

OPEN

Genome-wide identification and expression analysis of *bZIP* gene family in *Carthamus tinctorius* L.

Haoyang Li, Lixia Li, Guodong ShangGuan, Chang Jia, Sinan Deng, Muhammad Noman, Yilin Liu, Yongxin Guo, Long Han, Xiaomei Zhang, Yuanyuan Dong, Naveed Ahmad, Linna Du✉, Haiyan Li✉ & Jing Yang✉

The basic leucine zipper (bZIP) is a widely known transcription factors family in eukaryotes. In plants, the role of bZIP proteins are crucial in various biological functions such as plant growth and development, seed maturation, response to light signal and environmental stress. To date, bZIP protein family has been comprehensively identified in *Arabidopsis*, castor, rice, ramie, soybean and other plant species, however, the complete genome-wide investigation of *Carthamus tinctorius*-bZIP family still remains unexplained. Here, we identified 52 putative *bZIP* genes from *Carthamus tinctorius* using a draft genome assembly and further analyzed their evolutionary classification, physicochemical properties, Conserved domain analysis, functional differentiation and the investigation of expression level in different tissues. Based on the common bZIP domain, CtbZIP family were clustered into 12 subfamilies renamed as (A–J, S, X), of which the X is a unique subfamily to *Carthamus tinctorius*. A total of 20 conserved protein motifs were found in CtbZIP proteins. The expression profiling of CtbZIP genes deciphered their tissue-specific pattern. Furthermore, the changes in CtbZIP transcript abundance suggested that their transcription regulation could be highly influenced by light intensity and hormones. Collectively, this study highlights all functional and regulatory elements of bZIP transcription factors family in *Carthamus tinctorius* which may serve as potential candidates for functional characterization in future.

Abbreviations

BP	Biological process
bZIP	Basic region/leucine zipper
Ct	<i>Carthamus tinctorius</i> L.
CC	Cellular component
GRAVY	Grand average of hydropathicity
HMM	Hidden Markov model
Kda	Kilo Dalton
MF	Molecular function
MW	Molecular weight
RPKM	Reads per kilobase per million mapped reads
TFs	Transcription factors

Transcription factors (TFs) are regulated through sequence-specific DNA-binding proteins which interact with relevant cis-acting elements¹. The interaction alters transcription activity and stimulates or suppresses gene expression². As switch of gene expression, TFs play important regulatory roles in almost all processes of plant life³. Functional characterization of TFs is in-depth analyzed through biological processes and transcriptional

College of Life Science, Engineering Research Center of the Chinese Ministry of Education for Bioreactor and Pharmaceutical Development, Jilin Agricultural University, Changchun 130118, China. ✉email: dulinna0918@163.com; hyl99@163.com; yangjing5122010@163.com

regulatory networks. Thus, TFs are significant components of abiotic stress signaling pathways. Among them, the basic leucine zipper (bZIP) also accounts for a large and diverse TF family. All bZIP TFs consist of two structural components: a basic region (N-x7-R/K-x9) for sequence-specific DNA binding, and a leucine zipper. In *Arabidopsis*, 78 bZIP genes have been reported⁴, whereas 89 in *Oryza sativa*⁵, 55 in *Vitis vinifera*⁶, 125 in *Zea mays*⁷, 247 in *Brassica napus* L.⁸, 92 in *Sorghum vulgare*⁹ and 131 in *Glycine max*¹⁰. In general, it is known that the putative bZIP genes have been classified as tens of groups on the basis of sequence similarities of their basic regions and conserved motifs. *AtbZIP* gene family was classified into 13 subfamilies in *Arabidopsis thaliana*⁴ while the predicted bZIP proteins of *Oryza sativa* into 11 groups based on DNA-binding specificity and amino acid sequence¹¹. Although the number of groups of the bZIP family in *Oryza sativa* (11 groups) and *Arabidopsis thaliana* (13 groups) is different, however affiliation's relationship is similar. The interspecies clustering also indicates that homologous bZIP from both species fell into common group³.

