Fan S *et al.* (2021)
Notulae Botanicae Horti Agrobotanici Cluj-Napoca
Volume 49, Issue 3, Article number 12489
DOI:10.15835/nbha49312489
Research Article

Genome-wide investigation of *Hydroxycinnamoyl CoA: Shikimate Hydroxycinnamoyl Transferase* (*HCT*) gene family in *Carthamus tinctorius L.*

Sun FAN[1a], Naveed AHMAD[1,3b], Jin LIBO[2,4c], Zhang XINYUE[1], Ma XINTONG[1], Nguyen Q.V. HOANG[1], Ali I. MALLANO[3], Wang NAN[1], Yang ZHUODA[1], Liu XIUMING[1*], Yao NA[1*]

[1]*Jilin Agricultural University, Ministry of Education Engineering Research Center of Bioreactor and Pharmaceutical Development, Changchun 130118, China; porker0927@outlook.com; naveedjlau@gmail.com; 2778935775@qq.com; mxt7142020@163.com; viethoangjlau@gmail.com; 2124136734@qq.com; Xiuming1211@163.com (*corresponding author); nayao1103@jlau.edu.cn (*corresponding author)*
[2]*Wenzhou University, Institute of Life Sciences, Wenzhou 325035, China;*
[3]*Shandong Academy of Agricultural Sciences, Biotechnology Research Center, Shandong Provincial Key, Laboratory of Crop Genetic Improvement, Ecology and Physiology, Jinan 250100, China; iamallano@sau.edu.pk;*
[4]*Wenzhou University, Biomedical Collaborative Innovation Center of Zhejiang Province, Wenzhou, Zhejiang 325035, China*
[a,b,c]*These authors contributed equally to this work.*

Abstract

Hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl transferase (HCT) is mainly associated with monolignol biosynthesis, a central precursor to producing guaiacyl and syringyl lignins in plants. However, the explicit regulatory mechanism of HCT-mediated monolignol biosynthesis in plants still remained unclear. Here, the genome-wide analysis of the *HCT* gene family in *Carthamus tinctorius* as a target for understanding growth, development, and stress-responsive mechanisms was investigated. A total of 82 *CtHCT* genes were identified and characterized. Most of the CtHCTs proteins demonstrated the presence of two common conserved domains, including HXXXD and DFGWG. In addition, the conserved structure of protein motifs, PPI network, cis-regulatory units, and gene structure analysis demonstrated several genetic determinants reflecting the wide range of functional diversity of CtHCT-encoding genes. The observed expression analysis of *CtHCT* genes in different flowering stages under normal conditions partially highlighted their putative roles in plant growth and development pathways. Moreover, *CtHCT* genes appeared to be associated with abiotic stress responses as validated by the expression profiling in various flowering phases under light irradiation and MeJA treatment. Altogether, these findings provide new insights into identifying crucial molecular targets associated with plant growth and development and present practical information for understanding abiotic stress-responsive mechanisms in plants.

Keywords: abiotic stress; *Carthamus tinctorius*; expression diversity; *HCT* gene family; monolignol biosynthesis

Received: 31 August 2021. Received in revised form: 21 Sep 2021. Accepted: 22 Sep 2021. Published online: 27 Sep 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

Carthamus tinctorius is commonly known as safflower or 'bastard saffron', which belongs to the Asteraceae family of the plant kingdom. The increasing demand for its oilseed, which is extremely rich in conjugated linoleic acid, has attracted the attention of plant biologists worldwide. Safflower's oilseed consists of 80% of octadecadienoic acid, which helps regulate the rate of cholesterol and avert diseases related to cardiovascular channels (Roh *et al.*, 2004). *C. tinctorius* is radically recommended for its medicinal and economic value in the mainland of China and west Asia. Following multiple phytochemical and pharmacologic investigations on safflower petals, it was discovered that the vital component that provides abundant resources of pharmacogenetic importance is flavonoid. Almost over 5000 types of phenolic compounds and lignin derivatives exist across the plant kingdom in which safflower shares a remarkable reservoir of flavonoids. The widely distributed classes of flavonoids in *C. tinctorius* mainly include carthamin chalcone glycoside, kaempferol glucosides, hydroxylsafflor yellow A&B, and quercetin glucosides (Ye and Gao, 2008; Zhang *et al.*, 2011).

Hydroxycinnamoyl CoA: Shikimate Hydroxycinnamoyl Transferase (*HCT*) gene family is widely known as acyl-CoA-dependent transferases, including various enzymes that utilize the commonly used donner molecule hydroxycinnamoyl-CoAs which catalyse a group of reactions and substrates (Chiang *et al.*, 2018). HCT synthesizes p-coumaroyl shikimate by transferring the p-coumaroyl group from the acyl donor p-coumaroyl-CoA to the acyl acceptor shikimate. It is an essential enzyme in the phenylpropanoid metabolism, conserved across all land plants (Chao *et al.*, 2021; Weng and Chapple, 2010). The downstream pathway catalyses the conversion of phenylalanine into a variety of hydroxycinnamic acids, which are the key precursor molecules of flavonoids, hydroxycinnamic acid conjugates, and lignins (Wang *et al.*, 2015). The metabolic pathways towards lignin and chlorogenic acid (CGA) presumably share common intermediates and enzymes. Various studies in the dicots *Solanum lycopersicum* (tomato), *Solanum tuberosum* (potato), *Nicotiana tabacum* (tobacco), and *Cynara cardunculus* (globe artichoke) have suggested different routes for CGA biosynthesis in plants (Escamilla-Treviño *et al.*, 2014; Payyavula *et al.*, 2015; Sonnante *et al.*, 2010). The enzyme (HCT) catalyzes CGA formation from caffeoyl CoA and quinic acid in the first pathway. To produce the caffeoyl CoA substrate, 4-coumaroyl CoA is converted to 4-coumaroyl shikimate by (HCT), this shikimate ester receptor molecule is hydroxylated by the enzyme 4-coumaroyl shikimate 3'-hydroxylase (C3'H), and the caffeoyl shikimate generated is converted to caffeoyl CoA by an HCT enzyme acting in the reverse direction. The second suggested route proceeds by synthesizing 4-coumaroyl quinate by HCT or HQT enzymes, followed by hydroxylation of the coumaroyl moiety by C3'H (Eudes *et al.*, 2016; Tsai *et al.*, 2006; Wagner *et al.*, 2007).

In vascular plants, the phenylpropanoid pathway is required to synthesize many metabolites, including lignin, which provides mechanical strength to vascular tissues and defense against various stresses (Vanholme *et al.*, 2019; Wang *et al.*, 2015). For instance, low temperature, high salinity, drought, mechanical injury, abscisic acids (ABAs), salicylic acid (S.A.), and hydrogen peroxide induce HcHCT expression in *Hibiscus cannabinus*. HcHCT increases abiotic stress tolerance in plants (Chowdhury *et al.*, 2012). In *Cucumis sativus*, the HCT expression was reduced with pectinase treatment and directing the phenylpropanoid pathway to generate H-lignin caused p-coumaraldehyde accumulation (Varbanova *et al.*, 2011). *HCT* is generally a conserved gene family among higher plants (Tohge *et al.*, 2013; Xu *et al.*, 2009). The comprehensive genome-wide characterization of HCT gene family as well as focusing on different structural components and functionally active sites would enables us to gain deeper understanding of HCT utilization during specialized metabolism in plants. In the current study, the structural and functional dynamics of the *HCT* gene family in *C. tinctorius* was unveiled by conducting genome-wide identification and expression analysis under normal sporadic conditions. This work will also provide unique insights into the underlying regulatory mechanism of plant growth and development under abiotic stress conditions.

Materials and Methods

Plant materials and treatment conditions

The seeds of 'JiHong' No. 1 variety of *C. tinctorius* were purchased from the Tacheng seed company, Xinjiang province of China, and then grown in the experimental station of Jilin Agricultural University under control conditions at 23 ± 2 °C. The flowering development period in *C. tinctorius* was recorded approximately 100 days from the date of cultivation. The flower samples from the bud, initial, full, and fade flowering stages were collected on the 99th day, 120th day, 140th day, and 160th day, respectively. In the case of light treatment, the healthy plants of *C. tinctorius* after the initiation of flowering were allowed to grow under the induction of normal light radiation (16.8 MJ/m^2) and weak intensity of light irradiation (4.6 MJ/m^2) maintained in the experimental station of the laboratory. For MeJA treatment, healthy flowering plants of *C. tinctorius* were treated with (100 μM solution) once daily for 7 days. The flower's petals from each flowering stage were collected and immediately placed in liquid nitrogen and preserved at – 80 °C until their subsequent use.

Genome-wide identification and sequences retrieval of CtHCTs

The Hidden Markov model (HMMsearch) of the HCT domain (PF02458) at the Pfam database, accessible at http://pfam.xfam.org/ (Finn *et al.*, 2015), was screened to investigate distribution of *CtHCTs* in the *C. tinctorius* genome. Moreover, we screened the entire set of CtHCT protein sequences for the existence of HXXXD and transferase domains using the online server of MARCOIL available at (*http://toolkit.tuebingen.mpg.de/marcoil*). The non-redundant protein sequences lacking the two domains of HCT were deleted from the analysis. After the assembly of CtHCT sequences, the genomic and protein sequences of HCTs were collected from *Arabidopsis thaliana, Cynara cardaunculus*, *Helianthus annuus*, *Lactuca sativa*, and *Artemisia annuua*. The HCTs sequences from *A. thaliana* and other plants were extracted from the Arabidopsis Information Resource (TAIR) (*http://www.Arabidopsis.org/*), NCBI (*https://www.ncbi.nlm.nih.gov/*), Phytozome (*https://phytozome.jgi.doe.gov/pz/portal.html*), and Plantgrn noble (*http://plantgrn.noble.org/*) respectively. Lastly, a dataset including 82 clean sequences of CtHCTs and 259 members of HCTs from the other five plants were assembled for further bioinformatics analyses. The physicochemical properties of 82 putative CtHCT proteins were investigated using different online tools. The theoretical isoelectric point (P.I.), protein lengths, and molecular weight (M.W.) of the obtained proteins were analyzed using ExPASyProtParam online tool (available online: at *https://web.expasy.org/protparam/*). The subcellular localization prediction of each gene was predicted using the cello web server (*http://cello.life.nctu.edu.tw/*) and WoLF PSORT (*https://wolfpsort.hgc.jp/*).

Phylogenetic reconstruction of HCT proteins

The full-length amino acid sequences of the 82 CtHCT proteins obtained from the *C. tinctorius* genome were subjected to multiple sequence alignment using Clustal W (2.0). The CtHCT sequences were numbered from CtHCT001-082 following their identification order. To analyse the evolutionary relationship and divergence of CtHCTs, an unrooted neighbour-joining phylogenetic tree with 1000 bootstrap method was generated together with 259 HCT sequences from *A. thaliana*, *C. cardaunculus*, *H. annuus*, *L. sativa*, and *A. annuua* using MEGA X software version 4.1 (Tamura *et al.*, 2011). The classifications of subfamilies were further analysed for genome-wide comparison.

Analyses of conserved protein motifs and PPI network

The clear sequences of CtHCTs were added to multiple sequence alignments in Clustal W (2.0) software to investigate the conserved amino acid composition and the presence of conserved protein motifs. The distribution and composition of the conserved protein motifs in CtHCTs were comprehensively investigated by adding each CtHCT protein sequence to MEME web server Version 4.8.1; available at

http://meme.nbcr.net/meme/cgi-bin/meme.cgi) using the default settings. The logos of these identified motifs were extracted from the MEME server. The graphical representation of protein motifs was edited in EvolView v.2 (*http://www.evolgenius.info/*). Furthermore, the prediction of protein interactive network of the putative CtHCT proteins was also investigated by uploaded CtHCT sequences to the online web server of STRING database version 10 (*https://string-db.org/*). The hierarchical network of interactor proteins associated with upstream and downstream regulation CtHCTs was created and exported from the STRING database.

Analysis of gene structure, cis-acting units, and Go term enrichment of CtHCTs

The gene structure organization, including exon and intron and UTR region along with the length of *CtHCT* genes, was examined from the C.D.sC.D.s and genomic sequences of *CtHCT* genes with the help of GSDS (Gene Structure Display Server (*http://gsds.cbi.pku.edu.cn/index.php*) according to the instructions given by (Hu *et al.*, 2014). Furthermore, to investigate the cis-regulatory units of the *CtHCT* promoter, the 2kb upstream 5` UTR flanking sequence of each gene was analyzed using the online webtool of PLACE (*https://sogo.dna.affrc.go.jp/*). In addition, the G.O. term analysis of *C. tinctorius* HCTs was determined with the help of Blast2GO software available at (*https://www.blast2go.com/*) following the instructions given by (Conesa and Götz, 2008). For this purpose, the full-length amino acid residues of CtHCT proteins were added to Blast2Go, and functional annotation of different categories was then identified.

Expression analysis of the putative CtHCTs

The experimental tissues of flower petals at different stages (bud, initial, fade, and full) were pulverized entirely in liquid nitrogen and then collected into centrifuge tubes. The total RNA extraction was performed using RNA ISOplus reagent (Takara Bio Co., Beijing, China), following the manufacturer's protocol. RNA quality was confirmed using OD260/280 concentrations through NanoDrop 2000 (ThermoFisher Scientific, Beijing, China) and 1% agarose gel electrophoresis. The first-strand cDNA templates were synthesized from the RNA isolated from each flowering stage using the reverse transcription kit (PrimeScript RT reagent kit with gDNA Eraser (Takara, Japan), following the instruction of the manufacturer's protocols. The quantitative real-time PCR assay was carried out to determine the transcription level of *CtHCT* genes using SYBR® Premix Ex Taq™ (TaKaRa). The system of StratageneMx3000P (Stratagene, CA, USA) was utilized to conduct qRT-PCR experiments. The relative expression level of *CtHCTs* at the bud stage was normalized to the housekeeping gene 18s ribosomal RNA expression. The fold change ratio was calculated according to the 2-ΔΔCT method (Livak and Schmittgen, 2001). Each experiment was repeated in three independent biological replicates. The gene-specific primers synthesized for each *CtHCT* are listed in (Table S1).

