



Article

Genome-Wide Identification, Expression Analysis, and Subcellular Localization of *Carthamus tinctorius* bHLH Transcription Factors

Yingqi Hong^{1,2}, Naveed Ahmad^{1,2}, Yuanyuan Tian^{1,2}, Jianyu Liu^{1,2}, Liyan Wang^{1,2}, Gang Wang^{1,2}, Xiuming Liu¹, Yuanyuan Dong¹, Fawei Wang¹, Weican Liu¹, Xiaowei Li¹, Xu Zhao³, Na Yao^{1,*} and Haiyan Li^{1,*}

¹ Ministry of Education Engineering Research Center of Bioreactor and Pharmaceutical Development, Jilin Agricultural University, Changchun 130118, China; hyq985676490@163.com (Y.H.); naveedjlau@gmail.com (N.A.); yuanyuantian@163.com (Y.T.); ljiy20180436@163.com (J.L.); wly19940427@163.com (L.W.); 13842847523@163.com (G.W.); xiuming1211@163.com (X.L.); yydong@aliyun.com (Y.D.); 13596076186@163.com (F.W.); liuweican602@163.com (W.L.); xiaoweili1206@163.com (X.L.)

² College of Life Sciences, Jilin Agricultural University, Changchun 130118, China

³ Jilin Institute for Food Control, Changchun 130033, China; plnto@163.com

* Correspondence: yaona1103@126.com (N.Y.); hyl99@163.com (H.L.); Tel.: +0431-84532885 (H.L.)

Received: 24 April 2019; Accepted: 18 June 2019; Published: 21 June 2019



Abstract: The basic helix–loop–helix (bHLH) family is the second largest superfamily of transcription factors that belongs to all three eukaryotic kingdoms. The key function of this superfamily is the regulation of growth and developmental mechanisms in plants. However, the bHLH gene family in *Carthamus tinctorius* has not yet been studied. Here, we identified 41 bHLH genes in *Carthamus tinctorius* that were classified into 23 subgroups. Further, we conducted a phylogenetic analysis and identified 10 conserved protein motifs found in the safflower bHLH family. We comprehensively analyzed a group of bHLH genes that could be associated with flavonoid biosynthesis in safflower by gene expression analysis, gene ontology annotation, protein interaction network prediction, subcellular localization of the candidate CtbHLH40 gene, and real-time quantitative expression analysis. This study provides genome-wide identification of the genes related to biochemical and physiological processes in safflower.

Keywords: bHLH family; *Carthamus tinctorius*; gene expression; flavonoid biosynthesis; subcellular localization

1. Introduction

Transcription factors (TFs) are substantial key elements of various physiological and biochemical pathways in higher plant development. Regulation of genes related to growth, stress response of plants, and other developmental processes is mainly regulated by TF activation and repression [1,2]. The second largest TF family across eukaryotic kingdoms is the basic helix–loop–helix (bHLH) superfamily. Based on the highly conserved bHLH domain, the bHLH superfamily is categorized into two main groups: the basic region and the HLH region [3,4]. The basic region is a DNA-binding interface located next to the N-terminus of the protein domain, whereas the HLH region serves as the dimerization domain [1] and consists of two amphipathic α -helices linked by a loop. Apart from the high frequency of bHLH conserved domains, some other motifs also occur within the bHLH superfamily [5,6]. Based on the highly conserved domains and their evolutionary relationships, this superfamily is mainly divided into 20–25 subfamilies in agronomically important plants [6,7]. A wide range of extensive

studies have been reported on the bHLH superfamily-related genes in plant species after genome sequences have been produced, including *Arabidopsis* [8], peanut [9], blueberry [10], tomato [11], Chinese cabbage [12], and apple [13]. Additionally, functional and structural characterizations of most plant bHLH proteins have been described in detail. These findings demonstrated that bHLH proteins are involved in various biochemical and physiological networks across the plant kingdom. Studies on bHLH revealed its partial involvement in iron uptake [14], response against salt and drought stress [13,15], tanshinone biosynthesis [16], and petal growth development. The most prominent studies on the bHLH family demonstrated its association with anthocyanin biosynthesis and regulation in flowering plants and fruits [17].

Carthamus tinctorius L. is the most extensively used flowering plant of family *Asteraceae* and is known for its medicinal value in West Asia and China. It is a rich resource of conjugated linoleic acid (CLA) and an essential herbal medicine with less side effects [18]. Safflower oil has been reported to minimize the risk of cardiovascular diseases and regulate cholesterol [19,20]. The biologically active ingredients of safflower petals are attributed to flavonoids that harbor diverse pharmacological effects for cancer, inflammation, hypertension, prevention of excessive oxidative stress, and blood flow regulation [21,22]. Of the 5000 types of plant flavonoids that have been discovered in a variety of plants to date, safflower shares a striking variety of important secondary metabolites, including quinochalcone glycoside, hydroxysafflor yellow A, safflor yellow B, and chalcone glycosides [23,24].

Safflower has a diploid genome that has not yet been sequenced; therefore, there is a lack of information on and understanding of the molecular mechanisms, functional diversity, and structural genomics in this important plant. Therefore, it is necessary to distinguish functional genes, gene expression factors, TFs, promoters, and chromosomal allocations. Our previous study [25] revealed a set of candidate genes in safflower that are potentially involved in the flavonoid metabolism. There are currently no comprehensive reports available on the safflower bHLH superfamily genes. In the present study, we aimed to investigate the comprehensive genome wide identification of the *Carthamus tinctorius* bHLH transcription factors family followed by spatiotemporal expression analysis and GFP tagging of the candidate genes. The objectives of this work were to: (1) extensively characterize the physicochemical properties of 41 safflower bHLH genes; (2) conduct a phylogenetic analysis of these genes and identify bHLH protein motifs; (3) conduct gene expression analysis of the selected 41 bHLH genes at different flowering stages and in different tissues; (4) produce expression profiles of the bHLH gene family using quantitative real-time PCR (qRT-PCR) analyses at various developmental stages; and (5) infer subcellular localization of a candidate gene, CmbHLH40, via GFP fusion protein assay. Our study provides an extensive genome-wide investigation and expression analysis of the safflower bHLH superfamily, which might have a crucial role in functional characterization of TFs involved in flavonoid metabolism.

2. Results and Discussion

2.1. Genome-Wide Identification and Phylogenetic Analysis of Safflower bHLH Genes

To identify bHLH Genes in safflower, an in silico search from safflower genome was performed based on keywords (*Carthamus tinctorius* L.) and a Pfam ID search of safflower genome database, identification of homologous *Arabidopsis* bHLH amino acid sequence, and a Pfam domain search of the conserved domain database (CDD). Based on the information available on genome sequencing, total of 89 safflower bHLH genes were obtained by sequencing. Some of them had short sequences and no full-length sequence and some have long sequences. After eliminating these invalid genes, the remaining genes were compared by BLAST, and the genes with very low homology ratio were removed. Some of them did not have the typical domain of bHLH. Finally, 41 genes were obtained. These 41 genes were divided into 15 groups (groups 1–15) (Table 1) based on conserved domains and sequence similarity based on phylogenetic analysis. The largest group was group 9, which contained seven members, whereas the smallest groups were groups 3, 5, 6, 12, 13, and 15, with only one member.

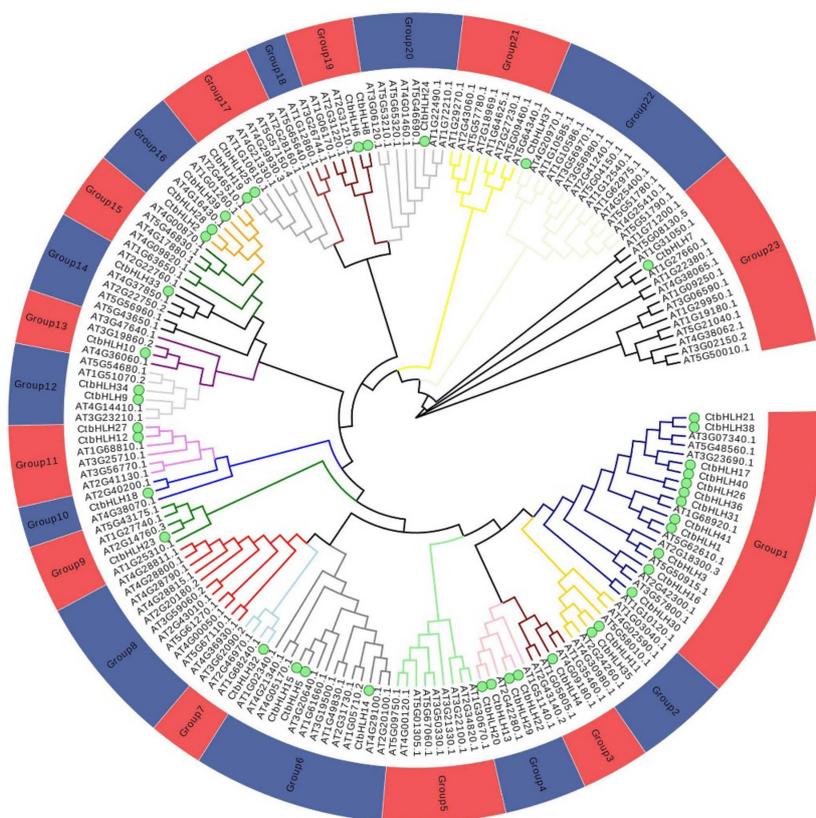
Table 1. Complete list of safflower bHLH protein family genes and their physiochemical properties. Information provided includes gene names, gene IDs, theoretical subcellular localization, coding sequence (CDS), protein length, molecular weight (MW; Da), and theoretical isoelectric points (PI). These data were obtained from ExPASy (available online: <http://web.expasy.org/protparam/>).