In plants, bZIP TFs regulate many transcriptional response pathways in multiple biological processes. They regulate the development of tissues and organs, including seed maturation and germination, embryogenesis¹², blooming¹³ and photomorphogenesis¹⁴. In addition, bZIP TFs are involved in responses to abiotic and biotic stresses such as extreme temperatures, water deficit, high osmolarity and salinity and defence from pathogens^{15,16}. *AtbZIP* TFs as TGA2, TGA5 and TGA6 regulate salicylic acid-dependent mechanisms and activate jasmonic acid and ethylene-dependent pathway in abiotic stress, while ABF3 and ABF4, play essential roles in ABA stress responses¹⁷. HY5 and HYH, the main regulators of photomorphogenesis, mediate the light response in *Arabidopsis thaliana*¹⁸. Similarly, OsBZ8 is induced by ABA and mediates salt resistance¹⁹, while LIP19 functions as a molecular switch for low-temperature signal transduction in rice²⁰.

Carthamus tinctorius L. is an annual Asteraceae plant and an economic crop grown for edible oil extracted from its seed. It is suitable to grow in environments with insufficient moisture²¹ thus can be grown on marginal agricultural lands that are suitable for few other crops^{20,21}. It is a minor oilseed crop in terms of production among unexploited crops, which not only has high ornamental value, but also has other important practical value. Safflower seeds are rich in oil and contain high levels of unsaturated fatty acids, vitamin E and oryzanol. The application of safflower seed oil is very extensive in fuel industry, cardiovascular health care and production of pharmaceuticals as a plant factory. Therefore, genetic improvement of safflower is necessary to increase its acceptability and utility as an important oilseed crop, but due to its limited gene information resource, this demand still persists. To accelerate the improvement of safflower, its genome was sequenced recently. Development of safflower gene resources has the potential to speed up the process of molecular breeding, and can be used to investigate patterns of genome evolution.

Complete genomic information of safflower has not yet been uncovered fully, therefore, demands the analysis of molecular mechanisms, network regulation and functional diversity. A draft genome assembly of safflower was produced covering 866 million bp after sequencing a single, short insert library to ~21 × depth²². The full-length transcriptome sequencing of safflower generated 10.43 GB clean data from which 38,302 redundant sequences were captured²³. We accomplished the de novo transcriptome assembly of safflower from which we identified putative oleosin genes²⁴ by Solexa-based deep sequencing and investigated some genes related to the biosynthesis of safflower yellow²⁵. In addition, we have also sequenced the genome of safflower (Accession: PRJNA399628 ID: 399628). In this study, we screened fifty-two bZIP family genes from safflower genome database, named *CtbZIP1-52*. We not only analyzed the structural characteristics of *CtbZIP* family genes, identified *CtbZIP* motifs and constructed a phylogenetic tree, but also speculated their network regulating relationship and functional diversification among the members of *CtbZIP* family. The expression patterns of 52 *CtbZIP* genes in various tissues and different developmental stages were predicted by RPKM values, and the accuracy of expression profiles was verified by RT-qPCR. This study provides a comprehensive genome-wide investigation and expression analysis of *CtbZIP* family of safflower which would be important for functional characterization of *CtbZIP* TFs involved in biological processes and transcriptional regulatory networks, and then lay a foundation for molecular breeding of safflower in the future.

Results

Genome-wide identification of *CtbZIP* TFs. Through in silico analyses (detailed in methods), a total of 52 members in safflower bZIP gene family were identified. Based on the splicing results of genomic sequence, we sequentially sorted 52 *CtbZIP* proteins according to unigene number from small to large provisionality and named them *CtbZIP1-52* (Table 1). Their name, ID, ORF (open reading frame) length and polypeptide length as well as conserved domain position are mentioned in Table 1 while molecular weight, PI (Isoelectric Point) and Grand Average of Hydropathicity (GRAVY) are given in supplementary file S1, Table S1. The predicted molecular weights lie between 11.98 kDa to 86.19 kDa while PI ranges from 4.86 to 9.78. *CtbZIP38* gene has the shortest conserved domain with 21 amino acids, whereas *CtbZIP10* possesses the longest domain (74 amino acids). All negative GRAVY values indicate their hydrophilic nature.