Statistical analysis

The results were calculated as mean ± S.D. with three replicates. The variations between means of each group were assessed by carried through a one-way analysis method of the variance with the help of (Statistix 8.1). P-value equal to 0.05 was kept statistically significant.

Results

Identification and characterization of CtHCTs in C. tinctorius

To identify all candidate *HCT* genes in *C. tinctorius*, we conducted comprehensive searches using the hidden Markov model (HMMsearch) against the *C. tinctorius* genome. Furthermore, the set of CtHCT sequences were re-investigated for the existence of HXXXD and functional domain of transferase using the online webserver of MARCOIL. Based on the information obtained, a total of 105 *HCTs* were identified in *C. tinctorius* genome. Among these, 23 *CtHCTs* sequences were eliminated from the analysis due to the

insufficient data and absence of the active HCT functional domains. These 82 *CtHCT* genes were distributed unevenly on different chromosomes, demonstrating the formation of different clusters. The assembly and organization of *CtHCTs* in *C. tinctorius* revealed that evolutionary events such as tandem duplication and genome repetition might participate in the origin *HCT* gene family. The 82 CtHCT-encoded amino acids were renamed from CtHCT001-CtHCT082. The length of these peptides ranged from 305aa-903aa; molecular weights ranged from 33.88kDa to 99.69kDa, with an average of 50.40kDa (Table 1). The isoelectric points (pI) values fall under 5.00 to 9.03, with an average of 6.51. In addition, the subcellular localization predictions were investigated of which, most of the candidate CtHCTs proteins were localized to the plasma membrane, cytoplasm, nucleus, and mitochondria, respectively. All CtHCTs proteins showed thermal stability due to their aliphatic matching indexes with other globular proteins.

Table 1. Physicochemical properties of CtHCT protein identified in *C. tinctorius*

Gene Name	Gene Ids	Physical position		Properties of HTC proteins			Subcellular localization
		Start position (bp)	End Position (bp)	Protein length (aa)	pI	Molecular weight (kDa)	
CtHCT001	CCG000516.1	746114	747439	441	6.79	49.84	Cytoplasmic
CtHCT002	CCG001378.2	5909	7267	452	6.33	50.12	Cytoplasmic
CtHCT003	CCG001379.1	24296	25537	413	5.71	45.97	Nuclear
CtHCT004	CCG001803.1	211037	212986	481	5.54	54.10	Cytoplasmic
CtHCT005	CCG002140.1	229582	232205	423	5.63	46.67	Lysosomal
CtHCT006	CCG002236.2	198227	202026	506	5.72	56.51	Cytoplasmic
CtHCT007	CCG002377.1	590115	591774	454	6.35	50.80	Mitochondrial
CtHCT008	CCG002378.1	600655	602022	455	5.73	50.64	Cytoplasmic
CtHCT009	CCG002482.1	863044	868400	472	5.72	52.98	Cytoplasmic
CtHCT010	CCG002510.1	97232	98599	455	6.09	50.74	PlasmaMembrane
CtHCT011	CCG003257.1	1176385	1177676	403	6.13	44.89	Chloroplast
CtHCT012	CCG003633.1	311714	313105	463	5.97	51.40	Cytoplasmic
CtHCT013	CCG003634.1	335577	336980	467	5.76	51.82	Cytoplasmic
CtHCT014	CCG003759.1	1127462	1128874	470	6.29	52.96	PlasmaMembrane
CtHCT015	CCG003761.1	1408445	1409809	454	6.17	50.55	PlasmaMembrane
CtHCT016	CCG003763.1	1440050	1441414	454	5.91	50.52	PlasmaMembrane
CtHCT017	CCG003818.1	124158	126495	349	6.63	39.07	Cytoplasmic
CtHCT018	CCG003842.1	948116	950791	434	6.09	47.93	PlasmaMembrane
CtHCT019	CCG004580.1	901107	902447	446	6.49	49.88	Mitochondrial
CtHCT020	CCG004641.1	26129	27547	472	5.79	52.90	Plasma Membrane
CtHCT021	CCG004799.1	1370991	1372205	404	8.33	45.31	Plasma Membrane
CtHCT022	CCG005260.1	27335	28777	480	7.18	53.05	Chloroplast
CtHCT023	CCG005367.1	39596	40963	455	5.64	50.54	Plasma Membrane
CtHCT024	CCG005369.1	56258	57616	452	6.02	50.30	Plasma Membrane
CtHCT025	CCG006502.1	1101963	1103285	440	8.40	48.75	Mitochondrial
CtHCT026	CCG006503.1	1110821	1112152	443	7.55	48.85	Cytoplasmic
CtHCT027	CCG006546.1	534080	536081	425	5.71	47.01	Cytoplasmic
CtHCT028	CCG007462.1	1385829	1387238	469	7.97	52.15	Chloroplast
CtHCT029	CCG007907.1	1096935	1098290	451	6.16	50.05	Nuclear
CtHCT030	CCG007909.1	1136763	1138121	452	6.20	49.86	Nuclear
CtHCT031	CCG008674.1	83325	101842	465	5.71	52.11	Plasma Membrane
CtHCT032	CCG009857.1	106340	107755	443	5.98	50.01	Plasma Membrane
CtHCT033	CCG010073.1	1973978	1975333	451	6.72	50.06	Mitochondrial
CtHCT034	CCG010534.1	2308617	2309879	420	8.59	46.96	Plasma Membrane

CtHCT035	CCG010535.1	2313735	2314982	415	5.01	45.42	Chloroplast
CtHCT036	CCG010536.1	2321994	2323241	415	7.49	46.05	Plasma Membrane
CtHCT037	CCG010896.1	682701	685374	510	8.79	56.86	Plasma Membrane
CtHCT038	CCG012015.1	207979	209313	444	5.70	49.77	Cytoplasmic
CtHCT039	CCG012113.1	536639	537988	448	6.07	49.59	Plasma Membrane
CtHCT040	CCG012708.1	105808	107166	440	8.07	49.36	Nuclear
CtHCT041	CCG012709.1	115509	116870	453	7.63	50.97	Nuclear
CtHCT042	CCG012797.1	174029	182700	903	5.95	99.69	Nuclear
CtHCT043	CCG014917.1	199763	201881	532	7.31	59.09	Chloroplast
CtHCT044	CCG015116.1	1991299	1992639	446	6.18	50.10	Cytoplasmic
CtHCT045	CCG015117.1	2000394	2001758	454	6.07	50.75	Nuclear
CtHCT046	CCG017040.1	626714	628000	428	5.69	47.59	Plasma Membrane
CtHCT047	CCG017170.1	163122	165066	460	5.62	51.65	Cytoplasmic
CtHCT048	CCG017171.1	169537	171323	460	5.86	51.43	Cytoplasmic
CtHCT049	CCG017720.1	26911	48638	345	8.49	38.87	Extracellular
CtHCT050	CCG018048.1	11064	12464	466	6.37	52.30	Plasma Membrane
CtHCT051	CCG018068.1	325127	326672	368	5.87	40.58	Cytoplasmic
CtHCT052	CCG018070.1	334783	336192	469	5.81	52.04	Cytoplasmic
CtHCT053	CCG019384.1	387400	388767	455	6.19	50.25	Plasma Membrane
CtHCT054	CCG019474.1	81878	87176	549	9.01	61.69	Mitochondrial
CtHCT055	CCG020554.1	476616	480099	462	6.61	51.58	Cytoplasmic
CtHCT056	CCG020555.1	498145	504358	461	7.10	51.02	Cytoplasmic
CtHCT057	CCG020556.1	510109	513738	461	7.57	51.28	Cytoplasmic
CtHCT058	CCG020557.1	538870	542571	462	7.57	51.19	Plasma Membrane
CtHCT059	CCG020948.1	875842	877996	556	9.03	62.40	Plasma Membrane
CtHCT060	CCG021991.1	563097	563097	411	7.69	46.50	Plasma Membrane
CtHCT061	CCG022120.1	1791506	1795303	461	6.73	51.15	Mitochondrial
CtHCT062	CCG022233.1	165977	167374	465	6.41	52.09	Plasma Membrane
CtHCT063	CCG023912.1	2497	4184	462	5.46	52.08	Cytoplasmic
CtHCT064	CCG024342.1	1196607	1197980	440	5.39	48.19	Cytoplasmic
CtHCT065	CCG024712.1	4710	6089	459	6.38	51.50	Plasma Membrane
CtHCT066	CCG026284.1	1500263	1501561	432	5.72	47.60	Chloroplast
CtHCT067	CCG026329.1	341504	342844	446	7.93	49.34	Cytoplasmic
CtHCT068	CCG026331.1	353669	354908	380	7.05	41.79	Plasma Membrane
CtHCT069	CCG026332.1	361244	362584	446	7.94	49.42	Cytoplasmic
CtHCT070	CCG026516.1	2255826	2258720	441	6.46	48.48	Plasma Membrane
CtHCT071	CCG027548.1	331399	333241	439	5.26	48.67	Cytoplasmic
CtHCT072	CCG027974.1	29637	31019	460	6.23	50.90	Plasma Membrane
CtHCT073	CCG028279.1	172673	174025	450	5.41	49.92	Cytoplasmic
CtHCT074	CCG028280.1	178840	180192	450	8.12	49.67	Cytoplasmic
CtHCT075	CCG028313.1	266033	267402	305	6.90	33.88	Chloroplast
CtHCT076	CCG028841.1	239752	245127	430	6.67	48.16	Plasma Membrane
CtHCT077	CCG029651.1	263479	264894	471	6.03	52.98	Plasma Membrane
CtHCT078	CCG029784.1	1533004	1534407	446	5.78	50.21	Cytoplasmic
CtHCT079	CCG029982.1	1158515	1159906	463	5.83	51.19	Cytoplasmic
CtHCT080	CCG030854.1	299022	304069	441	5.30	48.86	Cytoplasmic
CtHCT081	CCG031372.1	30801	32141	446	5.80	49.28	Chloroplast
CtHCT082	CCG031547.1	3808	4998	396	5.00	44.43	Plasma Membrane

Phylogenetic analysis of CtHCTs

To further elucidate the evolutionary relationship of the HCT family in *C. tinctorius*, the protein sequences combined with 259 members of HCTs from *A. thaliana*, *C. cardaunculus*, *H. annuus*, *L. sativa*, and *A. annuua* were added to the alignments. The selected species were purposely nominated for genome-wide comparison because these plant species shared a relatively high frequency of presenting the *HCT* gene family across the plant kingdom (Figure S1). A neighbour-joining phylogenetic tree was constructed using the MEGA X software package with the 1000 bootstrap method. CtHCT family were clustered into six subfamilies designated as the group I, II, III, IV, V, and VI, indicating strong conservation with other species (Figure 1). Comparative phylogenetic studies suggested that most CtHCTs were assembled into group IV with AtHCTs and AuHCTs, which are mainly associated with the function of r hydroxycinnamates transfer to shikimate during monolignol formation. Similarly, the second most CtHCT containing group was group VI, along with HaHCTs and LsHCTs in abundance, corresponding to the strong conservation of the active sites for carbonyl group and shikimate binding sites. The smallest group was designated as group II containing only one member of CtHCT protein, clustered together with only one CcHCT and two members of AnHCT, suggesting a different catalytic activity and are most likely related to function other than lignin biosynthesis in plants.

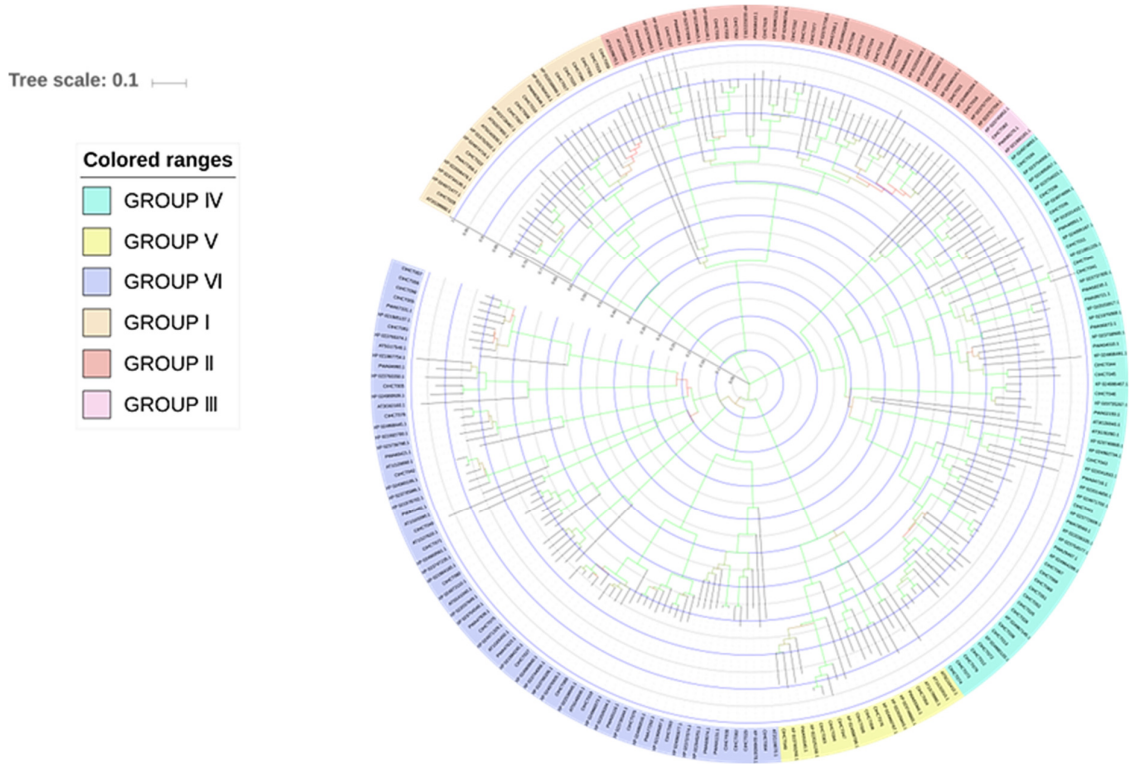


Figure 1. A neighbor-joining phylogenetic tree was generated from the full-length HCT proteins obtained from *A. thaliana*, *C. cardaunculus*, *H. annuus*, *L. sativa*, *A. annuua*, and *C. tinctorius*
The representation of nodes determines the percentage of bootstrap scores. Different background colours indicated the classification of HCTs from uncommon plant species into different subgroups.