Gene Name	Gene ID	Protein Length	PI	MW (Da)	Subcellular Localization	Instability Index	<i>Arabidopsis</i> Homology	GRAVY
CtbHLH01(4)	CCG019410.2	271	7.84	29436.75	Nuclear	59.80	AT5G62610.1	-0.732
CtbHLH02(8)	CCG019793.1	488	6.85	54381.40	Nuclear	51.53	AT4G16430.1	-0.542
CtbHLH03(2)	CCG020831.1	314	7.78	35586.93	Nuclear	46.88	AT1G30670.1	-0.770
CtbHLH04(14)	CCG021763.1	264	6.66	29051.21	Nuclear	43.29	AT4G09180.1	-0.272
CtbHLH05(11)	CCG022142.1	424	5.63	46461.98	Nuclear	61.08	AT3G20640.1	-0.770
CtbHLH06(9)	CCG023511.1	396	5.54	45156.48	Nuclear	43.12	AT2G31210.1	-0.620
CtbHLH07(7)	CCG028074.4	480	6.73	53210.85	Nuclear	48.98	AT1G27660.1	-0.476
CtbHLH08(9)	CCG028576.1	426	5.62	47924.37	Nuclear	37.73	AT1G06170.1	-0.654
CtbHLH09(7)	CCG029033.2	420	9.34	48353.34	Nuclear	51.47	AT3G23210.1	-0.712
CtbHLH10	CCG029306.2	294	5.93	32779.18	Nuclear	67.22	AT3G19860.2	-1.017
CtbHLH11(12)	CCG031433.1	373	5.88	38966.22	Nuclear	56.54	AT2G24260.1	-0.424
CtbHLH12(9)	CCG031433.1	313	6.64	34357.55	Nuclear	57.98	AT1G68810.1	-0.587
CtbHLH13(14)	CCG003415.1	398	5.61	44597.03	Nuclear	61.12	AT2G42280.1	-0.826
CtbHLH14(7)	CCG003696.1	362	6.53	39563.23	Nuclear	67.71	AT1G51140.1	-0.690
CtbHLH15(11)	CCG004828.1	363	6.33	39459.94	Nuclear	57.93	AT1G61660.1	-0.602
CtbHLH16(5)	CCG006936.1	320	5.21	35518.73	Nuclear	60.39	AT3G57800.1	-0.585
CtbHLH17(1)	CCG009097.1	513	5.61	56099.86	Nuclear	42.63	AT3G07340.1	-0.677
CtbHLH18(9)	CCG010966.1	249	9.15	27766.58	Nuclear	31.23	AT1G68810.1	-0.453
CtbHLH19(10)	CCG015095.1	437	7.81	49777.63	Nuclear	34.05	AT1G10610.1	-0.390
CtbHLH20(14)	CCG016487.1	388	5.95	43102.83	Nuclear	58.48	AT2G42280.1	-0.726
CtbHLH21(1)	CCG030113.1	474	6.34	51930.85	Nuclear	51.00	AT3G07340.1	-0.720
CtbHLH22(14)	CCG000235.1	401	6.07	44439.58	Nuclear	69.25	AT1G51140.1	-0.913
CtbHLH23(7)	CCG000365.1	283	5.59	31696.50	Nuclear	52.80	AT2G14760.3	-0.654
CtbHLH24(9)	CCG001914.1	312	7.05	34294.77	Nuclear	64.73	AT1G72210.1	-0.500
CtbHLH25(10)	CCG002853.1	414	7.96	46510.11	Nuclear	41.81	AT1G10610.1	-0.434
CtbHLH26(2)	CCG003971.1	353	6.47	39567.82	Nuclear	40.90	AT1G10120.1	-0.587
CtbHLH27(9)	CCG006284.1	375	6.37	42123.27	Nuclear	69.08	AT1G68810.1	-0.696
CtbHLH28(8)	CCG006712.1	488	6.85	54381.40	Nuclear	52.40	AT4G16430.1	-0.548
CtbHLH29(14)	CCG008781.1	428	6.63	48008.56	Nuclear	61.50	AT2G42280.1	-0.997
CtbHLH30(2)	CCG009709.1	364	8.13	40547.38	Nuclear	43.90	AT3G23690.1	-0.736
CtbHLH31(3)	CCG010368.1	544	5.42	58619.71	Nuclear	57.98	AT1G68920.1	-0.712
CtbHLH32(6)	CCG011082.1	706	5.95	77127.87	Nuclear	58.02	AT2G20180.2	-0.576
CtbHLH33(9)	CCG011764.1	336	7.72	37338.69	Nuclear	47.03	AT4G37850.1	-0.424
CtbHLH34(7)	CCG013712.2	250	5.64	27865.55	Nuclear	59.41	AT5G54680.1	-0.614
CtbHLH35(13)	CCG014470.1	382	5.35	39191.40	Nuclear	53.49	AT4G30980.1	-0.375
CtbHLH36(2)	CCG017266.1	376	5.46	42211.90	Nuclear	49.28	AT1G68920.1	-0.779
CtbHLH37(15)	CCG018780.1	186	7.03	21187.15	Nuclear	44.33	AT4G20970.1	-0.465
CtbHLH38(1)	CCG018923.1	460	6.68	50715.86	Nuclear	56.68	AT3G07340.1	-0.676
CtbHLH39(8)	CCG018976.1	617	9.04	69705.26	Nuclear	53.97	AT2G46510.1	-0.497
CtbHLH40(1)	CCG019036.1	519	6.31	57019.01	Nuclear	44.67	AT3G07340.1	-0.682
CtbHLH41(4)	CCG019190.1	298	5.67	32673.94	Nuclear	64.80	AT3G57800.1	-0.933
CtbHLH01(4)	CCG019410.2	271	7.84	29436.75	Nuclear	59.80	AT5G62610.1	-0.732

The physical and chemical parameters of most proteins in the same group were similar according to an analysis using the ProtParam online tool. The 41 CtbHLH-encoded amino acids ranged from 186–706; molecular weights ranged from 21.19 kDa (CtbHLH37) to 77.13 kDa (CtbHLH32), with an average of 43.38 kDa. The isoelectric point (pI) values ranged from 5.21 (CtbHLH16) to 9.34 (CtbHLH7), with an average of 6.64. A grand average of hydropathicity (GRAVY) index analysis showed that most of the CtbHLH proteins were hydrophilic. The most stable protein was CtbHLH18, which had a stability index of 31.23.

2.2. Phylogenetic Analysis

To subdivide the *Carthamus tinctorius* bHLH proteins into subfamilies and elucidate the evolutionary relationships between the *Carthamus tinctorius* bHLH proteins and those of *Arabidopsis*, multiple bioinformatics tools were used to construct a phylogenetic tree that consisted of 138 *Arabidopsis* bHLH proteins and 41 bHLH *Carthamus tinctorius* proteins (Figure 1). The aforementioned bHLH proteins clustered into 18 subfamilies, which is consistent with the tree topology and previous classification of the bHLH taxonomic group in genus *Arabidopsis* [3]. The 18 subfamilies were designated 1, 2, 3, 4, 6, 7, 9, 10, 11, 12, 13, 14, 16, 17, 19, 20, 22, and 23 in Figure 1. These subfamilies were consistent with those recently reported in some other plant species, including peach [26], tomato [11], Chinese cabbage [12], peanut [9], and wheat [27]. These findings revealed that bHLH proteins within the reported subfamilies that are present in dissimilar plant species could play a basic role in plant development and evolution. In addition, our findings were found consistent with the previous subfamily classification of the bHLH domain in other plants, where 15–25 subfamilies were reported [6]. However, taxonomic classification of the subfamilies has varied. For example, one group renamed the bHLH subfamilies using the English alphabet [4,13], and another group named them using Arabic numerals [3,28] however, most previous reports used the same pattern as Roman numerals [6,9–12,29]. Therefore, to classify the *Carthamus tinctorius* bHLH subfamilies, we adopted the Roman numeral-based taxonomic names in our analysis, following [3] and [6]. None of the *Carthamus tinctorius* bHLH proteins were sorted into subfamilies 5, 8, 15, 18, and 21, which indicated loss of those proteins throughout *Carthamus tinctorius* evolution. Overall, these results revealed that the absence of the non-conserved bHLH subfamilies in *Carthamus tinctorius* could have explicitly evolved to enhance organic processes of the plants or improve stress tolerance.