Classification of the *CtbZIP* proteins based on phylogram. We constructed a phylogenetic tree to elucidate the evolutionary relationship among bZIP TFs of *Carthamus tinctorius* L., *Arabidopsis thaliana*, *Oryza sativa* and *Ricinus communis* (Fig. 1). *Arabidopsis thaliana* bZIP family has been classified into 13 subfamilies⁴. The bZIP TFs of most of plant species are classified according to the subfamilies of *Arabidopsis*. For example, the bZIP proteins of *Oryza sativa* were divided into 10 subfamilies⁵, *Ricinus communis* into 9²⁶ and *Camellia sinensis* into 11²⁷. We divided the 52 *CtbZIP* TFs into 12 subfamilies (*CtbZIP*-A, *CtbZIP*-B, *CtbZIP*-C, *CtbZIP*-D, *CtbZIP*-E, *CtbZIP*-F, *CtbZIP*-G, *CtbZIP*-H, *CtbZIP*-I, *CtbZIP*-J, *CtbZIP*-S and *CtbZIP*-X) on the basis of the classification of *Arabidopsis*. However, *CtbZIP13*, *CtbZIP14*, *CtbZIP20* and *CtbZIP46* could not be aggregated into any subfamily thus were grouped together into a separate branch named as subfamily X. None of *CtbZIP*

S. no	Name	ID in genomic data	Length of ORF	Number of aa	Conserved domain position (aa)
1	CtbZIP1	cgc000574	2052	683	197–258
2	CtbZIP2	cgc001746	795	264	186–229
3	CtbZIP3	cgc001897	981	326	238–284
4	CtbZIP4	cgc001910	1,113	370	240–277
5	CtbZIP5	cgc002216	1,050	349	107–168
6	CtbZIP6	cgc002309	516	171	25–82
7	CtbZIP7	cgc002770	852	283	236–278
8	CtbZIP8	cgc003603	1,101	366	289–338
9	CtbZIP9	cgc005204	594	197	71–130
10	CtbZIP10	cgc005666	945	314	178–224
11	CtbZIP11	cgc007675	1,044	347	150–223
12	CtbZIP12	cgc010618	810	269	90–148
13	CtbZIP13	cgc011035	891	296	138–188
14	CtbZIP14	cgc012241	1,710	569	525–550
15	CtbZIP15	cgc012409	1,302	433	354–405
16	CtbZIP16	cgc012979	1,152	383	295–357
17	CtbZIP17	cgc013975	1,218	405	41–92
18	CtbZIP18	cgc014366	1,119	372	243–281
19	CtbZIP19	cgc014560	567	188	75–133
20	CtbZIP20	cgc014897	339	112	57–127
21	CtbZIP21	cgc015074	429	142	39–88
22	CtbZIP22	cgc015568	318	105	34–85
23	CtbZIP23	cgc016238	426	141	56–96
24	CtbZIP24	cgc016704	1,422	473	257–309
25	CtbZIP25	cgc016850	471	156	44–95
26	CtbZIP26	cgc016950	477	158	48–108
27	CtbZIP27	cgc016979	1,167	388	252–314
28	CtbZIP28	cgc017531	780	259	165–222
29	CtbZIP29	cgc017533	750	249	165–209
30	CtbZIP30	cgc017757	465	154	33–77
31	CtbZIP31	cgc017772	417	138	27–85
32	CtbZIP32	cgc018261	1,245	414	309–371
33	CtbZIP33	cgc018785	1593	530	401–459
34	CtbZIP34	cgc018897	1509	502	365–409
35	CtbZIP35	cgc019734	1,041	346	56–95
36	CtbZIP36	cgc020071	444	147	24–63
37	CtbZIP37	cgc021349	519	172	55–108
38	CtbZIP38	cgc022221	639	212	171–191
39	CtbZIP39	cgc022423	459	152	65–116
40	CtbZIP40	cgc022869	1,560	519	215–256
41	CtbZIP41	cgc023166	1,257	418	216–268
42	CtbZIP42	cgc023405	969	322	241–280
43	CtbZIP43	cgc025170	1,005	334	201–245
44	CtbZIP44	cgc025235	1755	584	432–490
45	CtbZIP45	cgc027083	1,197	398	92–125
46	CtbZIP46	cgc028775	627	208	93–130
47	CtbZIP47	cgc029577	390	129	34–93
48	CtbZIP48	cgc030605	876	291	240–283
49	CtbZIP49	cgc031009	594	197	84–134
50	CtbZIP50	cgc031237	960	319	149–205
51	CtbZIP51	cgc031238	780	259	88–145
52	CtbZIP52	cgc031837	636	211	84–138

Table 1. List of the identified safflower bZIP TFs and their attributes. *ORF* open reading frame, *aa* amino acids.