Motifs distribution and protein alignment of CtHCTs

The conserved domains of CtHCT-encoding proteins were identified by aligning the 82 CtHCT amino acid sequences using multiple pairwise alignments. The presence of the two prevalent HCT domains in *C. tinctorius* consisting of HXXXD and DFGWG (Figure S2) were observed. The existence of high-frequency conservation of specified amino acids within the HCT domains suggested crucial hallmarks for the catalytic

activity of CtHCT in *C. tinctorius*. Furthermore, the distribution of the conserved motif was screened out with the help of MEME web server with specified settings including the classic mood, selecting the protein standard alphabet, site distribution as zero or one occurrence per sequence, and 10 numbers of motifs. The results confirmed the presence of 10 conserved motifs unevenly organized across CtHCT proteins (Figure 2). For instance, motif 1, 3, 4, 5, and 7 were found in all subgroups suggesting that these motifs were most frequently conserved in CtHCT proteins.

Moreover, from these findings, we deduced that all closely related CtHCTs proteins clustered together might represent the common composition of these conserved motifs and acquire similar activity. However, motifs 2, 6, 8, 9, and 10 were diversely distributed across all CtHCTs proteins. For example, the occurrence of motif 2 was found absent in the five members of subgroup II, motif 6 were found in all subfamilies except some members of subfamily VI, motif 8 were found nearly in all members of subfamily VI except for subfamily II and IV whereas appeared in only one member of subfamily III, motif 9 were found abundantly in VI but absent in group II, while appeared in only one member of subfamily I and III. Similarly, motif 10 was found unique to group VI and absent in all other subgroups. The full-length logos of these motifs were enlisted in (Figure S3).

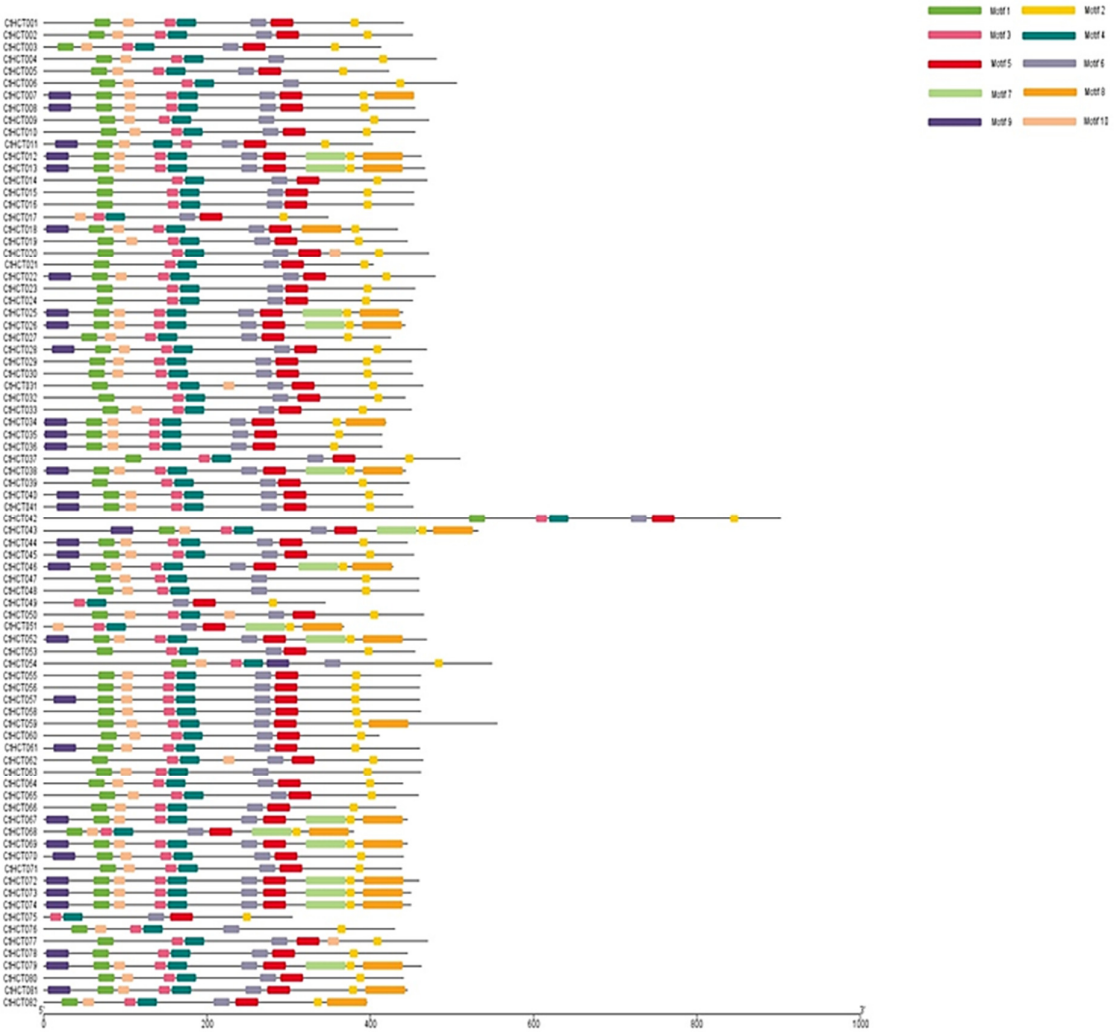


Figure 2. The distribution and organization of the conserved protein motifs found in CtHCT-encoding proteins

The analysis was carried out using the online tool of Multiple EM for Motif Elicitation (MEME version 4.8.1). Different groups of the motif were demonstrated with various colour patterns. The grey line showed the length of each sequence. Each block and colour represent the presence of different conserved motif at a specific location.

Functional protein interaction network

The interrelation of CtHCT with other proteins was investigated to link their interaction network involving different biological pathways. Identifying the protein-protein association network could play a fundamental role in predicting the possible function of the putative proteins. A total of 22 interactor proteins were predicted for the CtHCTs as shown in (Figure 3), some of them have already been determined experimentally, such as cinnamate 4-hydroxylase (C4H), p-coumarate 3-hydroxylase (C3H), 4-coumarate: coenzyme A ligase (4CL), and caffeoyl shikimate esterase (CSE). The role of these interactor proteins was most widely characterized in multiple pathways that occurred in plants during lignin and secondary metabolite biosynthesis. Furthermore, to understand the detailed topology of the interactor proteins with CtHCT proteins, the three-dimensional structures of these proteins were predicted and compared correspondingly determined by their genetically encoded amino acid sequences (Figure S4). As a result, understanding the relationship between 3D arrangements of amino acid sequence and protein structure allows us to draw a significant amount of information for functional prediction of novel protein from genome sequence data and the rational engineering of protein functions. Taken together, the protein-protein interaction network of putative CtHCTs could help us linking crucial biosynthetic pathways and routes leading to specialized metabolism in plants.

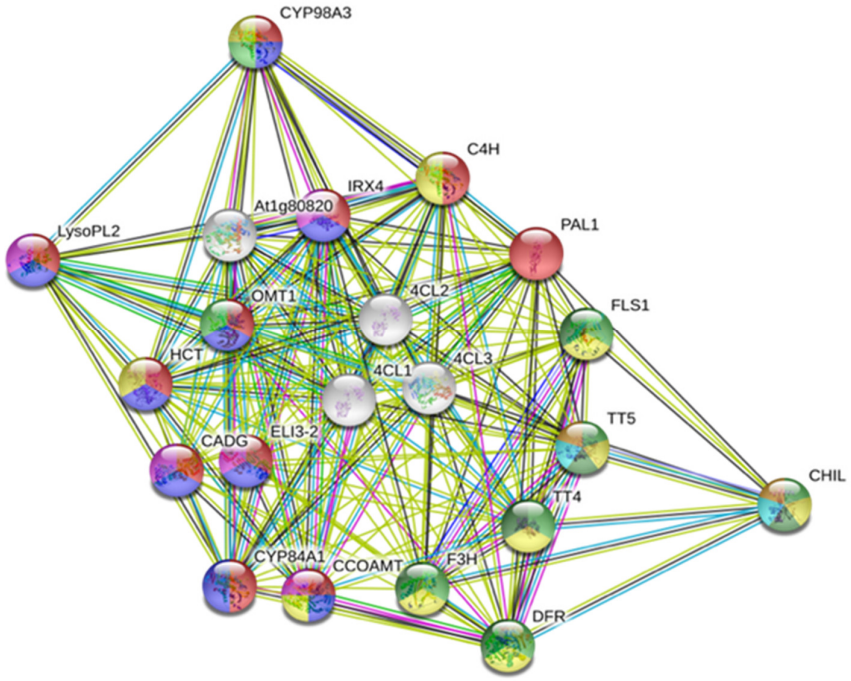


Figure 3. The predictive protein-protein interaction network of putative CtHCT-encoding proteins
The most substantial interactions were represented with red, blue, green, yellow, and grey colours. The figure was created with the online database of STRING version 10.

Gene structure organization of CtHCT genes

The investigation of the intron/exons organization of *CtHCT* genes was carried out to predict gene structure conservation and investigate their evolutionary relationships using the online GSDS tool. The average length of *CtHCT* genes ranged between 1000 bp (*CtHCT17*) to 2675 bp (*CtHCT042*). The structural organization of each *CtHCT* gene comprising exons (red), C.D.s (yellow) as well as 5` and 3` UTR regions (blue) is demonstrated in (Figure 4). These results suggested a variable trajectory in exon/intron numbers, C.D.s, and UTR regions were found even in the most closely related members of the same subgroup. For

example, *CtHCT018*, *CtHCT056*, *CtHCT057*, *CtHCT058*, and *CtHCT070* possess similar gene structures, however, the same members of group 6 contains different number of exons and introns including *CtHCT002*, *CtHCT005*, *CtHCT049*, *CtHCT064*, *CtHCT066*, and *CtHCT078*. The diversity in gene structures of *CtHCT* genes indicated multiple evolutionary mechanisms such as gene recombination, gene duplication, alternative splicing, and transposon, resulting in new gene structures and transcripts that form unique polypeptides with different biological functions.

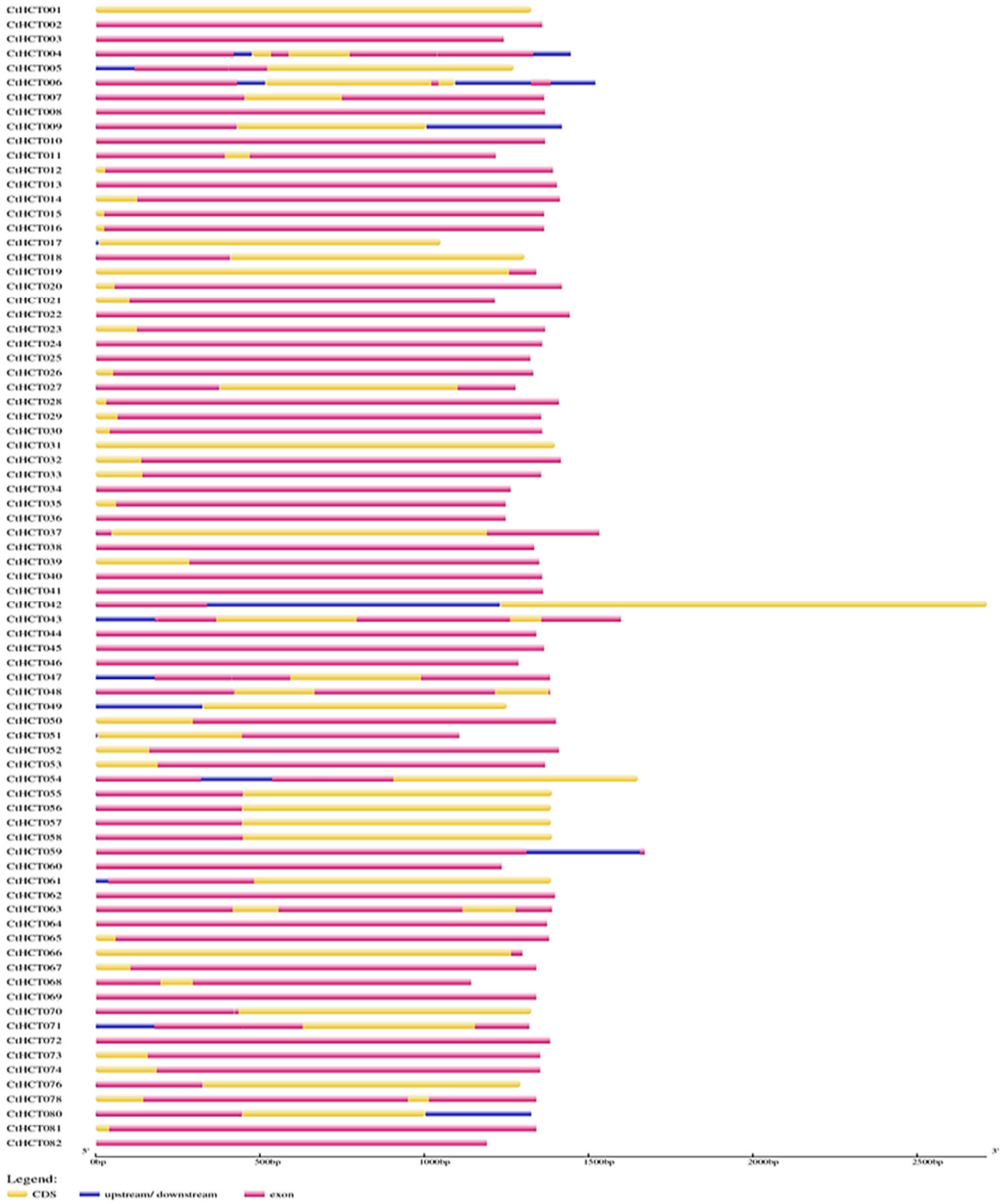


Figure 4. Gene structure organization of *CtHCT* genes
Different elements of the gene structure, including (5` and 3` UTR, Exons and CDS) of 82 putative HCT genes of *C. tinctorius* were crafted with the Gene Structure Display Server (GSDS; 2.0).

