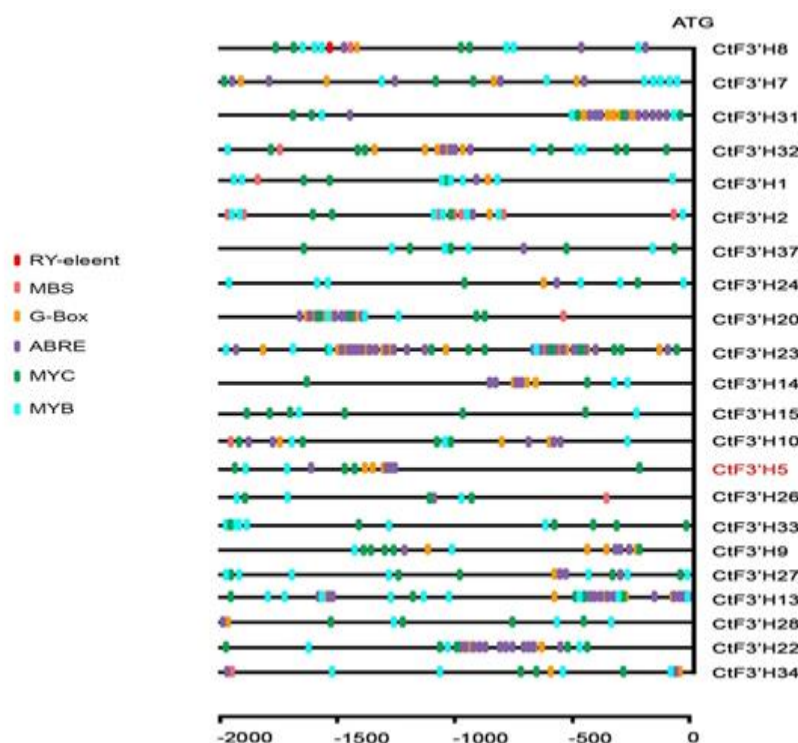


Figure 3. The graphical illustrations of key cis-regulatory elements detected in the promoter region of *CtF3'H* genes. Various groups of cis-elements were demonstrated with different colour boxes

Expression analysis of CtF3'H transcripts

The differential expression of *CtF3'H* genes identified from safflower was primarily computed with RPKM values obtained from RNA-seq data available online under the accession number PRJNA399628 in NCBI. Results implicated that all candidate *CtF3'H* transcripts were differentially expressed in different developmental stages of safflower including four different flowering stages (Figure 4A). The expression level of only a few *CtF3'H* transcripts such as (*CtF3'H5*, 20, 32) demonstrated strong expression in the shoot tissues. However, some *CtF3'H* transcripts, including *CtF3'H2*, 1, 2, 13, were abundantly expressed in the root and seeds tissue/organs (Figure 4B). On the contrary, the expression level of *CtF3'H33*, 20, 37, 28, 34 showed the highest expression in the intermediate seed maturation stage, besides (*CtF3'H23*, 24, 33) whereas the abundance of *CtF3'H22* transcript appears only in the seed tissues. To confirm the integrity of *CtF3'H* expression, we conducted the qRT-PCR analysis of 22 putative *CtF3'H* genes in four different flowering stages (Figure 4C). A diverse pattern of expression profiles for these *CtF3'H* genes was detected in different flowering stages of safflower. For instance, the relative expression level of *CtF3'H5* showed the highest expression at the full flowering stage followed by the initial stage. However, the expression of *CtF3'H5* was uniformly detected during the bud and fade stages (Figure 4D). These findings suggested insights into the correlation of *CtF3'H5*

with red and yellow colour transition during the full and initial flowering in safflower. Similarly, the highest expression of *CtF3'H10*, *CtF3'H13*, *CtF3'H23*, *CtF3'H27* and *CtF3'H34* was only observed in the fading stage. On the other hand, the expression pattern of *CtF3'H20*, *CtF3'H25*, *CtF3'H26*, *CtF3'H32* and *CtF3'H33* showed relatively higher expression level in full flowering stage compare to other stages. Together, the expression profiling and fold change ratio of *CtF3'H* genes during flower development suggested a variable transcriptional regulation of floral pigments during flowering maturation in safflower.