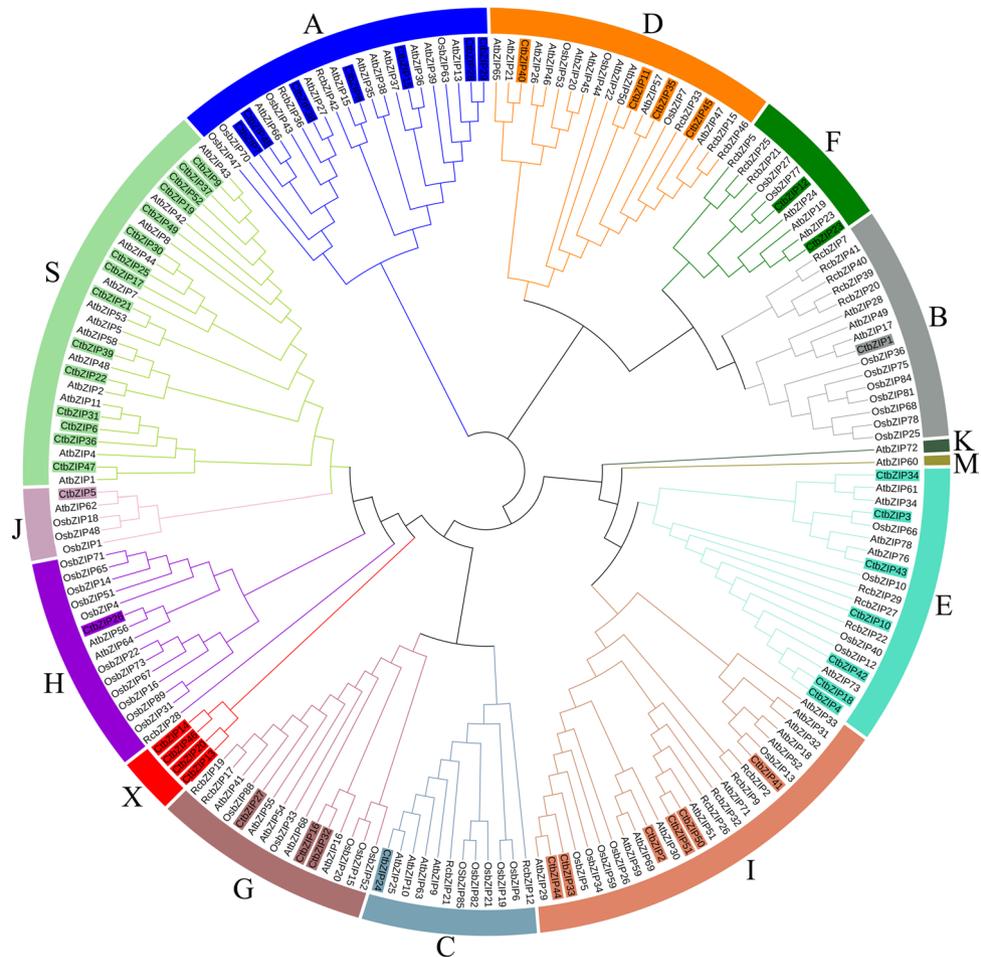


Figure 1. Cladogram of the CtbZIP TFs of Safflower, *Arabidopsis thaliana*, *Oryza sativa* and *Ricinus communis*. All bZIP TFs clustered into 12 distinct clades, marked by curves with different colors. Subfamily S has got the highest number (15) while J, H, C and B got the lowest (1) each, whereas 4 CtbZIPs exclusively clustered together in subfamily X.

proteins clustered into subfamily K and M indicating loss of these proteins throughout safflower evolution. A separate phylogenetic reconstruction elucidating the evolutionary relationship of Arabidopsis and safflower bZIP proteins is given in figure S1.