Cis-regulatory units of CtHCT genes

To explore the functional diversity and regulatory system of the *CtHCT* gene family, we investigated cis-regulatory elements within the promoter region of each gene. A total of 20 frequent cis-regulatory units were identified in the 2000 bp genomic sequences located upstream from the initiation codon or 5' untranslated region (5'UTR) of *CtHCT* genes. Most of the cis-elements found abundantly in the promoter region of *CtHCT* genes include hormonal responsive elements, particularly gibberellins, jasmonic acid, salicylic acid, abscisic acid, and auxin-responsive elements. Furthermore, diverse group of regulatory units related to tissue-specific expression were also detected, such as meristems, endosperm, root, and seed-specific regulatory units. Apart from that, various defense and abiotic stress-associated responsive elements, such as light and low temperature-responsive elements combined with cell cycle regulatory units and metabolic-related responsive units were also detected in the promoter region of *CtHCT* genes of *C. tinctorius* (Figure S5). These findings revealed the flexibility and functional diversity of *CtHCT* genes involving their potential roles in specialized metabolism and diverse biological activities in plants.

Functional annotation of CtHCT genes

The G.O. term analysis was performed to assign functional annotation to putative *CtCHCT*-encoding genes. All *CtHCTs* were divided into three functional categories, including biological processes (B.P.), molecular function (M.F.), and cellular component (CC). A bulk of *CtHCT* genes were enriched into biological processes term followed by molecular function and cellular component. The most enriched G.O. terms of biological processes contain biosynthetic, metabolic, and cellular processes which include biotic and abiotic stimuli, defense responses, cell wall organization and biogenesis, cellulose metabolic, and biosynthetic processes (Figure S6). In the molecular function term, the top-ranked G.O. terms of binding and catalytic activities were enriched which include enzyme inhibitor and regulator activity, molecular function regulator, copper ion binding, protein histidine kinase activity, carboxylic ester hydrolase activity, and pectinesterase activity. The most enriched G.O. terms such as the cell wall, external encapsulating structure, cell periphery, and respiratory chain were assigned to the cellular component category (Figure S6). These results emphasized the potential roles of *CtHCTs* in a variety of biosynthetic and crucial metabolic processes which may directly or otherwise participate in regulating plant responses against external stimuli.

HCT expression profiles at different flowering stages of C. tinctorius

The digital expression level of *CtHCT* genes was primarily calculated with FPKM (fragments per Kb per million reads) statistics of each *CtHCT* gene using the software package of featureCounts (v1.5,0-p3). The data was obtained from the whole transcriptome shotgun sequencing in four different flowering stages of *C. tinctorius*, including bud, initial, full, and fade stages. The data have been deposited in the public database of NCBI under the accession number (PRJNA399628). As described in (Figure S7), the expression level of *CtHCT* genes was divided into different cluster groups demonstrating the differential expression pattern of these transcripts at four flowering stages in *C. tinctorius*. Furthermore, to validate the biological expression level of *CtHCT* transcripts, we performed the qRT-PCR analysis of 20 genes in four different flowering stages of *C. tinctorius* (bud, initial, full, and fade). Expectedly, the expression patterns of these *CtHCT* genes across all flowering stages were found consistent with RNA-seq results. For instance, the expression trend of *CtHCT005*, *CtHCT019*, *CtHCT030*, *CtHCT055*, *CtHCT077* and *CtHCT081* was increased upwards at the initial and full phase of flower development (Figure 5). Similarly, the transcript abundance of the *CtHCT009*, *CtHCT028*, *CtHCT031* candidate genes was exhibited at the full and fade flowering phase. However, the relative transcript abundance of *CtHCT001*, *CtHCT028*, *CtHCT033*, and *CtHCT048* was detected only at the full flowering phase indicating their transcriptional regulation full bloom flowering period of *C. tinctorius.* On the contrary, the relative fold expression level of *CtHCT017*, *CtHCT027*, *CtHCT031*, *CtHCT039*, *CtHCT040* and *CtHCT059* was peaked at the fading stage of flower development (Figure 5).

Mostly, the *CtHCT* genes expression level in fading stages of flower development was four-fold higher than the rest of the three stages. The expression preference of *CtHCT* genes in different flowering phases suggested a significant correlation with flower developmental and regulation of secondary metabolism in plants.

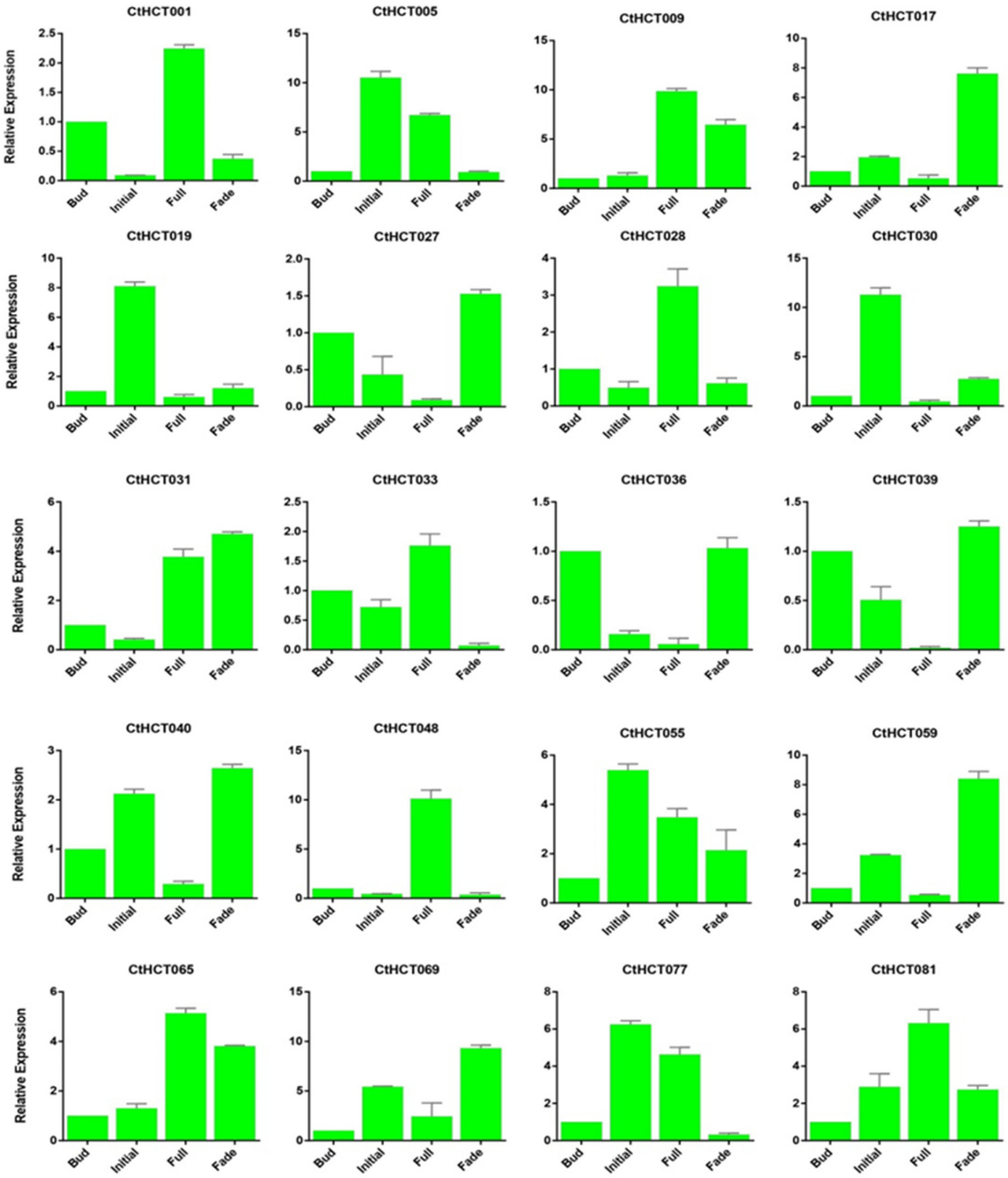


Figure 5. Expression profiling of *HCT* genes in *C. tinctorius* at four different flower developmental stages, including bud, initial, full, and fade

The 18s ribosomal RNA genes were used as a housekeeping gene in our analysis. The data was calculated using the 2-$\Delta\Delta^{CT}$ method.

Expression profiling of CtHCT genes under light irradiation

Here we analyzed the expression profiling of *CtHCT* genes at four different phases of flower development in *C. tinctorius* in response to normal and low light irradiation using the qRT-PCR assay. Following weak light irradiation (4.6 MJ/m^2), the transcription level of selected *CtHCT* genes was significantly induced through all flowering stages under investigation. However, the expression trend was found variable for each transcript than the normal light irradiation condition (16.8 MJ/m^2). Following weak light irradiation at bud flowering stage, the expression level of the most of CtHCT genes including *CtHCT017*, *CtHCT019*, *CtHCT030*, *CtHCT017*, *CtHCT039*, *CtHCT048*, *CtHCT055*, *CtHCT065*, and *CtHCT081* was significantly up-regulated indicating different folds of transcript abundance. However, *CtHCT001*, *CtHCT033*, and *CtHCT059* expression levels were down-regulated at the bud flowering stage (Figure 6).

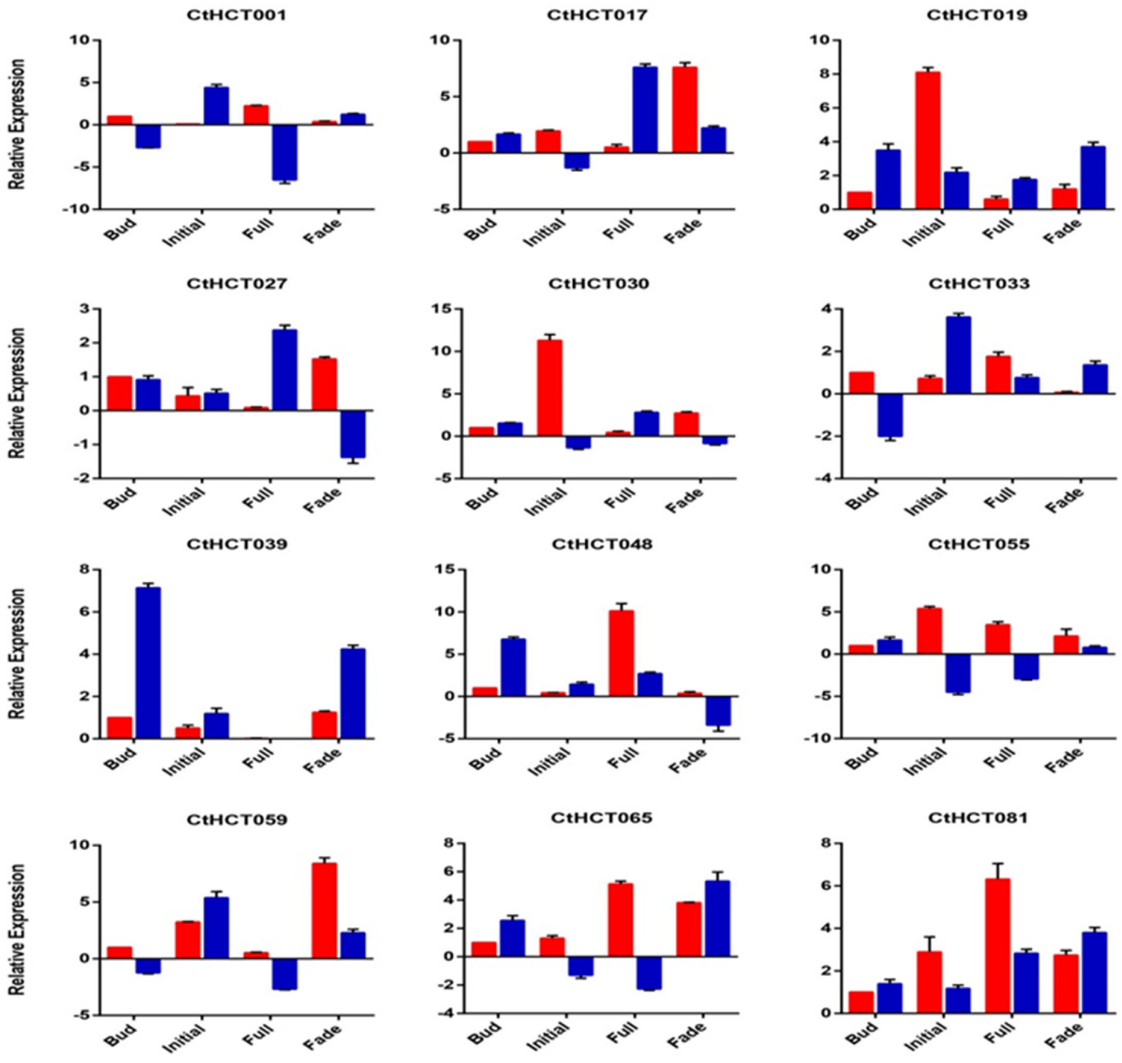


Figure 6. Expression profiling of *CtHCT* genes at four different flowering stages of *C. tinctorius* under light irradiation using qRT-PCR analysis

The red bars represent the control treatment group under normal light irradiation (16.8 MJ/m^2), whereas blue bars denote the treatment group induced with (4.6 MJ/m^2). The 18s ribosomal RNA genes were used as the housekeeping gene. The data was calculated using the 2-$\Delta\Delta^{CT}$ method.