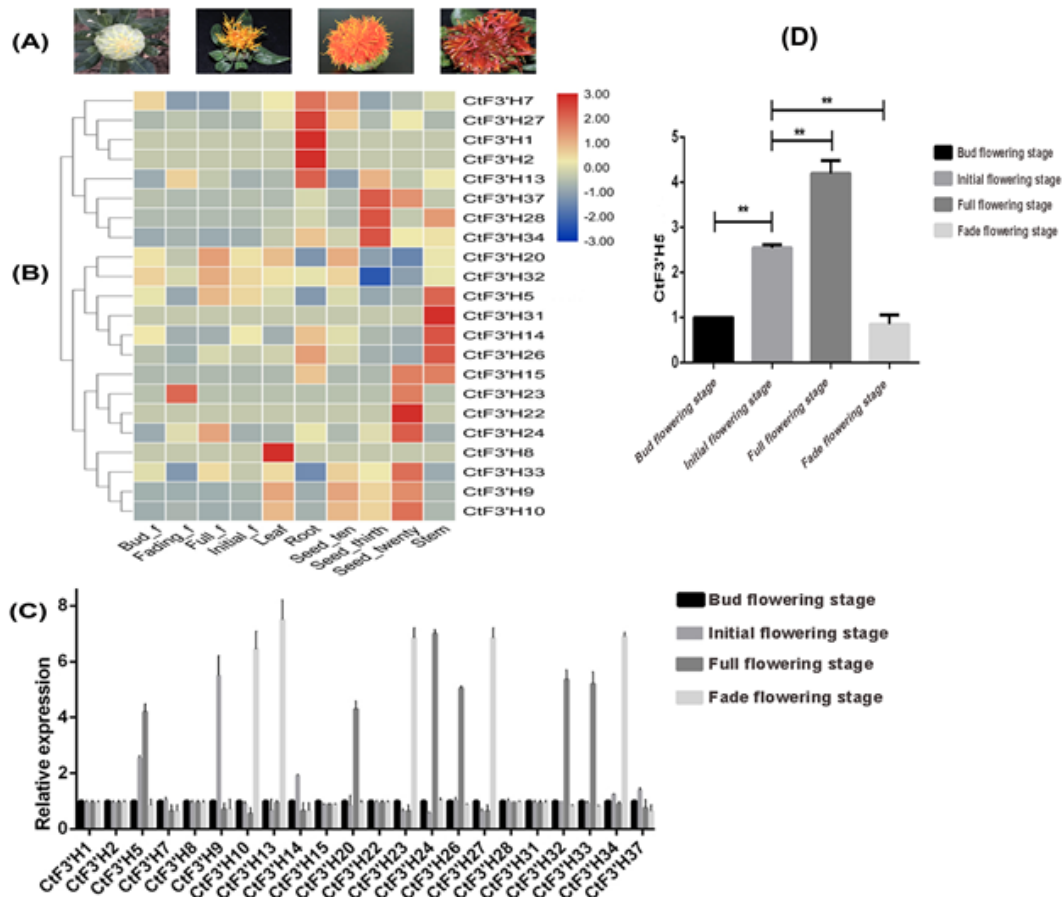


Figure 4. Expression profiling of *CtF3'H* genes in safflower (A) the illustration of four different flowering stages in safflower including bud, initial, full, and fade flowering. (B) the heat map representing the differential expression level of *CtF3'H* genes in different tissues of safflower. (C) qRT-PCR analysis of 22 putative *CtF3'H* genes in different flowering stages of safflower under normal conditions. Different background colours represent different periods of the safflower. Expression analysis was standardized using the *18S rRNA* as a reference gene. Error bars are estimated from differences in the expression patterns of three independent biological samples. (D) the relative expression level of *CtF3'H5* gene in four different flowering stages of safflower.

Molecular cloning, sequence analysis and subcellular localization of CtF3'H5

The full-length *CtF3'H5* target gene encoding 455 amino acids was cloned from safflower using Taq polymerase enzyme, and the result was confirmed on 1% agarose gel electrophoresis and Sanger sequencing. The sequencing results showed that the sequence was found consistent with the expected size of the *CtF3'H5* gene, indicating successfully cloning for further investigations (Figure 5A). The 3D structure prediction of *CtF3'H5* target protein was performed using the SWISS-MODEL online software (Figure 5B) (Lu *et al.*, 2018) by translating the genomic sequence of *CtF3'H5* with the help of ExPasy ProtParam 4.0 online tool. The theoretical molecular weight was 50.03 kD with the average total hydrophobic value -0,035, suggesting that

CtF3'H5 is a hydrophilic protein. The protein sequence of CtF3'H5 was further subjected to multiple sequence alignments along with other F3'H proteins obtained from various plant species, including *Callistephus chinensis*, *Camellia sinensis*, *Dahlia pinnata*, and *Centaurea cyanus* (Figure 5C). It was found that the amino acid sequence of CtF3'H5 contained the standard conserved "GGEK" sequence at positions 434 to 437, which is specific to the F3'H. The following conserved sequence, "LPPGP," starting at position 39, which is the main conserved domain of cytochrome P450 was also detected.

In addition, the subcellular localization prediction for CtF3'H5 protein was carried out by exploiting the online webserver of ProtComp version 9.0, which indicated the predominant signals to the plasma membrane. However, to confirm the subcellular localization of the CtF3'H5, we investigated the transient expression analysis using the fusion vector of pCAMBIA1302 and CtF3'H5 constructing pCAMBIA1302-CtF3'H5-GFP recombinant vector. The recombinant vector was transformed into the onion epidermal cells by the *Agrobacterium*-mediated infection system under controlled conditions and these results were analysed under a scanning electron microscope. Our findings marginally demonstrated that the GFP signals primarily appeared within the plasma membrane compared to other parts of the cell (Figures 6D-F). However, on the other hand, the GFP signals in the control vector indicated a dispersed pattern throughout the cell (Figure 6A-C).