Motif analysis of the CtbZIP proteins. Except the bZIP conserved domain, bZIP proteins usually contain other motifs which might bind potential functional sites thereby activating their function. Using ORF Finder tool of NCBI database, we found a complete open reading frame of all 52 CtbZIP transcripts. To find the conserved domains, Pfam database²⁸ showed one or more of the intact conserved domains (bZIP_1, bZIP_2 and bZIP Maf while 20 conserved motifs were identified using MEME software²⁹ the names and sequence logos of which are illustrated in Fig. 2. We counted the width and E value of each conservative motif using TB tools³⁰ (Fig. 3A), and the distribution number of motifs in each subfamily was depicted (Fig. 3B). In terms of size, motif 20 was the shortest (20 aa) while motif 3, 11, 12, 17 and 18 were longest having 50 aa each. The motif average width lied around 38 aa. Interestingly, motif1 and motif2 were recognized as bZIP conserved domains and could be found in all of the subfamilies, however some subfamilies also had unique motif compositions (Fig. 3B,C). For example, subfamily A possesses a unique motif6, whereas motif11 is unique to subfamily D, motif 17 in subfamily I and motif13 in subfamily S. For the safflower specific subfamily X, CtbZIP20 and CtbZIP46 specifically contain motif 19, which are associated next to the N-terminus of the amino acid sequence and substantially identical to the bZIP conserved domain (Fig. 3C). All of these motifs indicate the group-specific functions for members in each group.

Functional differentiation of CtbZIP TFs. Some motifs of bZIP TFs participate in a variety of physiological processes. To understand their function in the biological processes, we predicted the function of CtbZIP TFs in silico using Gene Ontology (GO) terms³¹. All of 52 CtbZIP TFs were analyzed, 45 of which categorized into three primary GO functional categories, biological processes (BP), molecular function (MF) and cellular

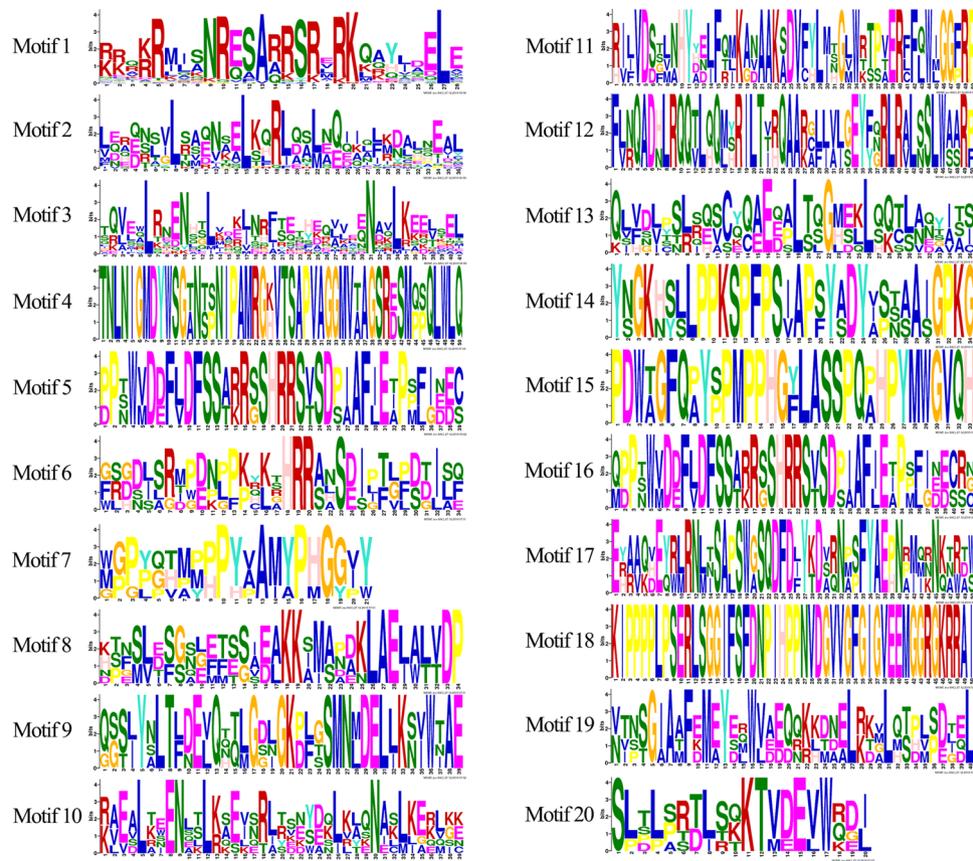


Figure 2. Sequence logos of CtbZIPs conserved motifs. From the multiply aligned protein sequences of CtbZIP, 20 conserved motifs were determined. Among them, Motif1 and Motif2 were common in all orthologs of the phylogenetic tree.