Moreover, after the induction of weak light irradiation at the initial flowering phase of *C. tinctorius*, the transcription level of *CtHCT001*, *CtHCT033*, *CtHCT039*, *CtHCT048*, *CtHCT059* was increased up to 2-3 folds, whereas the expression level was declined in case of *CtHCT017*, *CtHCT019*, *CtHCT030*, *CtHCT055*, *CtHCT065*, and *CtHCT081* genes.Under the same light intensity at the full flowering stage, the expression level of the most *CtHCT* genes showed a downwards trend, including in *CtHCT001*, *CtHCT033*, *CtHCT048*, *CtHCT055*, *CtHCT059*, and *CtHCT065*; however, the expression level was induced upwards in *CtHCT017*, *CtHCT019*, *CtHCT027*, and *CtHCT030*. During fading stage, the expression of CtHCT genes such as *CtHCT001*, *CtHCT019*, *CtHCT033*, *CtHCT039*, *CtHCT065*, and *CtHCT081* was up-regulated under weak light induction, whereas the expression pattern was declined in case of *CtHCT017*, *CtHCT027*, *CtHCT030*, *CtHCT048*, *CtHCT055*, and *CtHCT059* (Figure 6). These findings of the of *CtHCT* genes under weak light irradiation suggested positive insights into understanding the regulation of stress responses by activating their genetic machinery in combination with other possible factors that interconnect early stress responsive mechanisms.

Expression profiling of CtHCT genes under MeJA treatment

Methyl jasmonate is an important hormone involved in regulating various growth and defense-related signaling pathways in plants. Herein, the expression level of *CtHCT* genes under the MeJA treatment was investigated using qRT-PCR analysis. Following the results of qRT-PCR, a variable expression profile of *CtHCT* genes under MeJA stress was detected than the normal condition. Among 10 *CtHCT* genes, the expression level of *CtHCT001*, *CtHCT048*, and *CtHCT081* genes respectively up-regulated up to 2 folds at bud flowering and 2-4 folds at initial flowering under MeJA stress (Figure 7). Similarly, in combination with bud and initial flowering, the transcript abundance of *CtHCT059* also suggested an increased expression trend with a slight change displaying up-regulation at the full flowering stage. In addition, the expression level of *CtHCT017*, *CtHCT039*, and *CtHCT065* following induction with MeJA rises at both the initial and full phases of flower development. In contrast, the expression was down-regulated in other flowering parts. In contrast, the transcription of *CtHCT019* and *CtHCT030* under the same treatment of MeJA demonstrated an upward expression trend at the bud and complete flowering stages. It was down-regulated at the remaining flowering stages (Figure 7). Interestingly, the expression level of *CtHCT055* was down-regulated at all flowering phases. In contrast, the transcript abundance of *CtHCT027* was observed with significantly increased expression patterns at the full and fading stages of flower development. Based on the variable expression network of *CtHCT* genes under MeJA induction at different stages of flower transitions, it is suggested that the transcription level of *CtHCTs* may play a pivotal role in facilitating the adaptation of *C. tinctorius* in fluctuating environments.

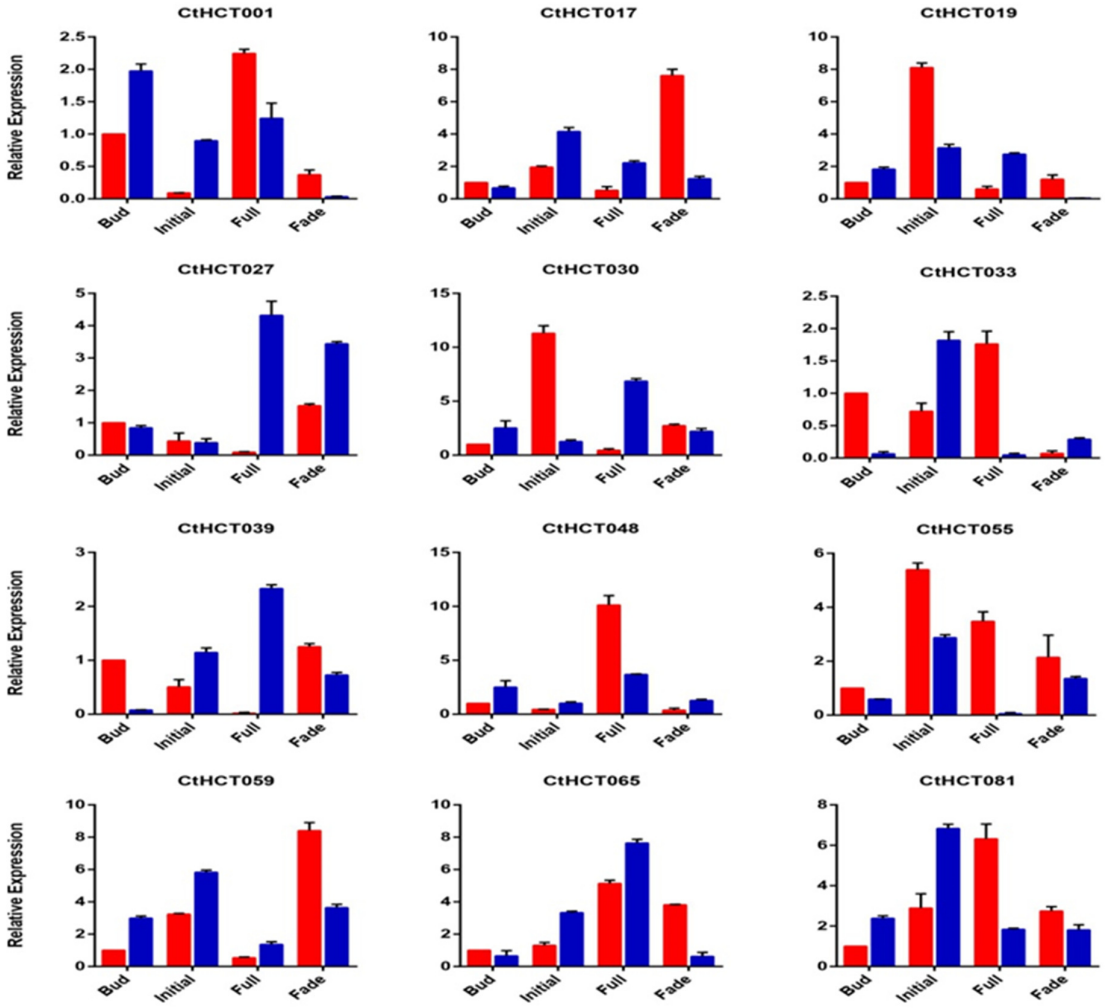


Figure 7. Expression profiling of *CtHCT* genes at four different flowering stages of *C. tinctorius* under methyl jasmonate treatment using qRT-PCR analysis

The red bars represent the control treatment group (no treatment), whereas the blue bars denote the treatment group induced with MeJA (100 µM). The 18s ribosomal RNA genes were used as a housekeeping gene. The data was calculated using the $2\text{-}\Delta\Delta^{CT}$ method.

Discussion

The *HCT* gene family is mainly involved during the regulatory mechanism of monolignol biosynthesis in different plant species (Shadle *et al.*, 2007; Sun *et al.*, 2018). Here, we extensively characterized the *HCT* gene family in *C. tinctorius* and provided a complete genome-wide overview of these genes alongside their structural and functional active sites interlinked with the regulation of abiotic stress responses in plants. In general, the assembly of the *HCT* gene family of *C. tinctorius* in comparison to other plants shared variable size in the total genome, including *Arabidopsis* (Initiative, 2000), strawberry (Shulaev *et al.*, 2011), pear (Wu *et al.*, 2013), apple (Velasco *et al.*, 2010), peach (Verde *et al.*, 2013). Further comparative analysis revealed that the occurrence of conserved amino acids at specific positions in *CtHCT* was found consistent with the *Arabidopsis* (D'Auria, 2006) and pear HCTs (Ma *et al.*, 2017). Moreover, the existence of other conserved amino acid residues indicating more than 60% similarity was consistent with *Populus nigra* (Vanholme *et al.*,

2013) and *Coffeaca nephora* (Lepelley *et al.*, 2007). These findings suggested that the high frequency of the conserved amino acid residues could be crucial for identifying putative function of *CtHCT* genes.

The investigation of the cis-elements in the promoter region of *CtHCT* genes revealed various critical regulatory units involved during the counter-response of plants against different abiotic stressors. The occurrence of these top-ranked abiotic stress-responsive elements in *CtHCT* genes includes abscisic acid, low temperatures, gibberellins, jasmonic acid, salicylic acid, auxin, defense and stress-responsive factors, and endosperm-specific expression. As found by (Dang *et al.*, 2011), these types of cis-regulatory elements were explicitly identified in pea plants against abiotic stress responses. In agreement with these findings, the results of of cis-elements of *CtHCT* genes also suggested important hallmarks involving plant adaptation to various abiotic stresses and signal transduction of hormones during plant growth and development. In addition to cis-regulatory elements, the enrichment of the *HCT*-encoding proteins in the core pathway of phenylpropanoid biosynthesis and other essential classes of secondary metabolites have been described in several plants, such as *Linum usitatissimum* (Tripathi and Agrawal, 2013), tobacco (Tamagnone *et al.*, 1998), and *Eucalyptus globules* (Shinya *et al.*, 2014). In this study, we also investigated that most of the *CtHCT*-encoding proteins were enriched in biosynthetic, metabolic, and cellular processes containing responses to biotic and abiotic stimuli, defense responses, cell wall organization and biogenesis, cellulose metabolic and biosynthetic processes, and external encapsulating structure organization. These findings strongly highlight essential clues into the putative role of *CtHCT* genes in secondary metabolism and understanding the abiotic stress resistance mechanism in *C. tinctorius*.

HCT genes played a critical role in plant growth and development-related activities. For example, the *HCT1* and *HCT2* of red clover plants were expressed in all tissues, including stems, leaves, and flowers, but was found higher in flowers than expression in unexpanded leaves, mature leaves, and stems (Sullivan, 2009). Similarly, *Populus trichocarpa* possesses seven *PtrHCTs* that can be expressed in the tissues of various plant parts and exhibit differences concerning their relative performance. In particular, *PtrHCT1* and *PtrHCT6* are primarily expressed in stem tissues, whereas *PtrHCT3* has a higher expression level in leaf tissues (Shi *et al.*, 2009). Another study revealed that the expression pattern of the *HcHCT* transcript was ubiquitous in all parts of a 4-week-old plant but was relatively high in roots and mature flowers. The highest *HcHCT* transcript was detected in young flowers and young leaves during flowering and leaf development (Chowdhury *et al.*, 2012). HcHCT showed high expression levels in flowers and roots, suggesting that HcHCT participates in the biosynthesis of secondary metabolites in floral and root tissues. Given the previous studies, the diverse expression profiling of *CtHCT* genes at four different flowering stages of *C. tinctorius* was also detected. The transcription regulation and expression preference of *CtHCT* genes at various flowering stages indicated a significant correlation of these transcripts with plant growth and development and regulation of secondary metabolism.

Furthermore, the natural exposure of plants to various abiotic and biotic stresses leads to generating several mechanisms in cell wall modification to protect themselves against these stresses. The potential of lignin to protects cell wall degradation by maintaining its polysaccharides level against pathogenic microbes and stress-induced degradation. Lignin act as an antioxidant in a plant, encountering heat stress and eventually increase plant tolerance against various stress conditions (Bhardwaj *et al.*, 2014). As a point of importance, the optimum light intensity and temperature required for germination and average growth of *C. tinctorius* are (16.8 MJ/m^2) and 35 °C (Torabi *et al.*, 2016). Fluctuation in light intensity and temperature can affect the physiological and development activities of *C. tinctorius* (Torabi *et al.*, 2016). In addition, MeJA treatment also demonstrated similar results in Norway spruce where the *HCT* genes showed induced expression level (Chowdhury *et al.*, 2012). The regulatory mechanism of abiotic stress tolerance in *C. tinctorius* by conducting the expression profiling of *CtHCT* genes under low light irradiation (4.6 MJ/m^2) and MeJA stress using qRT-PCR analysis was also examined. These results demonstrated that *CtHCT* transcription shared a diverse expression pattern through different flower development stages in *C. tinctorius* following induction with weak light irradiation and MeJA treatment. These data suggested that the up-regulation and down-regulation of *CtHCT* genes at

certain stages of flower development could be involved in the defense-related pathways in coordination with J.A. signalling pathways (Chowdhury *et al.*, 2012) that define early stress response to specific and broad-spectrum stress tolerance mechanisms. However, more efforts are still needed to elucidate the explicit functional role of *CtHCT* genes in *C. tinctorius*.