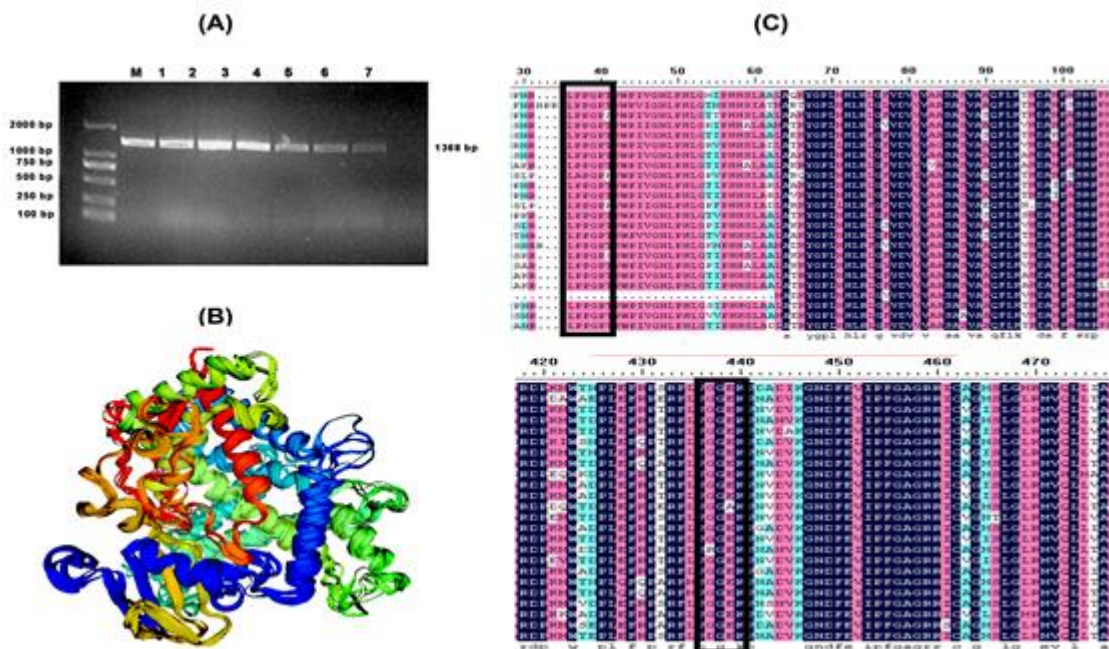


Figure 5. Molecular cloning and sequence analysis of the CtF3'H5 from safflower (A) PCR amplification of *CtF3'H5* gene; (B) Multiple sequence alignment of F3'H amino acid obtained from safflower and other plant species; (C) Prediction of the tertiary structure of the protein encoded by the *CtF3'H5*

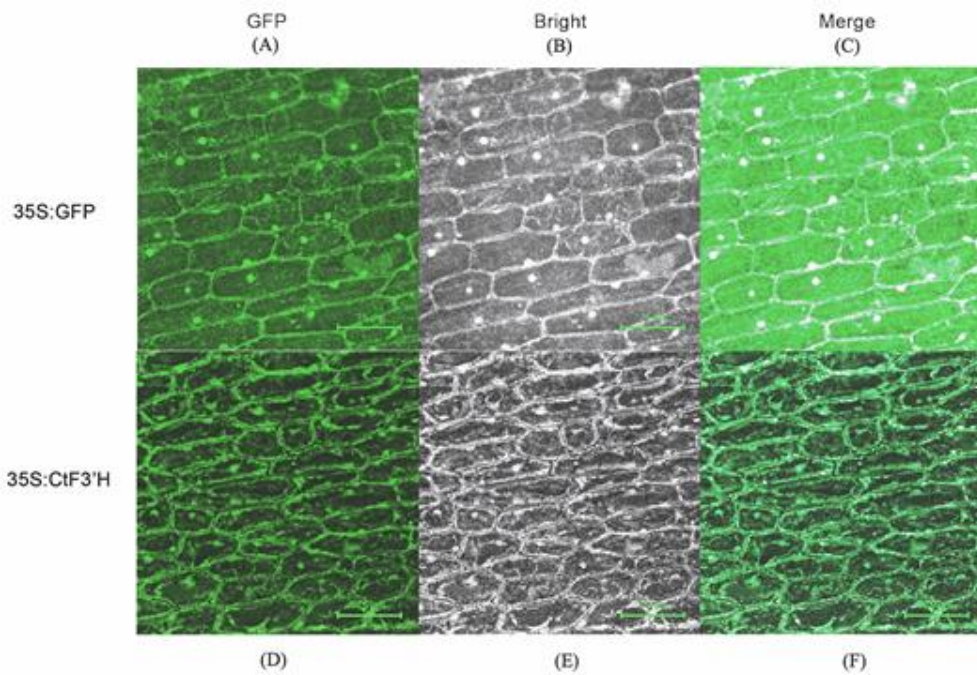


Figure 6. Subcellular localization of CtF3'H5 through GFP fluorescence under SEM (A–C) The GFP fluorescence result of empty vector pCAMBIA-GFP vector in onion (D–F) The GFP fluorescence result of pCAMBIA-CtF3'H5–GFP vector. Scar bar = 75 μ m.

Prokaryotic expression and Western Blot analysis of CtF3'H5

To investigate the heterologous expression of CtF3'H5 in bacterial system, the full-length cDNA sequence of CtF3'H5 was cloned and then ligated into pET28a+ prokaryotic expression vector. The construction of the recombinant vector (pET28a+-CtF3'H5) was transformed into bacterial cells of BL21-DE3 strain by heat and shock method (Figure 7A). Finally, the recombinant vector was confirmed with double restriction digestion system (Figure 7B), and sequencing. In the follow-up experiments, the heterologous expressed CtF3'H5 protein was induced by treating the recombinant bacterial cells of BL21 strain carrying the (pET28a+-CtF3'H5) vector with different concentrations of IPTG at different time periods. After the induction, the target protein product was confirmed with coomassie brilliant blue staining (SDS-PAGE), which resulted in a 50.3 kDa protein (Figure 7C). Noticeably, the use of different concentration of IPTG did not influence the protein expression at different time points. Similarly, the western blot hybridization method also detected the stably expressed CtF3'H5 protein at different IPTG concentrations. As described in figure 8D, we observed a unique band of the same size (50.3 kDa) and a change in the protein expression level when different IPTG concentrations were used. These results strongly suggested that the target CtF3'H5 protein was stably expressed in the bacterial system. The detection of CtF3'H5 protein confirmed by western blot analysis on a nylon membrane indicated that the expected size of the target protein was consistent with its theoretical molecular weight, however, the expression level could be marginally influenced when induced with different concentrations of IPTG.