components (CC) (Fig. 4). Among the 45 CtbZIPs, none was individually enriched into a certain GO functional category. Six CtbZIP TFs (13%) are enriched in three major categories CC, BP and MF while 39 CtbZIP TFs (87%) enriched in BP and MF. It can be seen that CtbZIP has many functions that affect the biological process of safflower. Besides, 45 CtbZIP TFs are classified into 13 subcategories, accounting for 57% of the enrichment data. The enrichment analysis showed that besides subfamily J, CtbZIP TFs of 11 subfamilies are enriched (Fig. 5A). At the same time, 6 subcategories are significantly enriched. The majority of CtbZIP TFs have DNA binding activity (Fig. 5B) and participate in the process of nitrogen metabolism. A number of CtbZIP TFs might respond to various abiotic stresses. All CtbZIP TFs have transcriptional regulatory activity, this allows them to regulate the growth and development of safflower. Based on these findings, the function of CtbZIPs may be associated with various biosynthetic and metabolic processes in response to abiotic and biotic stresses to affect the development of various tissues and organs.

Expression profiles and network analysis of CtbZIP TFs. The bZIP TFs are not only the most widely distributed and most conserved eukaryotic transcription factors, but their function is also diverse. The safflower bZIP TFs have a variety of functions and there are synergistic effects in the exercise of their functions. In order to explore the expression profiles and the interaction among the CtbZIP TFs, we analyzed their expression variation in different tissues, including roots, stems, leaves, flowers, DAF10-seeds, DAF13-seeds and DAF20-seeds by heatmap (Fig. 6). We noticed that *CtbZIP13* highly expresses in roots. *CtbZIP6* and 25 transcripts are abundant while that of *CtbZIP40*, 23 and 29 are less in stem. *CtbZIP13* and 25 have higher expression in flowers than in other samples. High expression of *CtbZIP5* is observed in DAF13-seeds. Similarly, *CtbZIP52* highly expresses in DAF20-seeds. However, the expression levels of *CtbZIP22* is almost the same in all of the 7 samples. The varied expression pattern indicates functional divergence of different groups of CtbZIP TFs. These results indicate that the functions of CtbZIP family are differentiated with differentiation in their expression.

We quantified the expressions of all 52 CtbZIP TFs in different tissues and seeds (of various developmental stages). The expression networks ($p \leq 0.05$) (Fig. 7) were constructed using BioLayout Express 3D 3.2 software³². The CtbZIP TFs are a complex family with 51 nodes and 1,199 edges. Among them, 43 transcripts (85%) are more tended to have associated expression and form a co-expression network whereas the other 8 transcripts also exhibit weak co-expression. The network is composed of 5 clusters; the largest cluster contains sixteen transcripts, while the smallest cluster contains eight. There is a certain degree of related expression trend between these clusters and this tendency was statistically significant. These results indicate that although the functions