Conclusions

This study provides the first comprehensive genome-wide analysis explaining various structural and functional components of *CtHCT* genes. From these findings, it was revealed that various conserved entities such as protein motifs, cis-acting elements, gene structure, and functional enrichments could be crucial in predicting the function of *CtHCT* genes during plant growth and development. In addition, a group of *CtHCT* genes showed preferential expression in developing flowers of *C. tinctorius* under normal and abiotic stress conditions, suggesting the regulation of stress responsive mechanism in flower tissues. Together, these findings could pave the wave for the discovery of key genes involved in lignin biosynthesis and the foundation for engineering C. tinctorius with enhanced lignin content.

Authors' Contributions

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 31771868, 31501366), the Science and Technology Development Project of Jilin Province (No. 20190201172JC,20190201175JC,20150623024TC-11), Basic Agricultural Science and Technology Program of Wenzhou Science & Technology Bureau (No. N20190006), and the 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

Bhardwaj R, Handa N, Sharma R, Kaur H, Kohli S, Kumar V, Kaur P (2014). Lignins and abiotic stress: an overview. Physiological Mechanisms and Adaptation Strategies in Plants Under Changing Environment. Springer, pp 267-296. *https://doi.org/10.1007/978-1-4614-8591-9_10*

Chao N, Qi Q, Li S, Ruan B, Jiang X, Gai Y (2021). Characterization and functional analysis of the Hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl transferase (HCT) gene family in poplar. PeerJ 9:e10741.

Chiang YC, Levsh O, Lam CK, Weng JK, Wang Y (2018). Structural and dynamic basis of substrate permissiveness in hydroxycinnamoyltransferase (HCT). PLoS Computational Biology 14(10):e1006511. *https://doi.org/10.1371/journal.pcbi.1006511*

Chowdhury EM, Choi BS, Park SU, Lim HS, Bae H (2012). Transcriptional analysis of hydroxycinnamoyl transferase (HCT) in various tissues of *Hibiscus cannabinus* in response to abiotic stress conditions. Plant Omics 5(3):305.

Conesa A, Götz S (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. International Journal of Plant Genomics 2008:619832. *https://doi.org/10.1155/2008/619832*

D'Auria JC (2006). Acyltransferases in plants: a good time to be BAHD. Current Opinion in Plant Biology 9(3):331-340. *https://doi.org/10.1016/j.pbi.2006.03.016*

Dang HQ, Tran NQ, Tuteja R, Tuteja N (2011). Promoter of a salinity and cold stress-induced MCM6 DNA helicase from pea. Plant Signaling and Behavior 6(7):1006-1008. *https://doi.org/10.4161/psb.6.7.15502*

Escamilla-Treviño LL, Shen H, Hernandez T, Yin Y, Xu Y, Dixon RA (2014). Early lignin pathway enzymes and routes to chlorogenic acid in switchgrass (*Panicum virgatum* L.). Plant Molecular Biology 84(4-5):565-576. *https://doi.org/10.1007/s11103-013-0152-y*

Eudes A, Pereira JH, Yogiswara S, Wang G, Teixeira Benites V, Baidoo EE, ... Loqué D (2016). Exploiting the substrate promiscuity of hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyl transferase to reduce lignin. Plant and Cell Physiology 57(3):568-579. *https://doi.org/10.1093/pcp/pcw016*

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, ... Sangrador-Vegas A (2015). The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research 44(D1):D279-D285. *https://doi.org/10.1093/nar/gkv1344*

Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2014). GSDS 2.0: an upgraded gene features visualization server. Bioinformatics 31(8):1296-1297. *https://doi.org/10.1093/bioinformatics/btu817*

Initiative AG (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408(6814):796. *https://doi.org/10.1038/35048692*

Lepelley M, Cheminade G, Tremillon N, Simkin A, Caillet V, McCarthy J (2007). Chlorogenic acid synthesis in coffee: An analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1, and CCoAOMT1 genes during grain development in *C. canephora*. Plant Science 172(5):978-996. *https://doi.org/10.1016/j.plantsci.2007.02.004*

Livak KJ, Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. Methods 25(4):402-408. *https://doi.org/10.1006/meth.2001.1262*

Ma C, Zhang H, Li J, Tao S, Qiao X, Korban SS, Zhang S, Wu J (2017). Genome-wide analysis and characterization of molecular evolution of the HCT gene family in pear (*Pyrus bretschneideri*). Plant Systematics and Evolution 303(1):71-90. *https://doi.org/10.1007/s00606-016-1353-z*

Payyavula RS, Shakya R, Sengoda VG, Munyaneza JE, Swamy P, Navarre DA (2015). Synthesis and regulation of chlorogenic acid in potato: Rerouting phenylpropanoid flux in HQT-silenced lines. Plant Biotechnology Journal 13(4):551-564. *https://doi.org/10.1111/pbi.12280*

Roh JS, Han JY, Kim, JH, Hwang JK (2004). Inhibitory effects of active compounds isolated from safflower (*Carthamus tinctorius* L.) seeds for melanogenesis. Biological and Pharmaceutical Bulletin 27(12):1976-1978. *https://doi.org/10.1248/bpb.27.1976*

Shadle G, Chen F, Reddy MS, Jackson L, Nakashima J, Dixon RA (2007). Down-regulation of hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase in transgenic alfalfa affects lignification, development and forage quality. Phytochemistry 68(11):1521-1529.

Shi R, Sun YH, Li Q, Heber S, Sederoff R, Chiang VL (2009). Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes. Plant and Cell Physiology 51(1):144-163. *https://doi.org/10.1093/pcp/pcp175*

Shinya T, Hayashi K, Onogi S, Kawaoka A (2014). Transcript level analysis of lignin and flavonoid biosynthesis related genes in *Eucalyptus globulus*. American Journal of Plant Sciences 5(18):2764. *https://doi.org/10.4236/ajps.2014.518293*

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, ... Mane SP (2011). The genome of woodland strawberry (*Fragaria vesca*). Nature Genetics 43(2):109. *https://doi.org/10.1038/ng.740*

Sonnante G, D'Amore R, Blanco E, Pierri CL, De Palma M, Luo J, ... Martin C (2010). Novel hydroxycinnamoyl-coenzyme A quinate transferase genes from artichoke are involved in the synthesis of chlorogenic acid. Plant Physiology 153(3):1224-1238. *https://doi.org/10.1104/pp.109.150144*

Sullivan M (2009). A novel red clover hydroxycinnamoyl transferase has enzymatic activities consistent with a role in phaselic acid biosynthesis. Plant Physiology 150(4):1866-1879. *https://doi.org/10.1104/pp.109.136689*

Sun CH, Yang CY, Tzen JT (2018). Molecular identification and characterization of hydroxycinnamoyl transferase in tea plants (*Camellia sinensis* L.). International Journal of Molecular Sciences 19(12):3938. *https://doi.org/10.3390/ijms19123938*

Tamagnone L, Merida A, Parr A, Mackay S, Culianez-Macia FA, Roberts K, Martin C (1998). The AmMYB308 and AmMYB330 transcription factors from *Antirrhinum* regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. The Plant Cell 10(2):135-154. *https://doi.org/10.1105/tpc.10.2.135*

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28(10):2731-2739. *https://doi.org/10.1093/molbev/msr121*

Tohge T, Watanabe M, Hoefgen R, Fernie AR (2013). The evolution of phenylpropanoid metabolism in the green lineage. Critical Reviews In Biochemistry and Molecular Biology 48(2):123-152. *https://doi.org/10.3109/10409238.2012.758083*

Torabi B, Soltani E, Archontoulis SV, Rabii A (2016). Temperature and water potential effects on *Carthamus tinctorius* L. seed germination: measurements and modeling using hydrothermal and multiplicative approaches. Brazilian Journal of Botany 39(2):427-436. *https://doi.org/10.1007/s40415-015-0243-x*

Tripathi R, Agrawal S (2013). Evaluation of changes in lipid peroxidation, ROS production, surface structures, secondary metabolites and yield of linseed (*Linum usitatissimum* L.) under individual and combined stress of ultraviolet-B and ozone using open top chambers. Indian Journal of Biochemistry and Biophysics 50(4):318-325.

Tsai CJ, Harding SA, Tschaplinski TJ, Lindroth RL, Yuan Y (2006). Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*. New Phytologist 172(1):47-62. *https://doi.org/10.1111/j.1469-8137.2006.01798.x*

Vanholme B, Cesarino I, Goeminne G, Kim H, Marroni F, Van Acker R, ... Pinosio S (2013). Breeding with rare defective alleles (BRDA): a natural *Populus nigra* HCT mutant with modified lignin as a case study. New Phytologist 198(3):765-776. *https://doi.org/10.1111/nph.12179*

Vanholme R, De Meester B, Ralph J, Boerjan W (2019). Lignin biosynthesis and its integration into metabolism. Current Opinion in Biotechnology 56:230-239. *https://doi.org/10.1016/j.copbio.2019.02.018*

Varbanova M, Porter K, Lu F, Ralph J, Hammerschmidt R, Jones AD, Day B (2011). Molecular and biochemical basis for stress-induced accumulation of free and bound p-coumaraldehyde in cucumber. Plant Physiology 157(3):1056-1066. *https://doi.org/10.1104/pp.111.184358*

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, ... Pruss D (2010). The genome of the domesticated apple (*Malus* × *domestica* Borkh.). Nature Genetics 42(10):833. *https://doi.org/10.1038/ng.654*

Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, ... Cattonaro F (2013). The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. Nature Genetics 45(5):487. *https://doi.org/10.1038/ng.2586*

Wagner A, Ralph J, Akiyama T, Flint H, Phillips L, Torr K, Nanayakkara B, Te Kiri L (2007). Exploring lignification in conifers by silencing hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyltransferase in *Pinus radiata*. Proceedings of the National Academy of Sciences 104(28):11856-11861. *https://doi.org/10.1073/pnas.0701428104*

Wang GF, He Y, Strauch R, Olukolu BA, Nielsen D, Li X, Balint-Kurti PJ (2015). Maize homologs of hydroxycinnamoyltransferase, a key enzyme in lignin biosynthesis, bind the nucleotide binding leucine-rich repeat Rp1 proteins to modulate the defense response. Plant Physiology 169(3):2230-2243.

Weng JK, Chapple C (2010). The origin and evolution of lignin biosynthesis. New Phytologist 187(2):273-285. *https://doi.org/10.1111/j.1469-8137.2010.03327.x*

Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan MA, Tao S, Korban SS, Wang H (2013). The genome of the pear (*Pyrus bretschneideri* Rehd.). Genome Research 23(2):396-408. *https://doi.org/10.1101/gr.144311.112*

Xu Z, Zhang D, Hu J, Zhou X, Ye X, Reichel KL, ... Gao P (2009). Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. BMC Bioinformatics 10(11):1-15. *https://doi.org/10.1186/1471-2105-10-S11-S3*

Ye SY, Gao WY (2008). Hydroxysafflor yellow A protects neuron against hypoxia injury and suppresses inflammatory responses following focal ischemia reperfusion in rats. Archives of Pharmacal Research 31(8):1010-1015. *https://doi.org/10.1007/s12272-001-1261-y*

Zhang Y, Guo J, Dong H, Zhao X, Zhou L, Li X, Liu J, Niu Y (2011). Hydroxysafflor yellow A protects against chronic carbon tetrachloride-induced liver fibrosis. European Journal of Pharmacology 660(2-3):438-444. *https://doi.org/10.1016/j.ejphar.2011.04.015*

Genome-wide investigation of *Hydroxycinnamoyl CoA: Shikimate Hydroxycinnamoyl Transferase* (*HCT*) gene family in *Carthamus tinctorius L.*

Supplementary files

Table S1. Primers set used in qRT-PCR analyses

Primer	Strand	Primer sequence (5'-3')
CtHCT001	Forward	CGGAACATCTTTCTGGCATT
	Reverse	ATTGTCATTTTCCCGTCGTC
CtHCT005	Forward	AGCTCACCATTTCCCCAATA
	Reverse	AAAGGCTGAAGATGGGGTTT
CtHCT009	Forward	TTCGAGGTCCATCCTCAAAC
	Reverse	ACAGTCCTTCCTCGTTGGTG
CtHCT017	Forward	CAGCCATCCAAGGATACGAT
	Reverse	CCTCGTCGTCTTCAAGATCC
CtHCT019	Forward	TGAGCTTAACGACGGGATCT
	Reverse	CTTTCCCCAAACCAAACTCA
CtHCT027	Forward	CTTATTCCGGCGGTTGATTA
	Reverse	GGCTGCAGTTCCAAGAAATC
CtHCT028	Forward	ACCCTCCCCAATCTCAAACT
	Reverse	TCGGAGACGGTAATTTGGTC
CtHCT030	Forward	TCCCCGATCTCGATACTTTG
	Reverse	TATTTTGCTCGCGATGTGTC
CtHCT031	Forward	CGATCCACCAGGTTTTCTTC
	Reverse	TGTTGGGTCCGTAACCATTT
CtHCT033	Forward	CCAAAGAGCAAGGACGTTGT
	Reverse	CCGGTAAAAGGCAAACTTCA
CtHCT036	Forward	GTTTCTTGTGGAAGGGGTCA
	Reverse	CCGAAGTCCATTCCGTAAAA
CtHCT039	Forward	ACTTCTCAGGGATGCCACAA
	Reverse	ACTCTCGGCACCTTTCAAGA
CtHCT040	Forward	CAGGCTTAGACCCGACCATA
	Reverse	CCTTCCGAATACGTCTCTGC
CtHCT048	Forward	AAATGTGGAGGCATGGTGAT
	Reverse	TCACCGACGTTTCTTTCTCC
CtHCT055	Forward	CAAGACTCCCTTGACCTTCG
	Reverse	GACCTATGGGCCATGTCATC
CtHCT059	Forward	ACCGTTGTTATCCGTTCAGG
	Reverse	CTCCCGTACATGTCGAACCT
CtHCT065	Forward	CTTCCCTTCCCATCTTGACA
	Reverse	GTTTCAACCTTGGCGTGTTT
CtHCT069	Forward	AGTTTCGAGCGGTCTTGTGT
	Reverse	CCATGGTTCCTTTGGAGCTA
CtHCT077	Forward	GGATCTCACCACCACCAAAC
	Reverse	GTCTGGCTAACCAGCAGAGG
CtHCT081	Forward	TGACGAATGTCCGAGTGAAC
	Reverse	CCCCATCTTACACCAACTCG