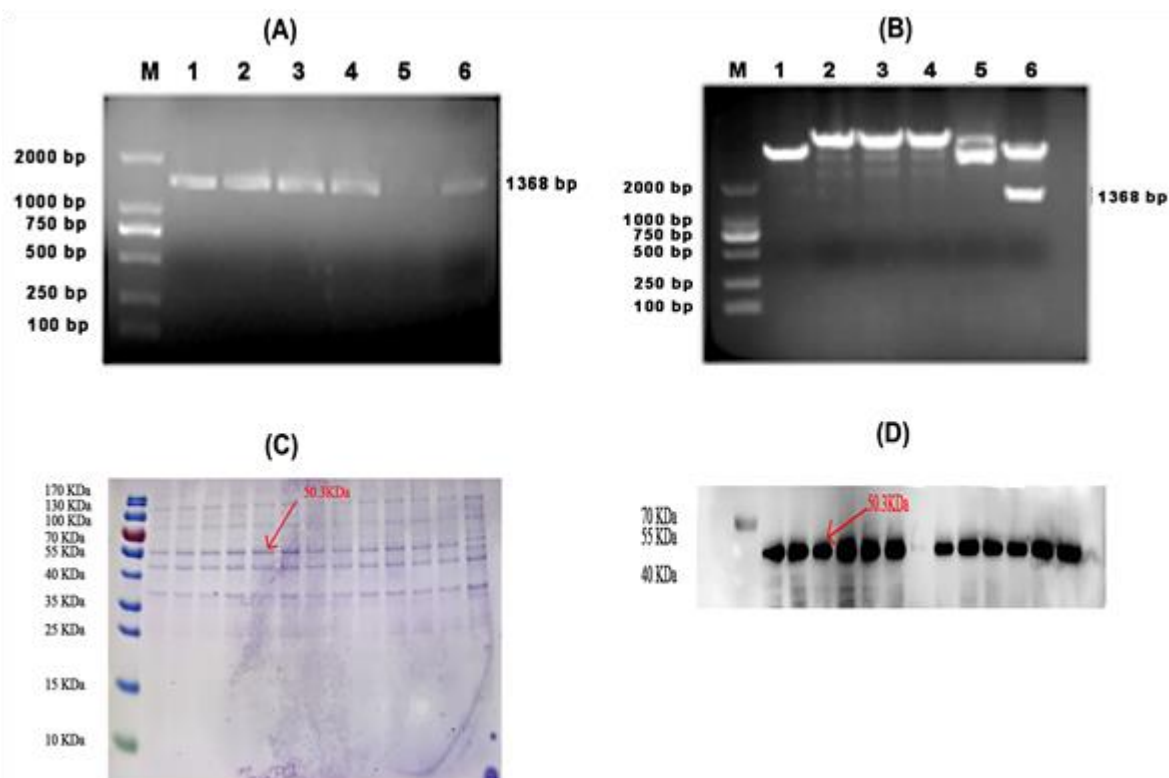


Figure 7. Heterologous expression and western blot hybridization of CtF3'H5 target protein (A) The real-time quantitative expression level of CtF3'H5 in four different flower developmental stages in safflower (B-C) Cloning and confirmation of PET28a+-CtF3'H5 recombinant plasmid using double restriction digestion system. (D) SDS-PAGE Analysis of heterologously expressed CtF3'H5 protein under various concentrations of 4 IPTG (E) Western Blot analysis of CtF3'H5 protein
The different concentrations of IPTG used in this experiment were as follow: L 0.2 IPTG; 0.4 IPTG, 0.5 IPTG, 0.6 IPTG, 0.7 IPTG, 0.8 IPTG.

Discussion

The flavonoid biosynthesis is one of the essential secondary metabolic pathways in higher plants (Liu *et al.*, 2019; Dermauw *et al.*, 2020). The regulation of the flavonoid pathway and its core structural genes have been shown to play an essential part in plant evolution, assisting plants in adapting to a variety of biotic and abiotic stresses. During flavonoid pathway, F3'H and F3'5'H enzymes play a vital role in determining the flower colour, seed pods, and the colour of stems and leaves in many plants (e.g, petunia, carnation, or rose). It has been revealed that the ratio of F3'H to F3'5'H determines the colour level of grapes (Nelson and Werck-Reichhart, 2011). Flavonoid 3'-hydroxylase is a critical enzyme in the flavonoid metabolism pathway, catalysing the hydroxylation of position 3' of the B-ring of naringenin and dihydro kaempferol. Naringenin and dihydrokaempferol are oxidized to form a series of essential intermediates in the flavonoid pathway. The structural stability and antioxidant function of these intermediates are closely related to the F3'H enzyme (Wei *et al.*, 2015). Several studies have identified naringenin as the optimal substrate for CsF3'H (Wang *et al.*, 2014; Lv *et al.*, 2017). However, a low-copy number of *F3'H* and *F3'5'H* genes have been reported in many plant species due to which the explicit regulatory path of these flavonoid hydroxylases still remained unclear. In this study, we presented a brief overview of the genome-wide identification and several structural and functional characteristics of the *CtF3'H* gene family in safflower. The process of lineage-specific evolution is the most widely observation in determining the evolution and diversification of plants. The numbers candidate F3'H