2.3. Motifs and Multiple Sequence Alignment of bHLH Proteins

To combine the phylogenetic tree parameters and thus the selection of the same taxonomic members, a tree was created using the 41 *Carthamus tinctorius* bHLH protein sequences (Figure 2). Although the order of the germline subfamily is slightly different from the order of the tree presented above (Figure 1), each member of the subfamily shared the same composition and pattern as found in the neighbor-joining tree of *Carthamus tinctorius* bHLH proteins. The conserved protein motifs of bHLH proteins were screened using the online tool, MEME version 4.8.1, by uploading 41 protein sequences of the *Carthamus tinctorius* bHLH superfamily according to [30]. Approximately all sequences were found to exhibit two types of highly conserved protein motifs with a similar width of 38 amino acids, which are demonstrated as red and green blocks, respectively (Figure 2). Furthermore, the locations of the motifs were investigated, and the majority of these motifs were present at the C-terminus of *Carthamus tinctorius* bHLH subfamilies 1, 3, 6, 8, and 14. Subfamily 9 shows the presence of some motifs adjacent to the N-terminus. In most of the cases, excluding a few bHLH proteins in safflower, these two conserved domains are located next to each other; however, some bHLH protein motifs were identified that were located far from each other. The maximum distance between the two conserved motifs was found in CtbHLH9. These findings are consistent with the previous findings on the bHLH domains in tomato [11] and wheat [27]. In this study, motif1 was indicated by red blocks and determined to be composed from basic residues and loop helix 1. However, motif2 was represented with green blocks and is composed of a loop and helix number 2. The variable loop length determines the gap between motif1 and motif2 in the bHLH protein superfamily. Motif 1 (in red) and motif 2 (in green) are composed of logos 1 and 2, respectively (Figure 2). Several reports have suggested conservation of motif1 and motif2 at the amino acid level in a variety of plants, leading to dimerization [11,27,31]. Moreover, with the help of multiple sequence alignment, we analyzed the various sequential features of 41 *Carthamus tinctorius* bHLH proteins (Figure 3). The occurrence of highly conserved motifs among same subfamily confirms the previous phylogenetic representation obtained from the proteins of bHLH domain. The corresponding position of each of these bHLH protein motifs were found conserved however, motif 1 and motif 2 was found dispersed among all subgroups. Motif 3 was present in all types of subgroups except for subfamily 7 and 8. Motif 6 was present in several proteins of subgroups 7, 8, 11, 12 and 13. The presence of motif 7 was found among the proteins of safflower bHLH from subgroup 1 and 3. Motif 8 and 9 were screened only in the proteins of 8, 9 and 10 subgroups of safflower bHLH superfamily.

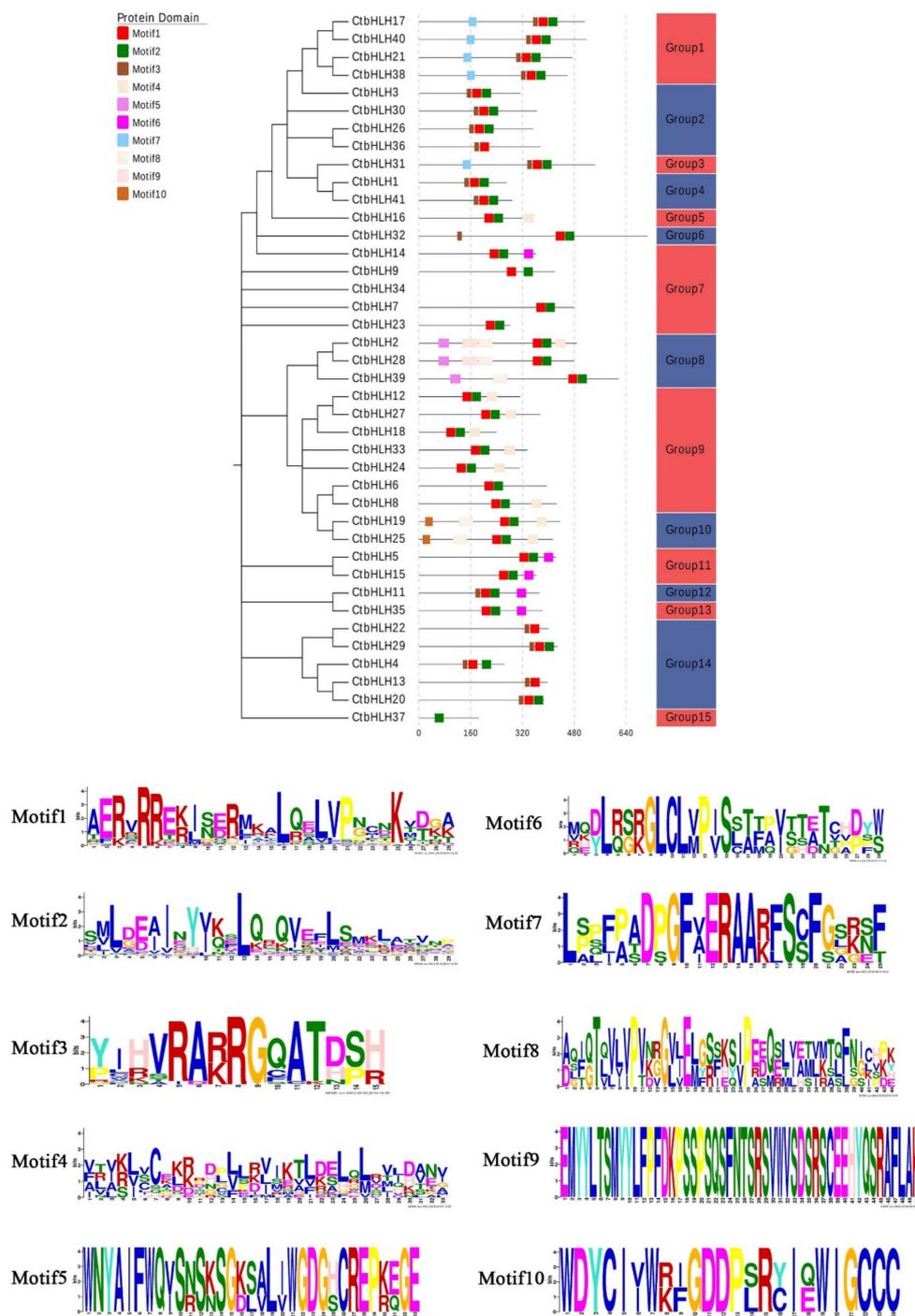


Figure 2. Motif distribution of safflower bHLH superfamily proteins. The phylogenetic tree was generated using the specific sequences of bHLH proteins obtained from *Carthamus tinctorius*. The online server of MEME server was used to identify the conserved motifs of *Carthamus tinctorius* bHLH superfamily genes. A total of 10 motifs were screened and visualized with different colors. The length of each grey line represents the relative length of the sequence. The occurrence of each block at different positions represents the location of the conserved motif to its matching one.

CtbHLH1	EEMKIKGDFVEGCN.....SVIGKLVDEIIN	195
CtbHLH2	HVIGGLNSQSIVS.....ALDQSNQDSSMLD	329
CtbHLH3	EEMKIKGDFVEGCD.....SVIGKLVDEIIN	202
CtbHLH4	DIRKKGDFVENMD.....QSNTHDMDFEAND	190
CtbHLH5	DRITAKQDFSPFG.....KTDTSVWEHATE	347
CtbHLH6	GRFEAKDFVENFA.....KADRASLNGYATE	239
CtbHLH7	DRIAAKQDFVAFPG.....KTDTSVWEHATE	399
CtbHLH8	DRYDAKDFVENFT.....KADRASVGDATK	260
CtbHLH9	DRFLEKDFSTDEGGE.....ANTDRIATGDAIR	310
CtbHLH10	SRQGSRRKSSEQGE.....SGSGKDVDSNDVA	230
CtbHLH11	ERMKIKGDFVENAN.....KTDTSVWEHATE	228
CtbHLH12	NHLAKRSLFSTIT.....KTDKQSLAEVIQ	171
CtbHLH13	ERMKIKGDFVENMD.....QSNTHDMDFEAND	382
CtbHLH14	DRITAKQDFSPFG.....KTDTSVWEHATE	255
CtbHLH15	DRVTAQDFSPFG.....KTDTSVWEHATE	284
CtbHLH16	ARMKIKGDFVEGCN.....KISGTMVDEIIN	200
CtbHLH17	ERMKIKGDFVEGCN.....SVIGKLVDEIIN	407
CtbHLH18	AHLHTRTVECTD.....KMDKATLAEVIR	121
CtbHLH19	DGLYTRALVEKIS.....KMDKATLAEVIR	287
CtbHLH20	ERMKIKGDFVENMD.....QSNTHDMDFEAND	363
CtbHLH21	ERMKIKGDFVEGCN.....SVIGKLVDEIIN	355
CtbHLH22	ERIKRKGDFENMD.....QSNTHDMDFEAND	381
CtbHLH23	ELRLIKGDFVNGT.....KVDISTMDEAVQ	243
CtbHLH24	EIEVTNVESSHANLK.....ILSKKRQQLKINA	251
CtbHLH25	DGLFTRSLVEKIS.....KMDRTRINGDATE	263
CtbHLH26	ERMKIKGDFVEGCN.....KITGKLVDEIIN	210
CtbHLH27	NHLAKRSLFSTIT.....KTDKQSLAEVIQ	229
CtbHLH28	HVIGGLNSQSIVS.....ALDQSNQDSSMLD	329
CtbHLH29	DMKKIKGDFETMD.....KQSTHDMDFEAND	396
CtbHLH30	ERMKIKGDFVEGCN.....KITGKLVDEIIN	224
CtbHLH31	ERMKIKGDFVEGCS.....SVIGKLVDEIIN	389
CtbHLH32	EMMRKKGDFIENCN.....KVDKASMDDEATE	459
CtbHLH33	GRFIAKSAVVEGLK.....KMDKASVGDATK	197
CtbHLH34	QQLKACCGPPTAFY.....PFAHSVMVPPFAV	212
CtbHLH35	ERMKIKGDFVENAN.....KTDTSVWEHATE	230
CtbHLH36	ERMKIKGDFVEGCN.....KITGKLVDEIIN	226
CtbHLH37	VRETGSSLENLILIT.....GVDFKFMETETIR	133
CtbHLH38	ERMKIKGDFVENCN.....SVIGKLVDEIIN	370
CtbHLH39	IDFTGALSRFSPVISRFMSGDSEISDVEASGRDERAAVSDD	439
CtbHLH40	ERMKIKGDFVEGCN.....SVIGKLVDEIIN	386
CtbHLH41	ERMKIKGDFVEGCN.....SVIGKLVDEIIN	224
Consensus		

Figure 3. Conserved amino acid residues and pairwise alignment of the safflower bHLH proteins. The colored boxes indicate 50% identity of amino acids.