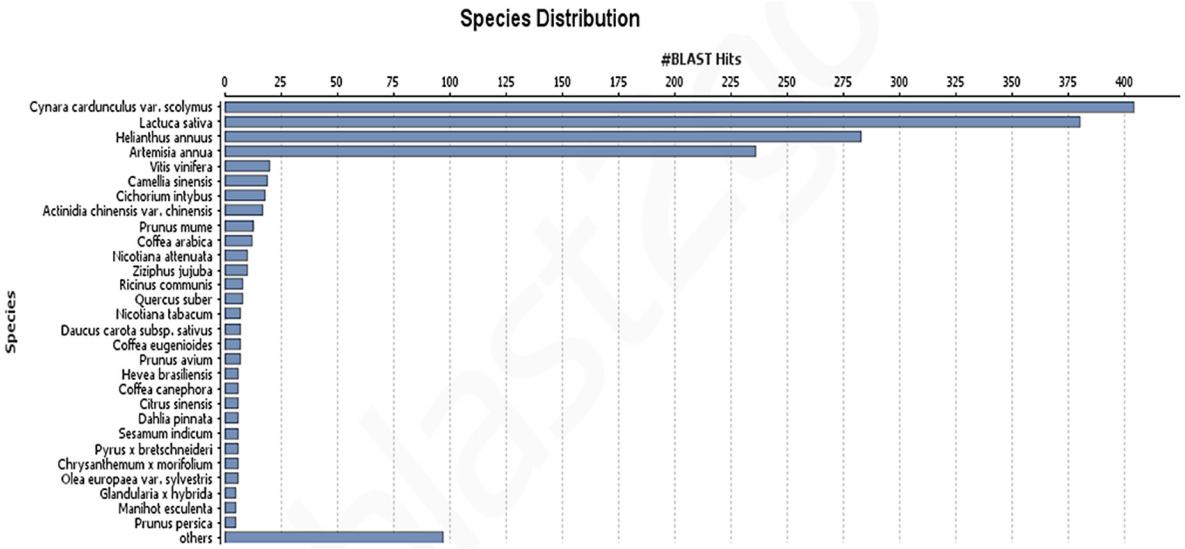


Figure S1. Distribution of HCT gene family shared by different plant species

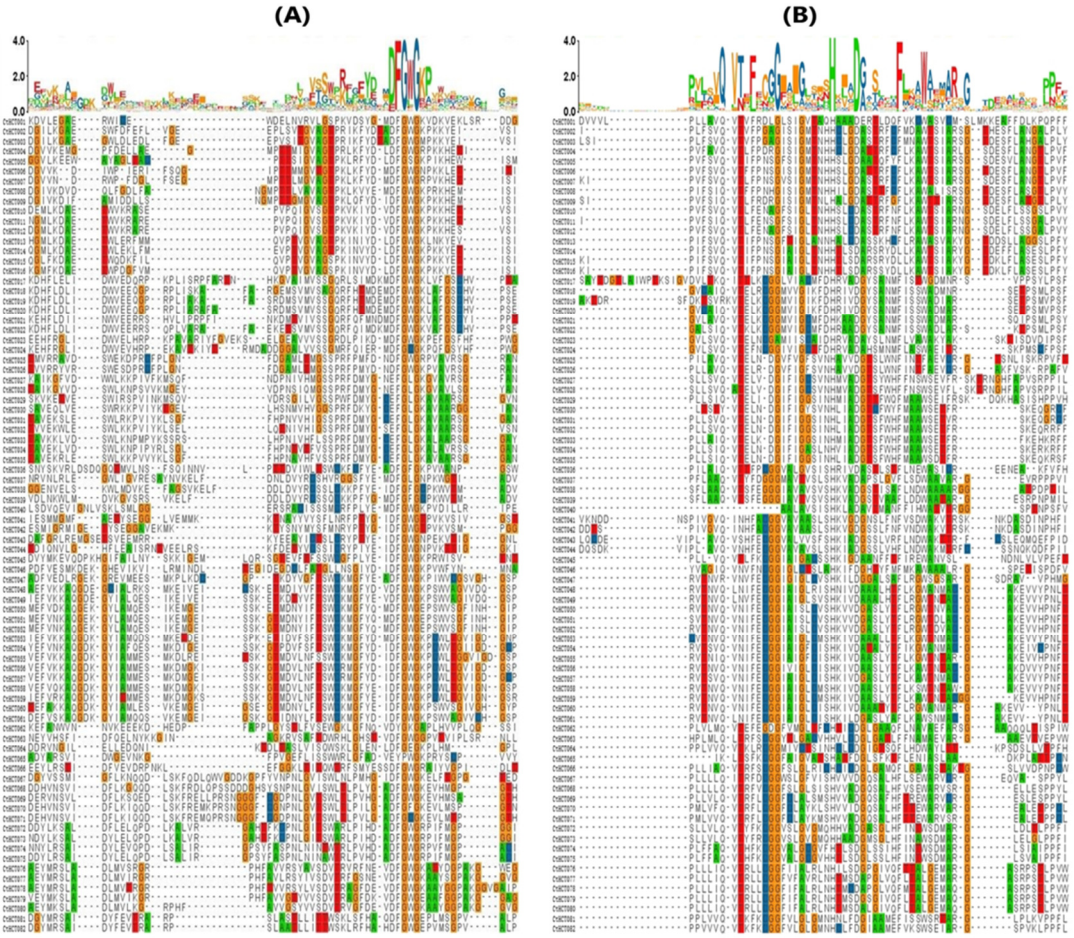
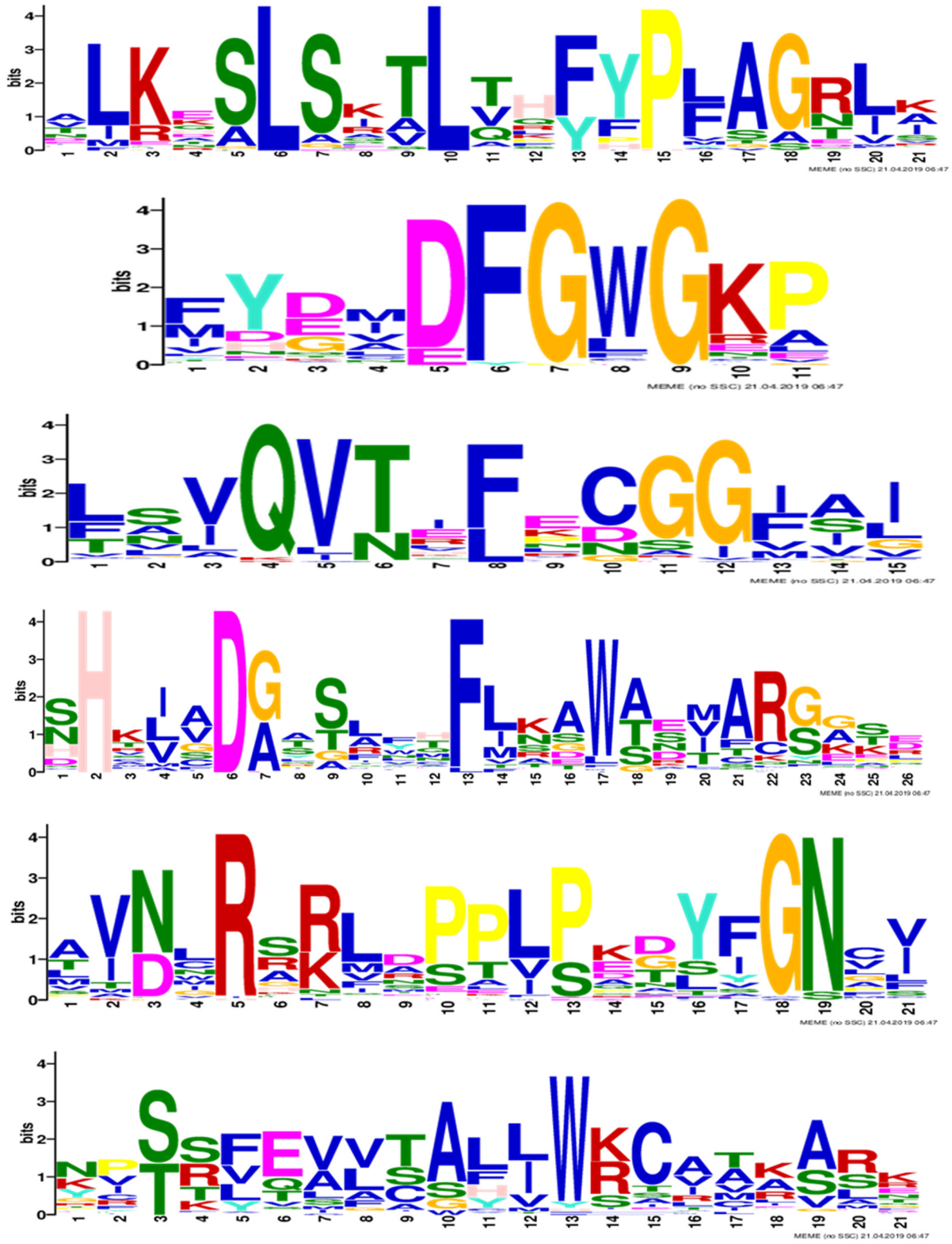


Figure S2. Multiple sequence alignment of the 82 CtHCT proteins

(A-B) The Logos of the two widespread conserved HCT domains of *Carthamus tinctorius* consisting HXXXD and DFGWG amino acids and the pairwise alignment of Carthamus tinctorius HCT proteins. The different colored boxes represent 60% similarity of the conserved amino acid residues


bits
MEME (no SSC) 21.04.2019 06:47

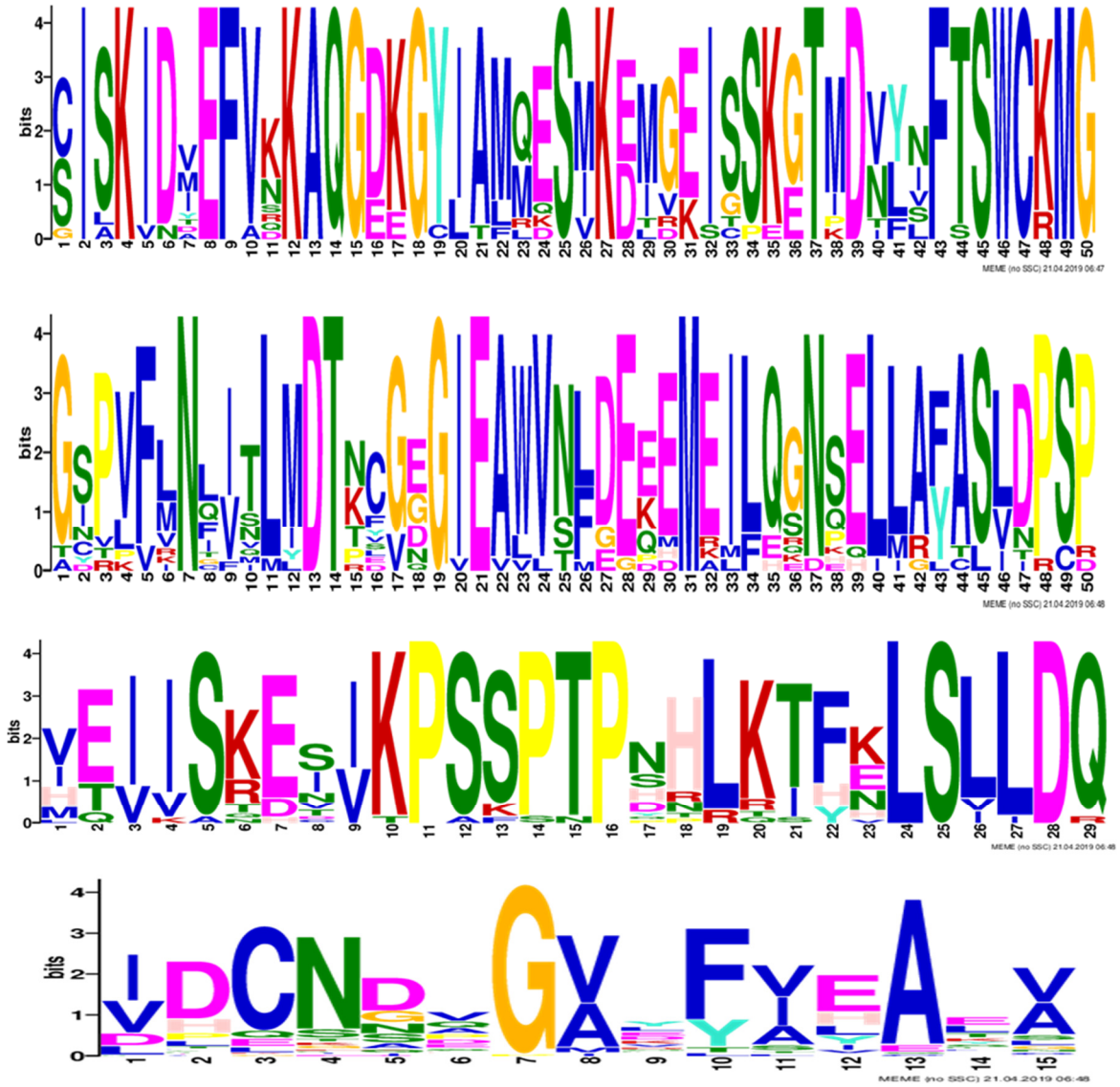


Figure S3. Logos of the 10 conserved protein motifs within CtHCT proteins obtained from the MEME online web-server