genes was found to be significantly unequal among different plants species. In *Triticum urartu*, only 2 copies of *F3'H* genes and in *Triticum aestivum* 9 copies of *F3'H* genes were identified. Similarly, in other monocot plants such *Sorgham bicolor* contained 5 copies *F3'H* genes and 4 copies of *F3'H* genes were found in *Brachypodium distachyon*. Here, we identified a total of 22 *CtF3'H* genes in safflower genome suggesting a large number of lineage-specific distribution of this important class of gene family. Another reason of vast *CtF3'H* expansion in the dicot safflower could be associated with tandem duplication and proximal duplication respectively. The Phylogeny analyses of *CtF3'H* proteins with other plants *F3'H* proteins identified explained the early evolution history and conservation of *CtF3'H* in safflower. Further classifications revealed that most of the *CtF3'H* proteins shared an identical pattern of evolutionary pattern with *S. miltiorrhiza* and *V. vinifera*, *P. hybrida* and *I. purpurea*. These findings were consistent with evolution of plant P450 family as described by (Ahmad *et al.*, 2020; Barvkar *et al.*, 2012).

Flavonoid biosynthesis is tightly regulated by different spatial and time signals that can limit the accumulation of these compounds in plants (Ahmad *et al.*, 2019; Musiol-Kroll *et al.*, 2019). Several studies have demonstrated the role of the *F3'H* gene in regulating plant metabolism and changes in flavonoid composition in flower petals (Ueyama *et al.*, 2002). The *F3'H*'s functional characterization has previously been reported in maize, where it regulates the onset of red aleurone colour (Sharma *et al.*, 2012). In rice, the two members of this family namely *CYP75B3* and *CYP75B4* underpin the synthesis of 3'-hydroxylated flavonoids and tricin (Park *et al.*, 2016). While in sorghum, it is associated with the formation of 3-deoxyanthocyanidins (Mizuno *et al.*, 2016). The functional divergence followed by gene duplication might also correlated with the transcriptional regulation of the *F3'H* genes, which may lead to a deeper understanding of these flavonoid hydroxylases in the downstream regulation of flavonoid pathway. Consequently, gene expression studies were extensively utilized to investigate the responsiveness of various flavonoid biosynthetic genes towards diverse biotic and abiotic stimuli. In this study, we extensively conducted the expression analysis of *CtF3'H* genes at different flowering stages using RNA-seq and qRT PCR analysis. Our findings confirmed significant differences during the expression of *CtF3'H* transcripts at different flowering stages. Initially, the expression level of most *CtF3'H* genes demonstrated down-regulation at bud flowering stage, in which the floral pigments was apparently lower. However, a diverse pattern of expression was observed the in the initial and full flowering stages, indicating a relatively higher expression than other flowering stages. The reason of the higher expression could be associated with the regulation and determination of red and yellow flowering colours. Nonetheless, the expression level of several *CtF3'H* genes including *CtF3'H1* and *CtF3'H2* did not showed significant changes in their expression at either of the flowering stages in safflower, suggesting the synergetic affects of possible mutations, which limits some steps in flavonoid biosynthesis (Jia *et al.*, 2019). Gene transcription divergence is also being discovered for distantly related *F3'H*s and *F3'5'H*s in numerous eudicot plants, including *F3'H*s in tea tree leaves (Wei *et al.*, 2015), confirming that functional diversity at the gene level of transcription is a typical finding after gene evolution via duplication.

We further characterized a putative *CtF3'H5* gene by cloning the full-length cDNA sequence and performed several molecular testing's to unleash the stability and reliability of this candidate gene in safflower. Various conserved amino acid sequences suggested that *CtF3'H5* protein is highly conserved in other plant species (Biasini *et al.*, 2014). The subcellular localization study is probably one of the essential links to estimate the final function of a target protein (Naqvi *et al.*, 2016). Hence, we experimentally investigated the localization of the *CtF3'H5* target protein using the onion epidermal cells through *Agrobacterium*-mediated transformation by transiently expressing the fusion construct of pCAMBIA1302-*CtF3'H5* containing the GFP cassettes. The result suggested that the target protein of *CtF3'H5* was detected in the plasma membrane of the onion cell epithelial cells, disagreeing with the previous development of (Chen *et al.*, 2017), who reported the *F3'H* gene on *Brassica napus* was localized into the endoplasmic reticulum. These differences may occur due to the chromosomal position of the gene, which varies in safflower as compare to *Brassica napus*. However, the function of this protein was generally ascribed to flavonoid biosynthesis.

The enzymatic activity of the Cytochrome P450 system exhibits a unique oxidative and reductive pattern. Nevertheless, to comprehend the vast substrate specificity of complex P450 proteins, establishing a stable and cost-effective expression system is challenging. The benefit of bacterial P450 expression systems has been covered by (Zelasko *et al.*, 2013), presenting a fast and accurate expression system. We demonstrated a capable and stable expression system for the production of heterologously expressed CtF3'H5 protein utilizing the prokaryotic expression system of bacteria. Our results identified an expected target protein of CtF3'H5 expression by a competent strain of bacteria (Bl21) comprising a 50.3 kDa size of band size in the SDS PAGE analysis. The level of CtF3'H5 target protein induction with IPTG indicated a stable production of the target protein; however, the use of different IPTG concentrations could affect the expression of our target protein (Wei and Chen, 2018), as described in our western blot hybridization analysis. The present work relates to the comprehensive genome-wide identification and understanding the evolutionary relationship of *CtF3'H*-encoding genes in safflower. We reported several regulatory networks such as the highlight conserved topology of CtF3'H including the identification of two highlight conserved regions "GGEK" and "LPPGP" underlying their functional importance. Similarly, the investigation of important cis-acting elements such as G-Box, MYB and MYC responsive elements, drought (MBS) and abscisic acid (ABRE) responsive elements highlighted significant insights into the underlying molecular regulation of CtF3'H genes in secondary metabolism and various abiotic stress-related pathways. The expression level of *CtF3'HS* during flower development revealed a differential pattern of regulation, suggesting their potential roles in flower development and secondary metabolism. Altogether, this study uncovered the fundamental groundwork on understanding the molecular mechanisms and regulation of *CtF3'H* underlying the hydroxylation of 3' B-ring containing flavonoid biosynthesis and also offers new avenues for future functional researches on F3'H-encoding genes in plants.