Our findings revealed the presence of two major conserved domains in bHLH proteins, including the basic region with helix1 and the loop region with helix2. The results were confirmed in MEME output, which also indicated that there were two highly conserved domains (Figure 3). These results were consistent with the result of the long distance between the two motifs in CtbHLH9, as described in Figure 2. Generally, the sequences of the basic region and the position of the two helix domains were more conserved compared with the sequences obtained from the loop region. The MEME and sequence alignment results further demonstrated that our previous hidden Markov model (HMM) findings showed high reliability. The results also showed that HMM can be highly precise for predicting a given superfamily.

2.4. Gene Expression Analysis

Based on the RNA-seq transcriptome data (PRJNA399628; submitted on august 23rd, 2017), the RPKMs of different tissues (roots, shoots, and leaves) and flower developmental stages (F0, F2, F4, F6, F8, F10, and F12 petals) were used to examine the gene expression profiles of the bHLH superfamily in safflower. The heat map shows that these CtbHLH genes clustered into two groups. In general, genes within the same subgroup showed similar expression patterns (Figure S1). The first group was highly expressed in the petals. The second group was highly expressed in the roots. The various expression patterns indicated functional divergence of different groups of CtbHLH members in safflower.

To further study the relationship between CtbHLHs and petal development, we isolated 19 genes with RPKM ≥ 5 at F0, F2, F4, F6, F8, F10, and F12 for further qRT-PCR analysis. Of these, four genes were highly expressed at F0 (Figure 4A), and three genes were highly expressed at F2 (Figure 4B), five genes were highly expressed at F4 (Figure 4C), and four genes were highly expressed at F8 (Figure 4D). In addition, three genes were highly expressed at F12 (Figure 4E). The qRT-PCR results showed that the fold-change values of CtbHLHs were reliable.

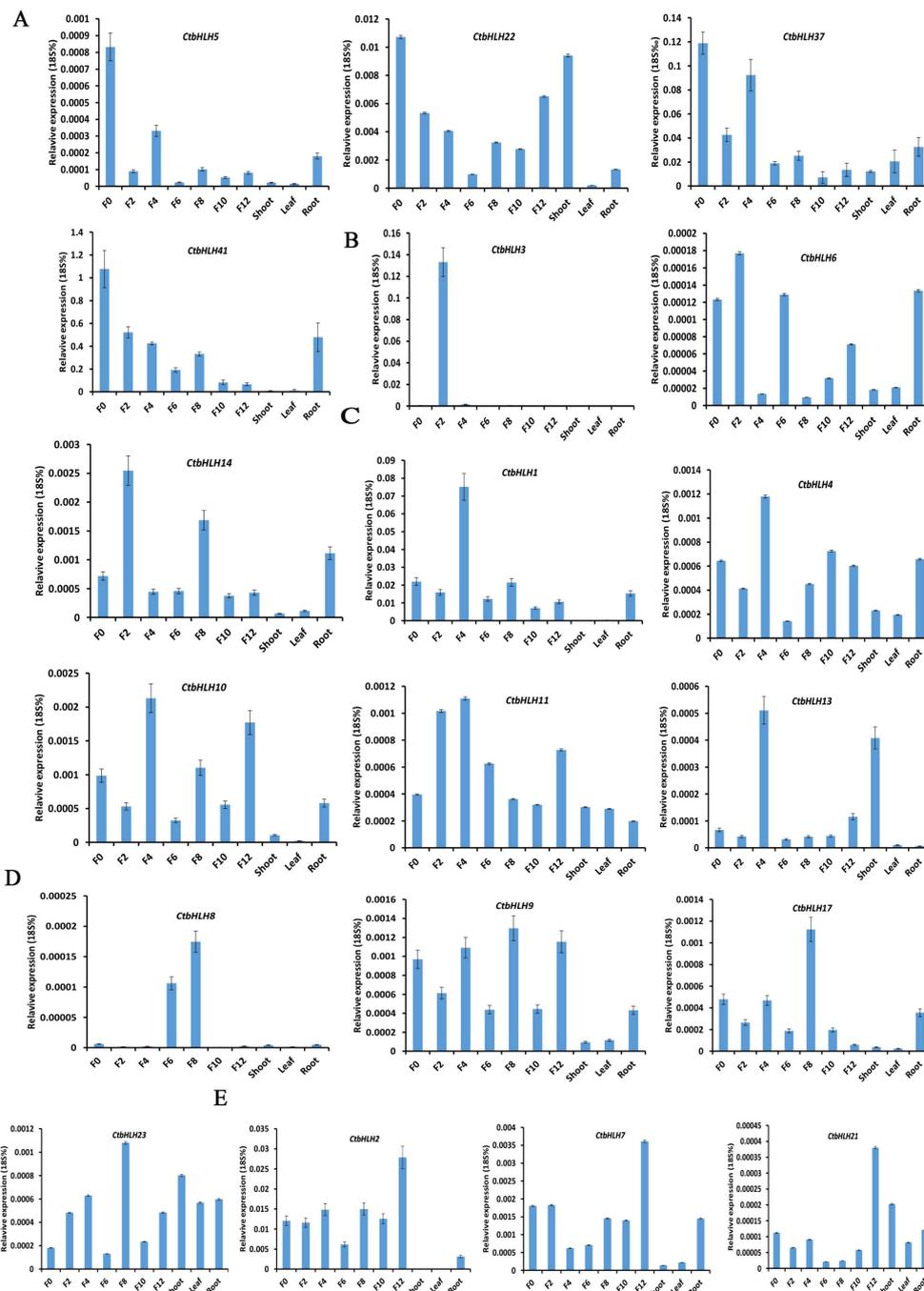


Figure 4. Expression profiles of 19 safflower bHLH genes during different developmental stages as determined by qRT-PCR analysis. **(A)** Includes genes that were highly expressed at F0. **(B)** Includes genes that were highly expressed at F2. **(C)** Temporal expression of genes at F4. **(D)** and **(E)** Transcript abundance of genes found at F8 and F12, respectively. Expression analysis was normalized using the 18S rRNA gene as an internal reference. Error bars were estimated from the differences in the expression patterns of three independent replicates.

2.5. Gene Ontology Annotations

The extremely deviating sequences present outside of the bHLH conserved domain suggested that bHLH proteins are actively participating in a variety of physiological processes. A large number of the *Arabidopsis* bHLH proteins were comprehensively investigated by [32–37]. Therefore, we demonstrated the biological processes associated with the safflower CtbHLH genes. For this purpose, we have used

the *Arabidopsis* as the model species to perform the GO annotation of the safflower bHLH proteins. As shown in Figure 5 the majority of the safflower CtbHLH proteins were found DNA binding. Most of these CtbHLH proteins were localized in the nucleus. However, there were also CtbHLH proteins located on membrane. The biosynthetic process as well as nuclear-based metabolic consequences were linked with the majority of CtbHLH proteins. Furthermore, a moderate number of CtbHLH proteins might respond to various abiotic stresses. The phenomenon of flower development, post-embryonic development and cell differentiation involves ~25–45 CtbHLH proteins. Thus, based on these findings, the multi functional role of CtbHLH proteins could be associated with various biosynthetic and metabolic processes, in response to abiotic and biotic stresses, in the development of various tissues and organs, in signal transduction and in the cell.