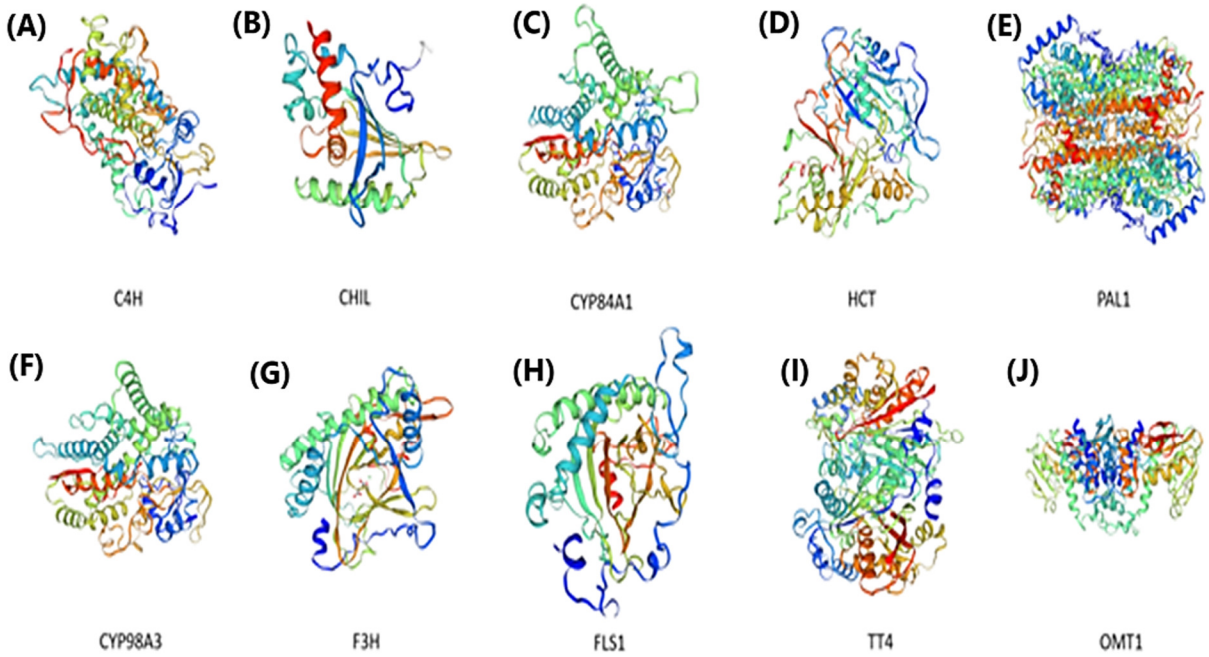


Figure S4. The three-dimensional structures of the ten putative interactor-proteins with a member of CtHCT putative protein

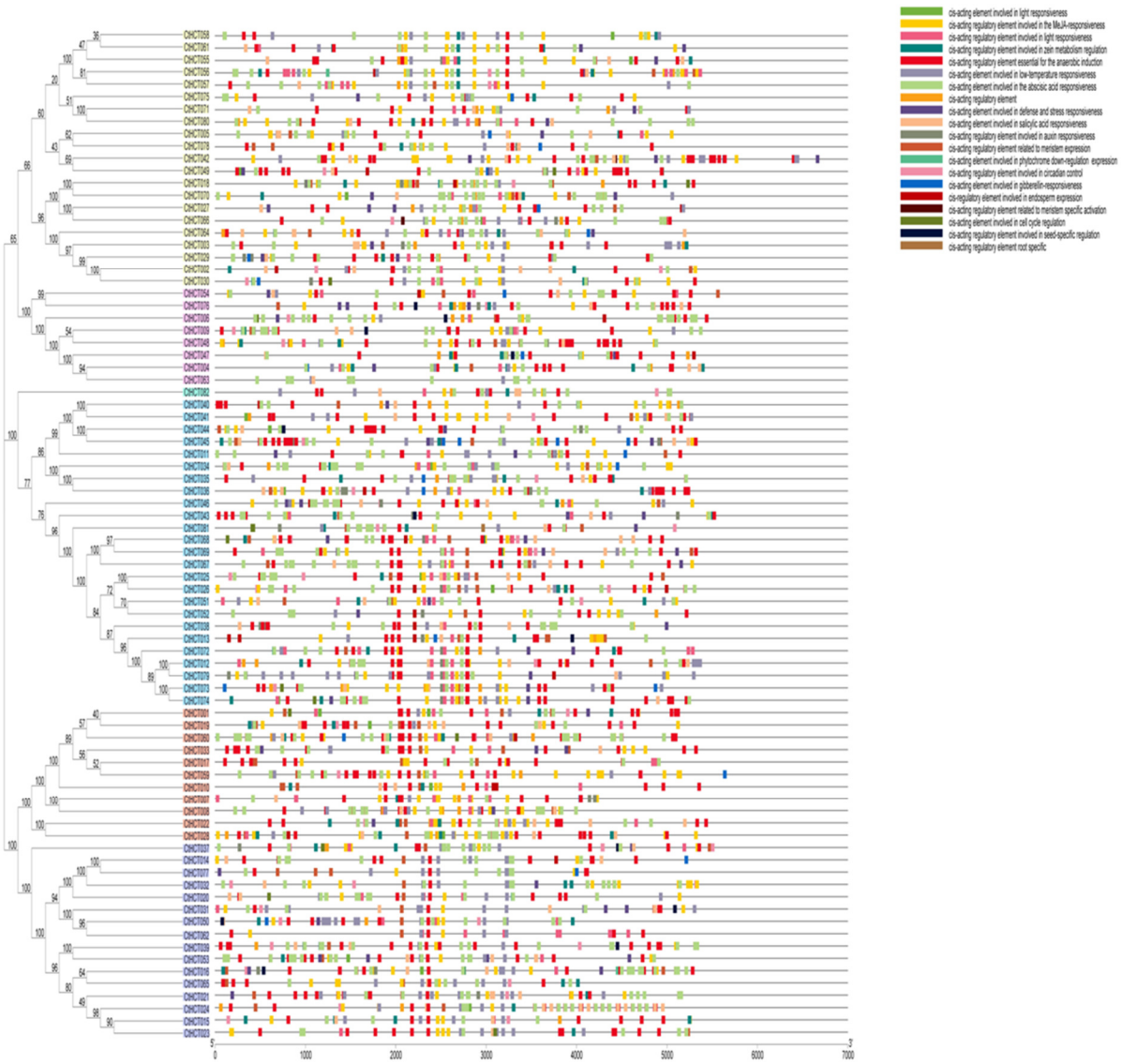


Figure S5. The graphical representation of the twenty conserved cis-regulatory elements identified in the 5' untranslated region (5' UTR) of the promoter sequence of CtHCT genes

Each colour represents a different type of cis-acting element within the promotor site of *CtHCT* genes.

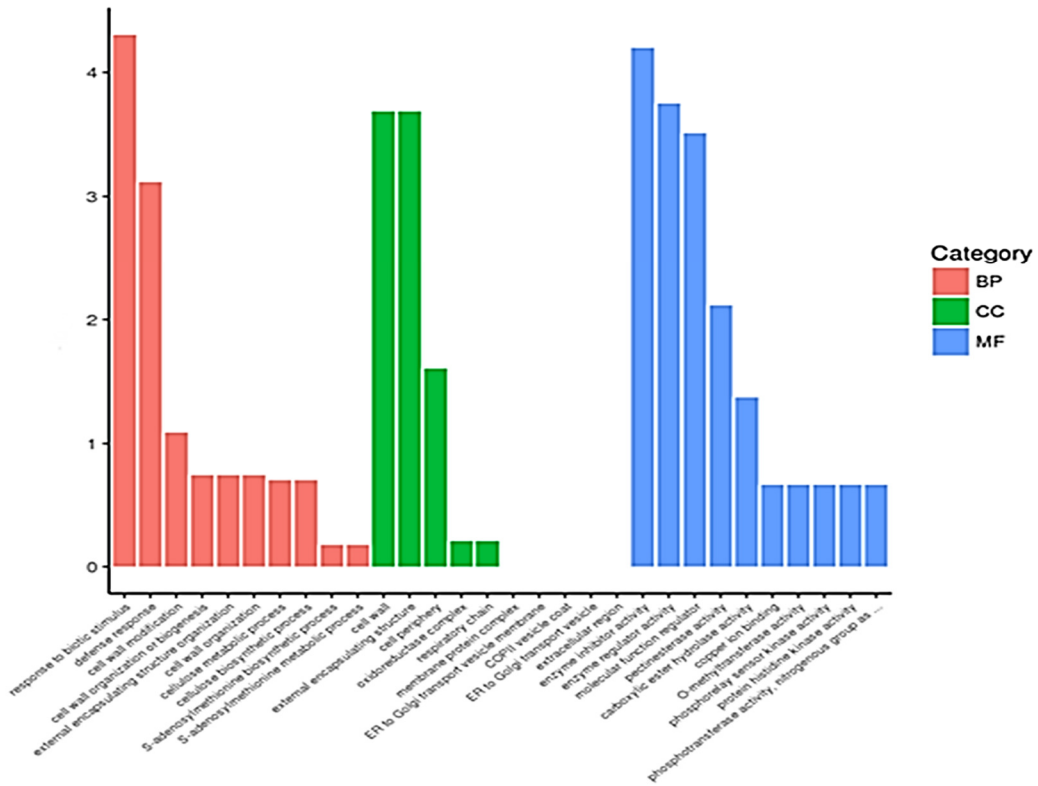


Figure S6. The Go term classification of *Carthamus tinctorius* HCTs

The annotation was investigated across three main categories: (a) biological process, (b) cellular component and (c) molecular function. The coloured boxes indicate pathways enrichment (in blue) and cellular component (yellow) with per cent score.

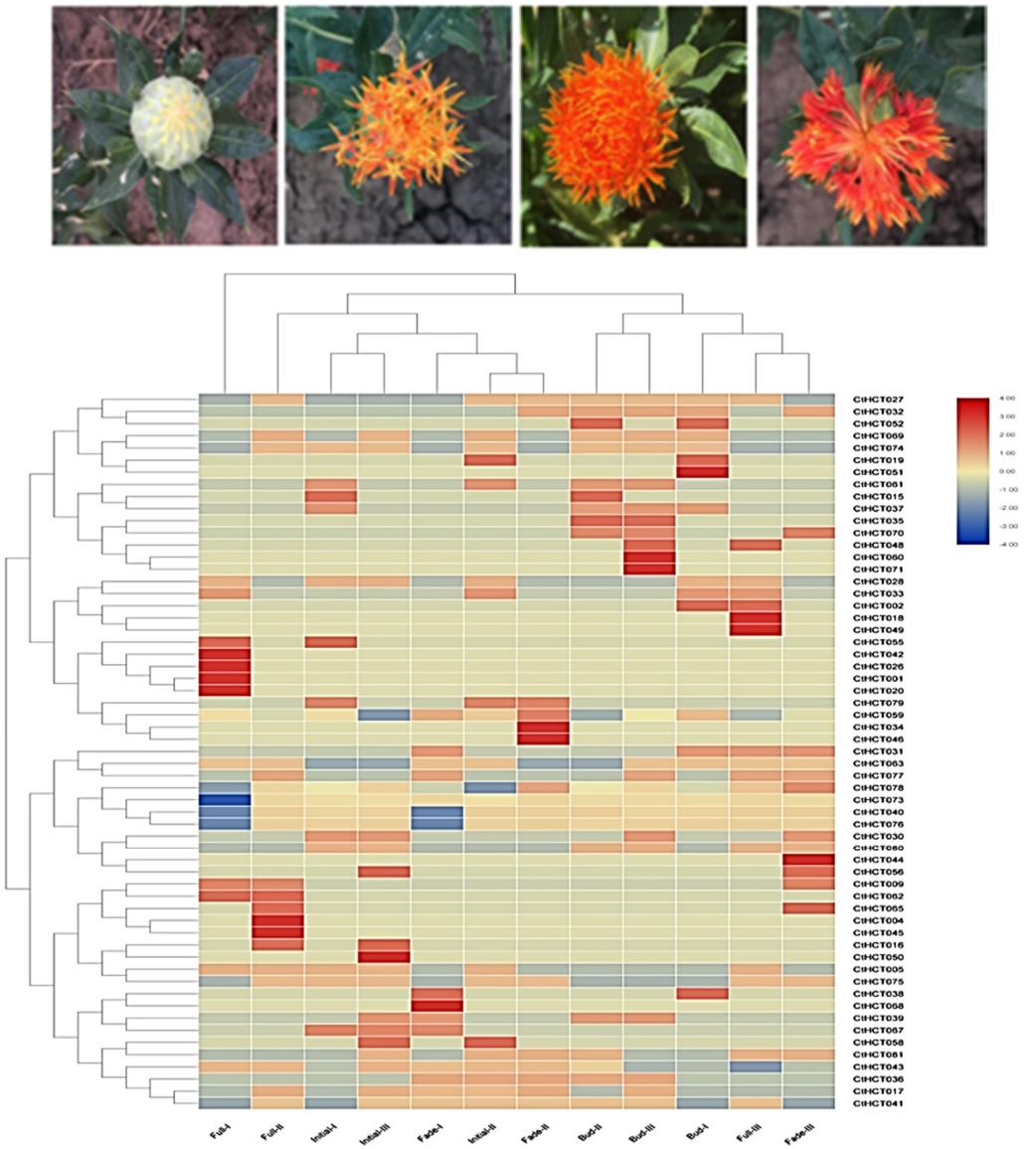


Figure S7. The four different flower developmental stages of *Carthamus tinctorius* and the expression profile of CtHCT genes

The heatmap was generated from the FPKM data obtained from RNA-seq data.