Conclusions

The current study focuses on the systematic genome-wide discovery, evolution and further characterization of the duplicated CtF3'H-encoding genes in safflower. The phylogenetic analysis, conserved topology, and expression diversity of CtF3'H in different flowering stages of safflower unveiled several important highlights into investigating the underlying function of these genes. In addition, we confirmed that these CtF3'H genes underwent robust evolutionary divergence that could be interconnected with the functional specialization of their duplicated members from their parent CYP450 gene families in safflower. Together, this research work provides the foundation for screening highly expressive CtF3'H genes for future functional studies and will help in the acceleration of molecular breeding programmes in plants.

Authors' Contributions

Conceptualization, LX, NA and JL; Data curation, NH, KJ and NA; Formal analysis, MX; Methodology, NH and NH; Software, ZX and WY; Supervision, LX, and JL; Validation, WN; Writing – original draft, NH, KJ, and NA; Writing – review & editing, NA, YN LX. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 31771868), the Science and Technology Development Project of Jilin Province (No. 20190201172JC, 20190201175JC,

20150623024TC-11), Student Innovation and Entrepreneurship Training Program of Jilin Agricultural University.

Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

References

- Ahmad N, Jianyu L, Xu T, Noman M, Jameel A, Na Y, ... Fawei W (2019). Overexpression of a novel cytochrome P450 promotes flavonoid biosynthesis and osmotic stress tolerance in transgenic Arabidopsis. *Genes* 10(10):756. <https://doi.org/10.3390/genes10100756>.
- Ahmad N, Li T, Liu Y, Hoang NQV, Ma X, Zhang X, Liu J, Yao N, Liu X, Li H (2020). Molecular and biochemical rhythms in dihydroflavonol 4-reductase-mediated regulation of leucoanthocyanidin biosynthesis in *Carthamus tinctorius* L. *Industrial Crops and Products* 156:112838. <https://doi.org/10.1016/j.indcrop.2020.112838>
- Alaiye A, Kaya E, Pnarbal MZ, Harmanc N, Cingi C (2020). An experimental comparison of the analgesic and anti-inflammatory effects of Safflower oil, benzydamine HCl, and naproxen sodium. *Journal of Medicinal Food* 23(8). <https://doi.org/10.1089/jmf.2019.0157>
- Ambreen H, Kumar S, Kumar A, Agarwal M, Jagannath A, Goel S (2018). *Carthamus tinctorius* association mapping for important agronomic traits in safflower (L.) Core collection using microsatellite markers. *Frontiers in plant science* 9:402. <https://doi.org/10.3389/fpls.2018.00402>
- Barvkar VT, Pardeshi VC, Kale SM, Kadoo NY, Gupta VS (2012). Phylogenomic analysis of UDP glycosyltransferase 1 multigene family in *Linum usitatissimum* identified genes with varied expression patterns. *BMC Genomics* 13(1):175-175. <https://doi.org/10.1186/1471-2164-13-175>
- Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, ... Bordoli L (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research* 42(W1):W252-W258. <https://doi.org/10.1093/nar/gku340>
- Castellarin SD, Di Gaspero G, Marconi R, Nonis A, Peterlunger E, Paillard S, ... Testolin R (2006). Colour variation in red grapevines (*Vitis vinifera* L.): genomic organisation, expression of flavonoid 3'-hydroxylase, flavonoid 3'5'-hydroxylase genes and related metabolite profiling of red cyanidin-/blue delphinidin-based anthocyanins in berry skin. *BMC Genomics* 7(1):1-17. <https://doi.org/10.1186/1471-2164-7-12>
- Cerqueira A, Santana I, Araújo J, Lima H, Batatinha M, Branco A, Santos Junior M, Botura M (2021). Homology modeling, docking, molecular dynamics and *in vitro* studies to identify *Rhipicephalus microplus* acetylcholinesterase inhibitors. *Journal of Biomolecular Structure & Dynamics* 1-11. <https://doi.org/10.1080/07391102.2021.1889666>
- Chen K, Liu H, Lou Q, Liu Y (2017). Ectopic expression of the grape hyacinth R2R3-MYB transcription factor gene, induces anthocyanin accumulation in tobacco. *Frontiers in Plant Science* 8:965. <https://doi.org/10.3389/fpls.2017.00965>.
- Cho H, Kim O, Lee Y, Kang L, Nguyen C, Ishihara A, Kim H (2017). Feruloylserotonin inhibits hydrogen peroxide-induced melanogenesis and apoptosis in B16F10 and SK-Mel-2 melanoma cells. *Biochemical and Biophysical Research Communications* 491(4):973-979. <https://doi.org/10.1016/j.bbrc.2017.07.158>
- Choudhary D, Pandey A, Adhikary S, Ahmad N, Trivedi R (2016). Genetically engineered flavonol enriched tomato fruit modulates chondrogenesis to increase bone length in growing animals. *Scientific Reports* 6:21668. <https://doi.org/10.1038/srep21668>
- De Palma M, Fratianni F, Nazzaro F, Tucci M (2014). Isolation and functional characterization of a novel gene coding for flavonoid 3'-hydroxylase from globe artichoke. *Biologia Plantarum* 58(3):445-455. <https://doi.org/10.1007/s10535-014-0424-7>
- Dermauw W, Van Leeuwen T, Feyereisen R (2020). Diversity and evolution of the P450 family in arthropods. *Insect Biochemistry and Molecular Biology* 127:103490. <https://doi.org/10.1016/j.ibmb.2020.103490>.