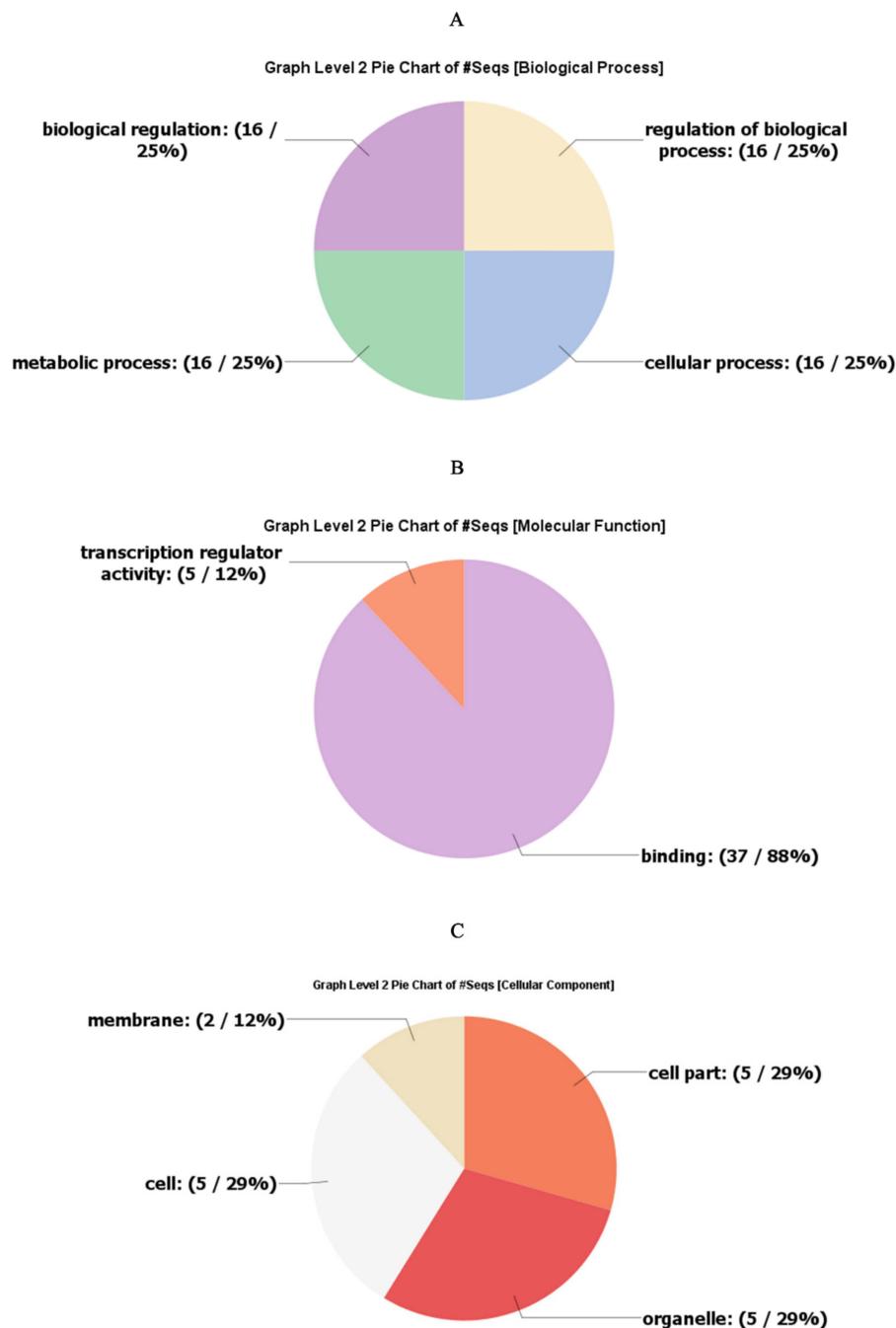


Figure 5. GO annotation analysis of CtbHLH transporter proteins. The annotation analysis included three major categories: (A) developmental process, (B) molecular function, and (C) cellular component.

2.6. Protein Clustering Networks

Multiple bHLH proteins can form homodimers or heterodimers that bind to DNA and regulate transcription of downstream targets; hence, protein–protein interactions have fundamental importance in bHLH protein function. Orthologous analysis of AtbHLHs was used to predict the interaction network of 41 candidate CtbHLHs. We validated 10 candidate proteins that interact with these CtbHLHs, which include a nuclear pore complex protein (1), TFs (5), E3 ubiquitin ligases (1), phytochromes (2), and repressors of jasmonic acid (1).

The independent protein interaction networks described in the Materials and methods demonstrated that the CtbHLH8 and CtbHLH16 proteins interacted with the DYT1 protein, which indicated complex regulation of certain proteins which are essential in early pollen maturation. The CtbHLH3 and CtbHLH9 proteins interact with the BTS protein and negatively regulate the response to iron deficiency and thus contributes to iron homeostasis. The CtbHLH39, CtbHLH2, and CtbHLH28 proteins interact with the TIFY10B protein and may elevate plant defense potential. The CtbHLH8 protein interacts with the AT1G59660 protein and may be involved in transport. Notably, five different proteins, including two phytochromes proteins (PHYA and PHYB) and three TF proteins (BZR1, HY5, and ABI3), were found to interact with the CtbHLH32 protein (Figure 6).

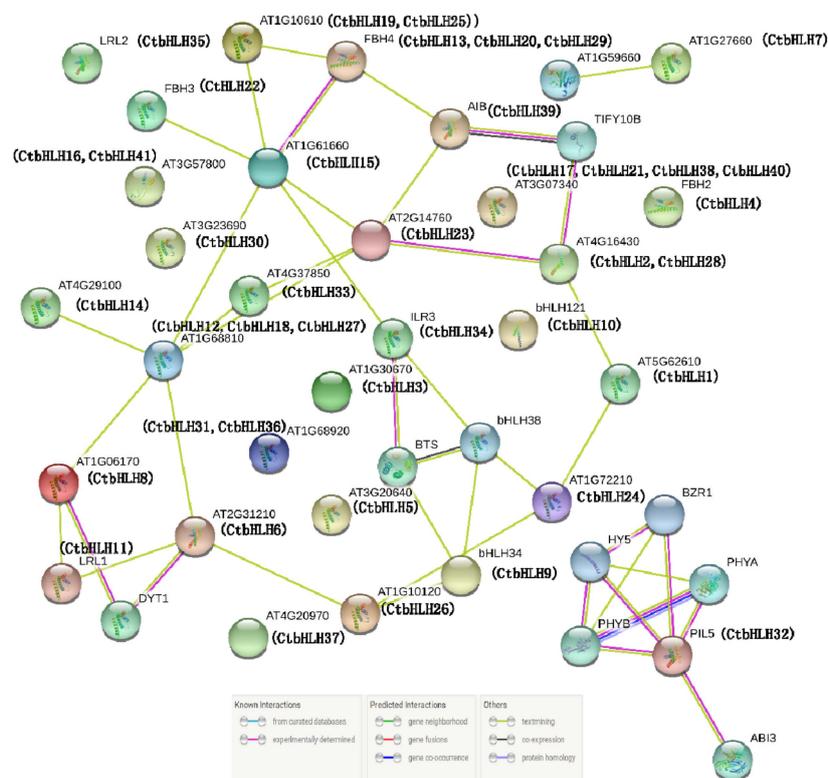


Figure 6. Protein association network for CtbHLHs according to information available about CtbHLH orthologs in *Arabidopsis*. The online tool STRING was used to predict the entire network. The red lines represents groups of proteins that are related to more than four other candidate bHLH proteins. The CtbHLH proteins are indicated in parentheses below the *Arabidopsis* orthologs.

2.7. Topology of Conserved Cis-Elements within the Promoter Regions of CtbHLH Genes

Along with evolutionary functional divergence, the putative gene families may also exhibit diverse patterns. To identify variations in gene regulation patterns, we screened all cis-elements within the promoter regions of the selected CtbHLH genes from *Carthamus tinctorius*. PlantCARE promoter sequence analysis identified multiple cis-elements known to participate in the biological responses to stress found in the promoter region of the CtbHLH sequence (2000 bp upstream of the transcription

start site, TSS), which was isolated from safflower genome database (GDB). The frequently predicted cis-elements in the promoter regions of most of the CtbHLH gene family are those associated with light response, which indicates that light signals may play an essential role in transcriptional regulation of the *Carthamus tinctorius* CtbHLH genes (Figure 7). For example, all 41 CtbHLH genes contained G-box cis-elements. The most commonly predicted cis-elements in the promoter regions of various CtbHLH genes were elements related to the light response, which indicated that light signals might play important roles in the transcriptional regulation of CtbHLH genes (Figure 7). All 41 of the CtbHLH genes contained the G-box cis-elements.