- Du C, Hou J, Wang C, Zhang M, Zheng Y, Yang G, Hu Y (2021). Effects of safflower yellow on cholesterol levels in serum and brain tissue of APP/PS1 mice. *Metabolic Brain Disease* 36(4):557-569. <https://doi.org/10.1007/s11011-021-00680-0>
- Gao, S, Xu X, Zeng W, Xu S, Lv Y, Feng Y, ... Chen J (2020). Efficient Biosynthesis of (2)-Eriodictyol from (2)-Naringenin in through a combination of promoter adjustment and directed evolution. *ACS Synthetic Biology* 9(12):3288-3297. <https://doi.org/10.1021/acssynbio.0c00346>
- Guo L, Gao L, Ma X, Guo F, Ruan H, Bao Y, ... Wang Y (2019). Functional analysis of flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylases from tea plant (*Camellia sinensis*), involved in the B-ring hydroxylation of flavonoids. *Gene* 717:144046. <https://doi.org/10.1016/j.gene.2019.144046>
- Han Y, Vimolmangkang S, Soria-Guerra RE, Rosales-Mendoza S, Zheng D, Lygin AV, Korban SS (2010). Ectopic expression of apple F3'H genes contributes to anthocyanin accumulation in the *Arabidopsis* tt7 mutant grown under nitrogen stress. *Plant Physiology* 153(2):806-820. <https://doi.org/10.1104/pp.109.152801>
- Hong Y, Ahmad N, Tian Y, Liu J, Wang L, Wang G, ... Liu W (2019). Genome-wide identification, expression analysis, and subcellular localization of *Carthamus tinctorius* bHLH transcription factors. *International Journal of Molecular Sciences* 20(12):3044. <https://doi.org/10.3390/ijms20123044>
- Iwashina T, Benitez ER, Takahashi R (2006) Analysis of flavonoids in pubescence of soybean near-isogenic lines for pubescence color loci. *Journal of Heredity* 97:438-443. <https://doi.org/10.1093/jhered/esl027>
- Jia Y, Li B, Zhang Y, Zhang X, Xu Y, Li C (2019). Evolutionary dynamic analyses on monocot flavonoid 3'-hydroxylase gene family reveal evidence of plant-environment interaction. *BMC Plant Biology* 19(1):1-16. <https://doi.org/10.1186/s12870-019-1947-z>
- Knowles P (1972). The plant geneticist's contribution toward changing lipid and amino acid composition of safflower. *Journal of the American Oil Chemists Society* 49(1):30-32. <https://doi.org/10.1007/BF02545134>
- Li H, Li L, Shang Guan G, Jia C, Deng S, Noman M, ... Zhang X (2020). Genome-wide identification and expression analysis of bZIP gene family in *Carthamus tinctorius* L. *Scientific Reports* 10(1):1-15. <https://doi.org/10.1038/s41598-020-72390-z>
- Lin K, Qin Z, Qu C, Chen X, Jiang Q, Li M, ... Li D (2021). Hydroxyl safflower yellow B combined with doxorubicin inhibits the proliferation of human breast cancer MCF-7 cells. *Oncology Letters* 21(5):426. <https://doi.org/10.3892/ol.2021.12687>
- Liu X, Ahmad N, Yang L, Fu T, Kong J, Yao N, Dong Y, Wang N, Li X, Wang F (2019). Molecular cloning and functional characterization of chalcone isomerase from *Carthamus tinctorius*. *AMB Express* 9(1):1-12. <https://doi.org/10.1186/s13568-019-0854-x>
- Lu R, Yuan T, Wang Y, Zhang T, Yuan Y, Wu D, ... Cheng Y (2018). Spontaneous severe hypercholesterolemia and atherosclerosis lesions in rabbits with deficiency of low-density lipoprotein receptor (LDLR) on exon 7. *EBioMedicine* 36:29-38. <https://doi.org/10.1016/j.ebiom.2018.09.020>
- Lv GY, Guo XG, Xie LP, Xie CG, Zhang XH, Yuan Y, ... Guo AG (2017). Molecular characterization, gene evolution, and expression analysis of the fructose-1, 6-bisphosphate aldolase (FBA) gene family in wheat (*Triticum aestivum* L.). *Frontiers in Plant Science* 8. <https://doi.org/10.3389/fpls.2017.01030>
- Min F, Sun H, Wang B, Ahmad N, Li H (2020). Hepatoprotective effects of hydroxysafflor yellow A in D-galactose-treated aging mice. *European Journal of Pharmacology* 173214. <https://doi.org/10.1016/j.ejphar.2020.173214>
- Mizuno H, Yazawa T, Kasuga S, Sawada Y, Kanamori H, Ogo Y, Hirai MY, Matsumoto T, Kawahigashi H (2016). Expression of flavone synthase II and flavonoid 3'-hydroxylase is associated with color variation in tan-colored injured leaves of *sorghum*. *Frontiers in Plant Science* 7:1718. <https://doi.org/10.3389/fpls.2016.01718>
- Musiol-Kroll EM, Tocchetti A, Sosio M, Stegmann E (2019). Challenges and advances in genetic manipulation of filamentous actinomycetes – the remarkable producers of specialized metabolites. *Natural Product Reports* 36(12). <https://doi.org/10.1039/C9NP00029A>
- Nagamatsu A, Masuta C, Matsuura H, Kitamura K, Abe J, Kanazawa A (2009). Down-regulation of flavonoid 3'-hydroxylase gene expression by virus-induced gene silencing in soybean reveals the presence of a threshold mRNA level associated with pigmentation in pubescence. *Journal of Plant Physiology* 166(1):0-39. <https://doi.org/10.1016/j.jplph.2008.02.002>
- Naqvi AAT, Anjum F, Khan FI, Islam A, Hassan MI (2016). Sequence analysis of hypothetical proteins from *Helicobacter pylori* 26695 to identify potential virulence factors. *Genomics Informatics* 14(3). <https://dx.doi.org/10.5808%2FGI.2016.14.3.125>

- Nelson D, Werck-Reichhart D (2011). A P450-centric view of plant evolution. *The Plant Journal* 66(1):194-211. <https://doi.org/10.1111/j.1365-313X.2011.04529.x>
- Panche A, Diwan A, Chandra S (2016). Flavonoids: an overview. *Journal of Nutritional Science* 5. <https://doi.org/10.1017/jns.2016.41>
- Park S, Choi MJ, Lee JY, Kim JK, Ha SH, Lim SH (2016). Molecular and biochemical analysis of two rice flavonoid 3'-hydroxylase to evaluate their roles in flavonoid biosynthesis in rice grain. *International Journal of Molecular Sciences* 17:1549. <https://doi.org/10.3390/ijms17091549>
- Ren T, Zheng P, Zhang K, Liao J, Xiong F, Shen Q, Ma Y, Fang W, Zhu X (2021). Effects of GABA on the polyphenol accumulation and antioxidant activities in tea plants (*Camellia sinensis* L.) under heat-stress conditions. *Plant Physiology and Biochemistry* 159:363-371. <https://doi.org/10.1016/j.plaphy.2021.01.003>
- Schrago C, Aguiar B, Mello B (2018). Comparative evaluation of maximum parsimony and Bayesian phylogenetic reconstruction using empirical morphological data. *Journal of Evolutionary Biology* 31(10):1477-1484. <https://doi.org/10.1111/jeb.13344>
- Sharma M, Chai C, Morohashi K, Grotewold E, Snook ME, Chopra S (2012). Expression of flavonoid 3'-hydroxylase is controlled by P1, the regulator of 3-deoxyflavonoid biosynthesis in maize. *BMC Plant Biology* 12:1-13. <https://doi.org/10.1186/1471-2229-12-196>
- Ueyama Y, Suzuki KI, Fukuchi-Mizutani M, Fukui Y, Tanaka Y (2002). Molecular and biochemical characterization of torenia flavonoid 3'-hydroxylase and flavone synthase II and modification of flower color by modulating the expression of these genes. *Plant Science* 163(2):253-263. [https://doi.org/10.1016/S0168-9452\(02\)00098-5](https://doi.org/10.1016/S0168-9452(02)00098-5)
- Wang X, Xu Y, Chu C, Li H, Mi J, Wen Z, Zhang S, Wang Q, Quan S (2019). Effect of safflower yellow on early type II diabetic nephropathy: a systematic review and meta-analysis of randomized controlled trials. *Journal of Pediatric Endocrinology & Metabolism* 32(7):653. <https://doi.org/10.1515/jpem-2018-0425>
- Wang YS, Xu YJ, Gao LP, Yu O, Wang XZ, He XJ, Jiang XL, Liu YJ, Xia T (2014). Functional analysis of Flavonoid 3',5'-hydroxylase from Tea plant (*Camellia sinensis*): critical role in the accumulation of catechins. *BMC Plant Biology* 14(1):347. <https://doi.org/10.1186/s12870-014-0347-7>
- Wei K, Chen H (2018). Global identification, structural analysis and expression characterization of cytochrome P450 monooxygenase superfamily in rice. *BMC Genomics* 19(1):35. <https://doi.org/10.1186/s12864-017-4425-8>
- Wei K, Wang L, Zhang C, Wu L, Li H, Zhang F, Cheng H (2015). Transcriptome analysis reveals key flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase genes in affecting the ratio of dihydroxylated to trihydroxylated catechins in *Camellia sinensis*. *PloS One* 10(9):e0137925. <https://doi.org/10.1371/journal.pone.0137925>
- Wu ZJ, Li XH, Liu ZW, Xu ZS, Zhuang J (2014). De novo assembly and transcriptome characterization: novel insights into catechins biosynthesis in *Camellia sinensis*. *BMC Plant Biology* 14(1):1-16. <https://doi.org/10.1186/s12870-014-0277-4>
- Zelasko S, Palaria A, Das A (2013). Optimizations to achieve high-level expression of cytochrome P450 proteins using *Escherichia coli* expression systems. *Protein Expression and Purification* 92(1):77-87. <https://doi.org/10.1016/j.pep.2013.07.017>



The journal offers free, immediate, and unrestricted access to peer-reviewed research and scholarly work. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles, or use them for any other lawful purpose, without asking prior permission from the publisher or the author.

License - Articles published in *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* are Open-Access, distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) License.

© Articles by the authors; UASVM, Cluj-Napoca, Romania. The journal allows the author(s) to hold the copyright/to retain publishing rights without restriction.