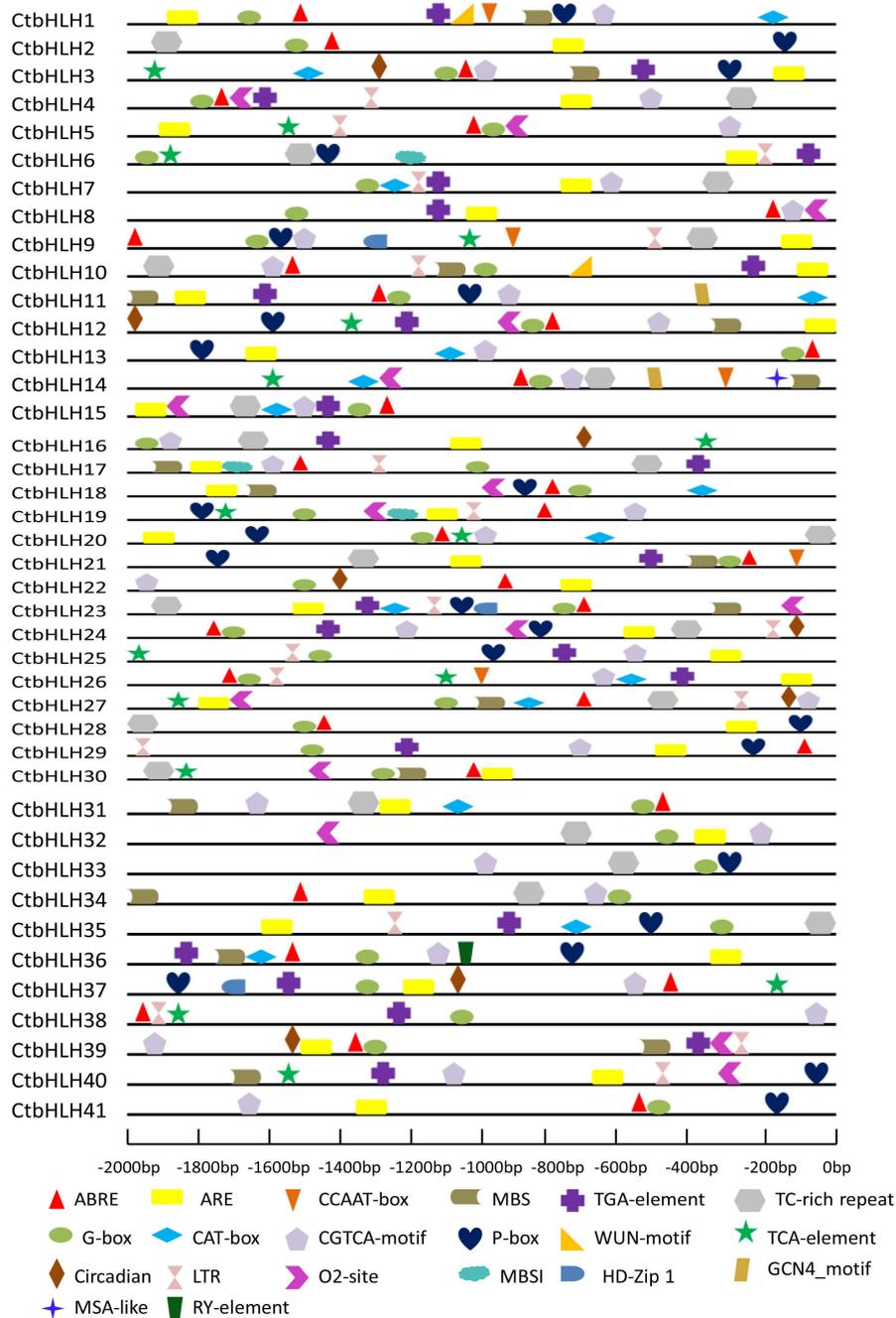


Figure 7. Illustration of 20 selected cis-regulatory units in all promoters of safflower bHLH genes. From the translation start site, the upstream region from -2000 to -200 shows the upstream region of each bHLH promoter. Several units, including ABRE, G-box, CGTCA-motif, P-box, and TC-rich repeats, exist in the promoter.

Cis-elements conserved in other major plant responses were also investigated inside the promoter regions of different CtbHLH genes, such as elements that are highly responsive to gibberellins (P-box), auxin (TGA element), jasmonic acid (CGTCA-motif), low temperatures (LTR element), salicylic acid (ARE and TCA elements), abscisic acid (ABRE element), defense and stress (TC-rich repeats), and endosperm-specific expression (RY element) (Figure 7).

2.8. Subcellular Localization of CmbHLH40

The hypothetical prediction of a candidate CmbHLH40 was initially estimated using the online software ProtComp version 9.0 (www.softberry.com), which was preliminarily determined to be localized to the nucleus. The transient expression assay was carried out which involved the transformation of the CmbHLH40-GFP-pCAMBIA1302 fusion construct into onion epidermal cells. It was found that the GFP signals were largely expressed inside the nucleus (Figure 8A–C). However, GFP alone exhibited a dispersed pattern throughout the whole cell. (Figure 8D–F).

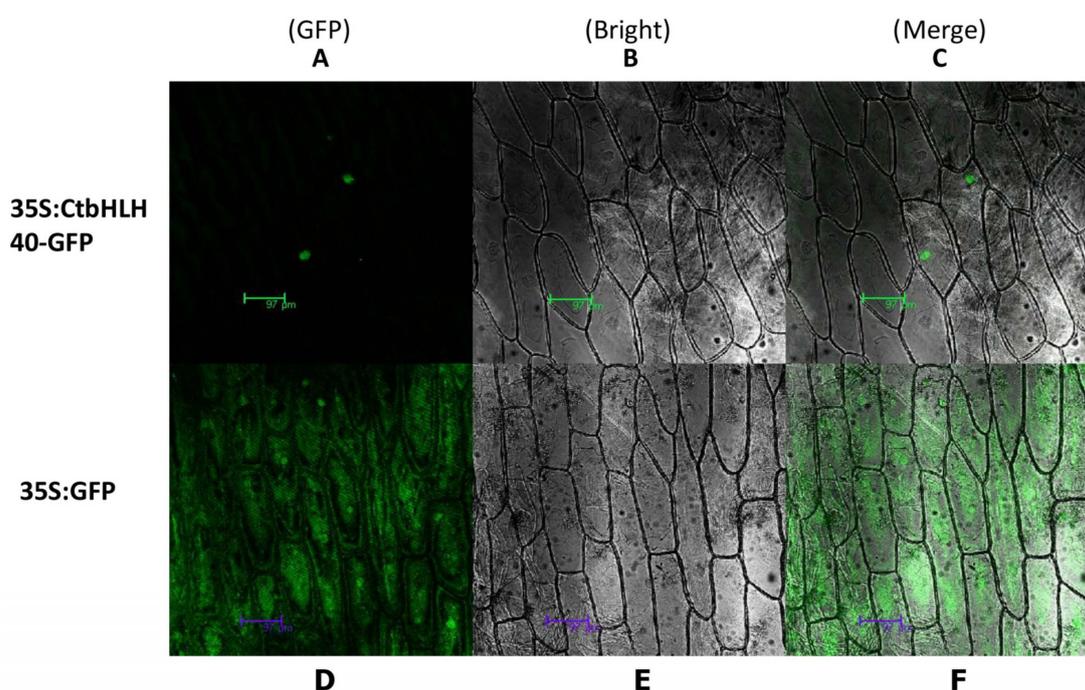


Figure 8. Subcellular compartment of CtbHLH40 based on transient expression in onion epidermal cells. (A–C) Onion cells introduced with 35S:CtbHLH40-GFP. (D–F) Induction of the control cells induced with 35S::GFP. (B–E) Onion epidermal cells under bright light to display the complete cell morphology. (A–D) Compilation on a dark field to visualize GFP signals. The images presented in (C) and (F) represent merging of (A) and (B), and (D) and (E), respectively. The fluorescence was detected with the help of Leica Microsystem CMS, D-68165 Mannheim. Scar bar = 97 μm .

3. Materials and Methods

3.1. Plant Materials

The “Jihong No.1” variety of safflower was used in this study, and the seeds were purchased from Tacheng, Xinjiang. At the end of April, 2018, the seeds were planted in the experimental field of the Engineering Research Center of Jilin Agricultural University Changchun, China, and safflower petals were collected at the flower bud stage (0 d) and 2 d, 4 d, 6 d, 8 d, 10 d, and 12 d after flowering, and quickly placed in the liquid nitrogen, wrapped in tin foil paper, and stored at $-80\text{ }^{\circ}\text{C}$ until next use. The plant expression vector pCAMBIA1302 was preserved in our laboratory.

3.2. Identification of Safflower bHLH Superfamily Genes

A comprehensive HMM of the bHLH domain (PF00010) was obtained from the online server of the Pfam database, which is available at <http://pfam.xfam.org/> [38]. We then scanned the entire genome of safflower using the *Carthamus tinctorius* keywords by using the HMMER software tools (<http://hmmer.org/>) and selecting the default cut off E-value as 0.01. Assembly of the complete set of bHLH proteins sequences obtained from candidate genes along with their ID numbers were identified from the Phytozome database using the previous HMMER results (<https://phytozome.jgi.doe.gov>). These putative bHLH protein sequences were subsequently processed through the online CD-search tool available at <https://www.ncbi.nlm.nih.gov/>. The conserved domain of the *Carthamus tinctorius* bHLH proteins were further identified with the help of the online version of InterProScan (<http://www.ebi.ac.uk/Tools/InterProScan/>) using the following parameters: Homologous superfamilies domain and repeats, detailed signature matches, residues annotation and other features.

3.3. Motif Elicitation and Phylogenetic Analysis of bHLH Proteins

The conserved motifs of herbaceous *Carthamus tinctorius* bHLH protein superfamily were identified by uploading 41 amino acid sequences into the online tool of MEME web server (Version 4.8.1; <http://meme.nbcr.net/meme/cgi-bin/meme.cgi>). The parameters were optimized as follows: zero or one, incidence of one motif per sequence; 10 bp, motif breadth range; and three other wide ranges of motifs were identified. All different parameters follow with the default values. Multiple sequence alignment using the 41 amino acid sequences of the *Carthamus tinctorius* bHLH protein superfamily was carried out in ClustalW (Version 2) [39]. The 138 taxon members' sequences of the *Arabidopsis* bHLH TF family were extracted from The *Arabidopsis* Info Resource (TAIR), which can be obtained from <http://www.arabidopsis.org/>. Visualization of the phylogenetic trees of the *Carthamus tinctorius* bHLH taxon proteins from *Arabidopsis* and *Carthamus tinctorius* were made with the ClustalW tool in combination with MEGA version 4.1 [39] using the neighbor-joining methodology, with 1000 bootstrap replicates. According to the subgroup taxonomic organization of the *Arabidopsis* bHLH proteins, a phylogenetic tree of *Carthamus tinctorius* and *Arabidopsis* was generated, following by the structural and functional classification into subfamilies.

3.4. Gene Ontology Annotation

GO analysis of the *Carthamus tinctorius* bHLH transporter protein superfamily was conducted using the Blast2GO tool (<https://www.blast2go.com/>) [40]. The full-length aa sequences of CtbHLH proteins were uploaded to Blast2Go for blast search followed by mapping and annotation. The analysis was disbursed in three categories: developmental process, molecular function, and cellular component.

3.5. Expression Analysis of the *Carthamus* bHLH Protein Superfamily

For this purpose, the safflower plants were maintained in the artificial chamber of our laboratory under controlled conditions (25°C, 16h light and 8h dark) until flowering. The safflower petals frozen at -80 °C were quickly placed in liquid nitrogen and ground, and then extracted by RNA according to the manufacturer's instructions. The extracted RNA was measured for purity and concentration of the extracted RNA by a NanoDrop 2000 ultraviolet spectrophotometer. The gene stability was detected by 1% agarose gel electrophoresis, and the RNA was stored at -80 °C. The cDNA was reverse transcribed according to the instructions of the reverse transcription kit, and the reverse-transcribed cDNA was preserved at -20 °C and then used for quantitative real-time PCR (qRT-PCR) analysis. A set of gene primers for the 41 safflower bHLH genes was synthesized based on the information obtained from the coding sequences in Primer Premier version 5 (Premier Biosoft, Palo Alto, CA, USA). To verify the integrity of primer specificity, each primer pair was generated away from the conserved domain of the genes. The internal reference genes of *Carthamus tinctorius* used in qRT-PCR analysis included 18S ribosomal RNA gene.

qRT-PCR was performed using SYBR[®] Premix Ex Taq[™] (Tli RNaseH Plus) from TaKaRa Biotechnology Co. Ltd. (Dalian, China) using the Applied Biosystems 7500 real-time PCR machine. Following the manufacturer's instructions, a 20- μ L reaction mixture was prepared with 1.0 μ L cDNA template, 0.4 μ M primers (F/R), 0.4 μ L ROX dye, 10 μ L master mix, and 7.8 μ L RNase-free water. PCR conditions were set according to the manufacturer's protocol for SYBR[®] Premix Ex Taq[™]. Each experiment was repeated for biological replicates of the petal samples. The relative transcript abundance values were calculated using the $2^{-\Delta\Delta C_t}$ method [41].

3.6. Protein Interaction Network Prediction and Functional Annotations Using STRING

The putative *Carthamus tinctorius* CtbHLH protein sequences were added to the online server STRING version 10 (<http://string-db.org>). The comparative organism was specified as *Arabidopsis thaliana*. The set of genes that exhibited the highest bit scores were utilized to create the hierarchical network of proteins. The annotation information of the functional domains was pasted manually from the blast results.

3.7. Promoter Sequence Analysis of CtbHLH

Conserved cis-elements within the promoter region of *Carthamus tinctorius* CtbHLH genes were studied by obtaining the 2000-bp upstream flanking sequence from each putative CtbHLH TSS from the online webserver Phytozome (<https://phytozome.jgi.doe.gov>). Promoter sequence analysis was carried out using PlantCARE.

3.8. Subcellular Localization of CtbHLH40

The pCAMBIA1302 vector with the CaMV35S promoter upstream of the multiple cloning site was used to identify the localization of CtbHLH40 using the transient expression system of the onion epidermal cells. For this purpose, we isolated the *SpeI* and *NcoI* digested CtbHLH40 gene and cloned into the pEASY-T1 vector and the positive bacteria was sent for sequencing (Shanghai Biotechnology Co., Ltd. (Shanghai, China)). After the sequencing, the gene fragment was digested with the aforesaid restriction enzymes and ligated into the *SpeI-NcoI* digested pCAMBIA1302 vector using T4 ligase. The recombinant plasmid pCtbHLH40-GFP was successfully transformed into onion epidermal cells, at the same time pCAMBIA1302 alone was also transformed to use as control [41]. GFP expression was analyzed by scanning confocal laser microscopy.

3.9. Statistical Analysis

All our findings were presented as mean \pm S.D. with three independent biological replicates. Differences between means of each group were assessed by one-way analysis of variance using Statistix 8.1 software. P-values equal to 0.05 was kept statistically significant.

4. Conclusions

Our study provides the first all-inclusive and systematic genome-wide analysis of the safflower bHLH superfamily. As a result, a set of 41 genes was characterized and classified into subfamilies. Comprehensive phylogenetics analysis of the 41 safflower bHLH proteins and 138 *Arabidopsis* bHLH proteins was carried out to identify the conserved homology between these proteins. Furthermore, protein motifs, their compositions, and their amino acid ratios within selected plant species and promoter characterization were also thoroughly investigated. Gene expression inferred by a heat map, extensive tissue-specific qRT-PCR analysis at various developmental stages, and subcellular localization of the candidate genes provided preliminary information for future investigation of the safflower bHLH gene family during flavonoid metabolism. The current investigation lays the highlights for functional characterization and confirmation of the safflower bHLH gene superfamily and improves our understanding of the bHLH gene superfamily in higher plants.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/12/3044/s1>.

Author Contributions: Conceptualization, Y.H.; Data curation, N.A. and W.L.; Formal analysis, Y.T., G.W. and J.L.; Funding acquisition, H.L.; Investigation, Y.H. and L.W.; Methodology, N.Y., X.Z. and X.L.; Software, Y.H. and X.L.; Writing original draft, Y.H. and N.A.; Writing review & editing, F.W., N.A. and Y.D.

Funding: This work was funded by grants from the National Natural Science Foundation of China (No. 31501366, 31771868), and the Science and Technology Development Project of Jilin Province (No. 20190201175JC, 20190201172JC, 20150623024TC-11). The 13th Five-Year Plan of Science and Technology Project of Jilin (Grant No. JJKH20190918KJ).

Acknowledgments: We thank Mallory Eckstut, from Liwen Bianji, Edanz Editing China (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

TFs	Transcription factors
bHLH	Basic helix–loop–helix
CLA	Conjugated linoleic acid
GRAVY	Grand average of hydropathicity
PI	Isoelectric points
Kda	Killi dalton
CDS	Coding sequences
HMM	Hidden Markov model
qRT-PCR	Quantitative real-time PCR
MW	Molecular weight
TSS	Transcription start site
GDB	Genome database
CDD	Conserved Domain

References

1. Massari, M.E.; Murre, C. Helix-loop-helix proteins: Regulators of transcription in eucaryotic organisms. *Mol. Cell. Biol.* **2000**, *20*, 429–440. [[CrossRef](#)] [[PubMed](#)]
2. Niu, X.; Guan, Y.; Chen, S.; Li, H. Genome-wide analysis of basic helix-loop-helix (bHLH) transcription factors in *Brachypodium distachyon*. *Bmc Genom.* **2017**, *18*, 619. [[CrossRef](#)] [[PubMed](#)]
3. Toledo-Ortiz, G.; Huq, E.; Quail, P.H. The Arabidopsis basic/helix-loop-helix transcription factor family. *Plant Cell* **2003**, *15*, 1749–1770. [[CrossRef](#)]
4. Li, X.; Duan, X.; Jiang, H.; Sun, Y.; Tang, Y.; Yuan, Z.; Guo, J.; Liang, W.; Chen, L.; Yin, J. Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis. *Plant Physiol.* **2006**, *141*, 1167–1184. [[CrossRef](#)] [[PubMed](#)]
5. Feller, A.; Machemer, K.; Braun, E.L.; Grotewold, E. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* **2011**, *66*, 94–116. [[CrossRef](#)] [[PubMed](#)]
6. Pires, N.; Dolan, L. Origin and diversification of basic-helix-loop-helix proteins in plants. *Mol. Biol. Evol.* **2009**, *27*, 862–874. [[CrossRef](#)]
7. Buck, M.J.; Atchley, W.R. Phylogenetic analysis of plant basic helix-loop-helix proteins. *J. Mol. Evol.* **2003**, *56*, 742–750. [[CrossRef](#)]
8. Carretero-Paulet, L.; Galstyan, A.; Roig-Villanova, I.; Martínez-García, J.F.; Bilbao-Castro, J.R.; Robertson, D.L. Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. *Plant Physiol.* **2010**, *153*, 1398–1412. [[CrossRef](#)]
9. Gao, C.; Sun, J.; Wang, C.; Dong, Y.; Xiao, S.; Wang, X.; Jiao, Z. Genome-wide analysis of basic/helix-loop-helix gene family in peanut and assessment of its roles in pod development. *Plos ONE* **2017**, *12*, e0181843. [[CrossRef](#)]
10. Yang, S.; LIU, H.-d.; Qiang, Z.; ZHANG, H.-j.; ZHANG, Z.-d.; LI, Y.-d.; WANG, H.-b.; LIU, F.-z. High-throughput sequencing of highbush blueberry transcriptome and analysis of basic helix-loop-helix transcription factors. *J. Integr. Agric.* **2017**, *16*, 591–604.

11. Sun, H.; Fan, H.-J.; Ling, H.-Q. Genome-wide identification and characterization of the bHLH gene family in tomato. *Bmc Genom.* **2015**, *16*, 9. [[CrossRef](#)] [[PubMed](#)]
12. Song, X.-M.; Huang, Z.-N.; Duan, W.-K.; Ren, J.; Liu, T.-K.; Li, Y.; Hou, X.-L. Genome-wide analysis of the bHLH transcription factor family in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *Mol. Genet. Genom.* **2014**, *289*, 77–91. [[CrossRef](#)] [[PubMed](#)]
13. Mao, K.; Dong, Q.; Li, C.; Liu, C.; Ma, F. Genome wide identification and characterization of apple bHLH transcription factors and expression analysis in response to drought and salt stress. *Front. Plant Sci.* **2017**, *8*, 480. [[CrossRef](#)] [[PubMed](#)]
14. Ogo, Y.; Nakanishi Itai, R.; Nakanishi, H.; Kobayashi, T.; Takahashi, M.; Mori, S.; Nishizawa, N.K. The rice bHLH protein OsIRO2 is an essential regulator of the genes involved in Fe uptake under Fe-deficient conditions. *Plant J.* **2007**, *51*, 366–377. [[CrossRef](#)] [[PubMed](#)]
15. Kavaz, M.; Baloglu, M.C.; Atabay, E.S.; Ziplar, U.T.; Dasgan, H.Y.; Ünver, T. Genome-wide characterization and expression analysis of common bean bHLH transcription factors in response to excess salt concentration. *Mol. Genet. Genom.* **2016**, *291*, 129–143. [[CrossRef](#)] [[PubMed](#)]
16. Zhang, X.; Luo, H.; Xu, Z.; Zhu, Y.; Ji, A.; Song, J.; Chen, S. Genome-wide characterisation and analysis of bHLH transcription factors related to tanshinone biosynthesis in *Salvia miltiorrhiza*. *Sci. Rep.* **2015**, *5*, 11244. [[CrossRef](#)] [[PubMed](#)]
17. Szecsi, J.; Joly, C.; Bordji, K.; Varaud, E.; Cock, J.M.; Dumas, C.; Bendahmane, M. BIGPETALp, a bHLH transcription factor is involved in the control of Arabidopsis petal size. *Embo J.* **2006**, *25*, 3912–3920. [[CrossRef](#)]
18. Roh, J.S.; Han, J.Y.; Kim, J.H.; Hwang, J.K. Inhibitory effects of active compounds isolated from safflower (*Carthamus tinctorius* L.) seeds for melanogenesis. *Biol. Pharm. Bull.* **2004**, *27*, 1976–1978. [[CrossRef](#)]
19. Li, X.Z.; Long, R.J.; Yan, C.G.; Lee, H.G.; Kim, Y.J.; Song, M.K. Rumen microbial response in production of CLA and methane to safflower oil in association with fish oil or/and fumarate. *Anim. Sci. J.* **2011**, *82*, 441–450. [[CrossRef](#)]
20. Cazzato, E.; Laudadio, V.; Corleto, A.; Tufarelli, V. Effects of harvest date, wilting and inoculation on yield and forage quality of ensiling safflower (*Carthamus tinctorius* L.) biomass. *J. Sci. Food Agric.* **2011**, *91*, 2298–2302. [[CrossRef](#)]
21. Thao, N.T.P.; Cuong, T.D.; Hung, T.M.; Lee, J.H.; Na, M.; Son, J.K.; Jung, H.J.; Fang, Z.; Woo, M.H.; Choi, J.S. Simultaneous determination of bioactive flavonoids in some selected Korean thistles by high-performance liquid chromatography. *Arch. Pharmacol Res.* **2011**, *34*, 455. [[CrossRef](#)] [[PubMed](#)]
22. Wu, C.; Zhang, J.; Zhou, T.; Guo, B.; Wang, Y.; Hou, J. Simultaneous determination of seven flavonoids in dog plasma by ultra-performance liquid chromatography–tandem mass spectrometry and its application to a bioequivalence study of bioactive components in *Herba Epimedii* and *Er-Xian* Decoction. *J. Pharm. Biomed. Anal.* **2011**, *54*, 186–191. [[CrossRef](#)] [[PubMed](#)]
23. Ye, S.-Y.; Gao, W.-Y. Hydroxysafflor yellow A protects neuron against hypoxia injury and suppresses inflammatory responses following focal ischemia reperfusion in rats. *Arch. Pharmacol Res.* **2008**, *31*, 1010–1015. [[CrossRef](#)] [[PubMed](#)]
24. Zhang, Y.; Guo, J.; Dong, H.; Zhao, X.; Zhou, L.; Li, X.; Liu, J.; Niu, Y. Hydroxysafflor yellow A protects against chronic carbon tetrachloride-induced liver fibrosis. *Eur. J. Pharmacol.* **2011**, *660*, 438–444. [[CrossRef](#)] [[PubMed](#)]
25. Liu, X.; Dong, Y.; Yao, N.; Zhang, Y.; Wang, N.; Cui, X.; Li, X.; Wang, Y.; Wang, F.; Yang, J. De Novo Sequencing and Analysis of the Safflower Transcriptome to Discover Putative Genes Associated with Safflor Yellow in *Carthamus tinctorius* L. *Int. J. Mol. Sci.* **2015**, *16*, 25657–25677. [[CrossRef](#)] [[PubMed](#)]
26. Zhang, C.; Feng, R.; Ma, R.; Shen, Z.; Cai, Z.; Song, Z.; Peng, B.; Yu, M. Genome-wide analysis of basic helix-loop-helix superfamily members in peach. *PLoS ONE* **2018**, *13*, e0195974. [[CrossRef](#)] [[PubMed](#)]
27. Guo, X.-J.; Wang, J.-R. Global identification, structural analysis and expression characterization of bHLH transcription factors in wheat. *BMC Plant Biol.* **2017**, *17*, 90. [[CrossRef](#)] [[PubMed](#)]
28. Chen, Y.-Y.; Li, M.-Y.; Wu, X.-J.; Huang, Y.; Ma, J.; Xiong, A.-S. Genome-wide analysis of basic helix-loop-helix family transcription factors and their role in responses to abiotic stress in carrot. *Mol. Breed.* **2015**, *35*, 125. [[CrossRef](#)]
29. Wang, R.; Zhao, P.; Kong, N.; Lu, R.; Pei, Y.; Huang, C.; Ma, H.; Chen, Q. Genome-wide identification and characterization of the potato bHLH transcription factor family. *Genes* **2018**, *9*, 54. [[CrossRef](#)]

30. Bailey, T.L.; Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in bipolymers. *1994*, *2*, 28–36.
31. Heim, M.A.; Jakoby, M.; Werber, M.; Martin, C.; Weisshaar, B.; Bailey, P.C. The basic helix–loop–helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. *Mol. Biol. Evol.* **2003**, *20*, 735–747. [[CrossRef](#)] [[PubMed](#)]
32. Wang, N.; Cui, Y.; Liu, Y.; Fan, H.; Du, J.; Huang, Z.; Yuan, Y.; Wu, H.; Ling, H.-Q. Requirement and functional redundancy of Ib subgroup bHLH proteins for iron deficiency responses and uptake in *Arabidopsis thaliana*. *Mol. Plant* **2013**, *6*, 503–513. [[CrossRef](#)] [[PubMed](#)]
33. Zhu, E.; You, C.; Wang, S.; Cui, J.; Niu, B.; Wang, Y.; Qi, J.; Ma, H.; Chang, F. The DYT 1-interacting proteins bHLH 010, bHLH 089 and bHLH 091 are redundantly required for *Arabidopsis* anther development and transcriptome. *Plant J.* **2015**, *83*, 976–990. [[CrossRef](#)] [[PubMed](#)]
34. Fernández-Calvo, P.; Chini, A.; Fernández-Barbero, G.; Chico, J.-M.; Gimenez-Ibanez, S.; Geerinck, J.; Eeckhout, D.; Schweizer, F.; Godoy, M.; Franco-Zorrilla, J.M. The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. *Plant Cell* **2011**, *23*, 701–715. [[CrossRef](#)] [[PubMed](#)]
35. Nesi, N.; Debeaujon, I.; Jond, C.; Pelletier, G.; Caboche, M.; Lepiniec, L. The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in *Arabidopsis* siliques. *Plant Cell* **2000**, *12*, 1863–1878. [[CrossRef](#)] [[PubMed](#)]
36. Jiang, Y.; Yang, B.; Deyholos, M.K. Functional characterization of the *Arabidopsis* bHLH92 transcription factor in abiotic stress. *Mol. Genet. Genom.* **2009**, *282*, 503–516. [[CrossRef](#)] [[PubMed](#)]
37. Tu, Y.; Liu, F.; Guo, D.; Fan, L.; Zhu, Z.; Xue, Y.; Gao, Y.; Guo, M. Molecular characterization of flavanone 3-hydroxylase gene and flavonoid accumulation in two chemotyped safflower lines in response to methyl jasmonate stimulation. *Bmc Plant Biol.* **2016**, *16*, 132. [[CrossRef](#)] [[PubMed](#)]
38. Finn, R.D.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Mistry, J.; Mitchell, A.L.; Potter, S.C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A. The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **2015**, *44*, D279–D285. [[CrossRef](#)]
39. Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **2011**, *28*, 2731–2739. [[CrossRef](#)]
40. Conesa, A.; Götz, S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genom.* **2008**, *2008*, 619832. [[CrossRef](#)]
41. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **2001**, *25*, 402–408. [[CrossRef](#)] [[PubMed](#)]