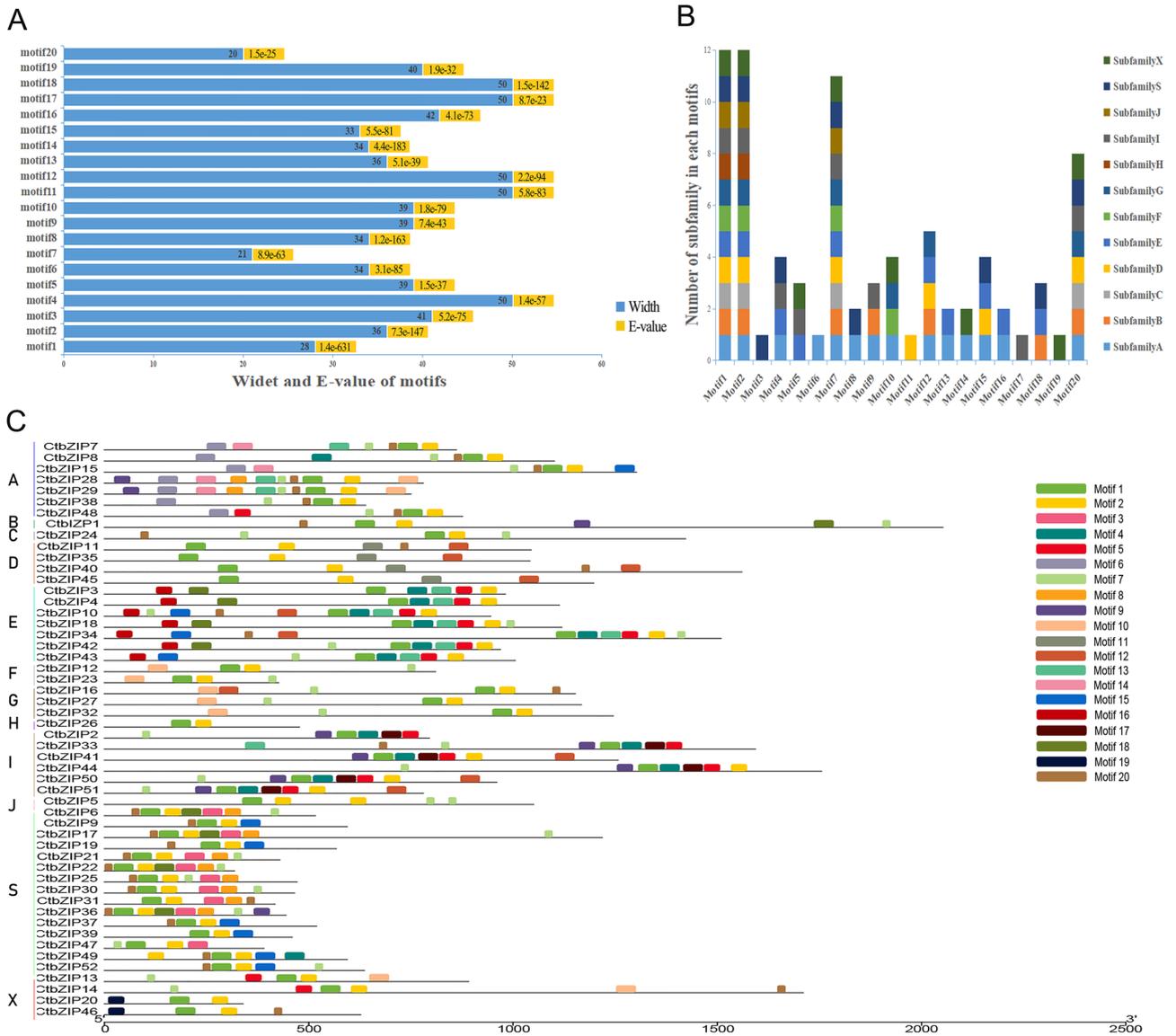


Figure 3. Conserved domain analysis of CtbZIP proteins in 12 subfamilies. **(A)** Width and E-value of sequence logos for 20 motifs. The blue portion indicates width and orange indicates E-value. **(B)** The number of subfamily in each motif. **(C)** Distribution of conserved structures in all 52 CtbZIP proteins.

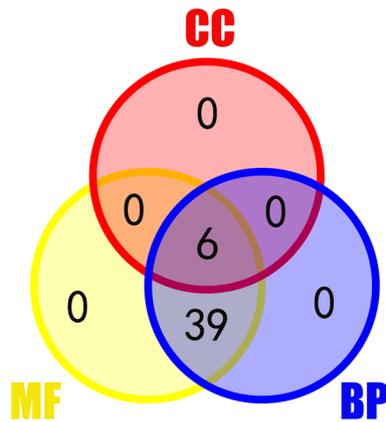


Figure 4. Venn diagram of the functional categorization of CtbZIP TFs. BP denotes biological process, MF stands for molecular function and CC for cellular component.

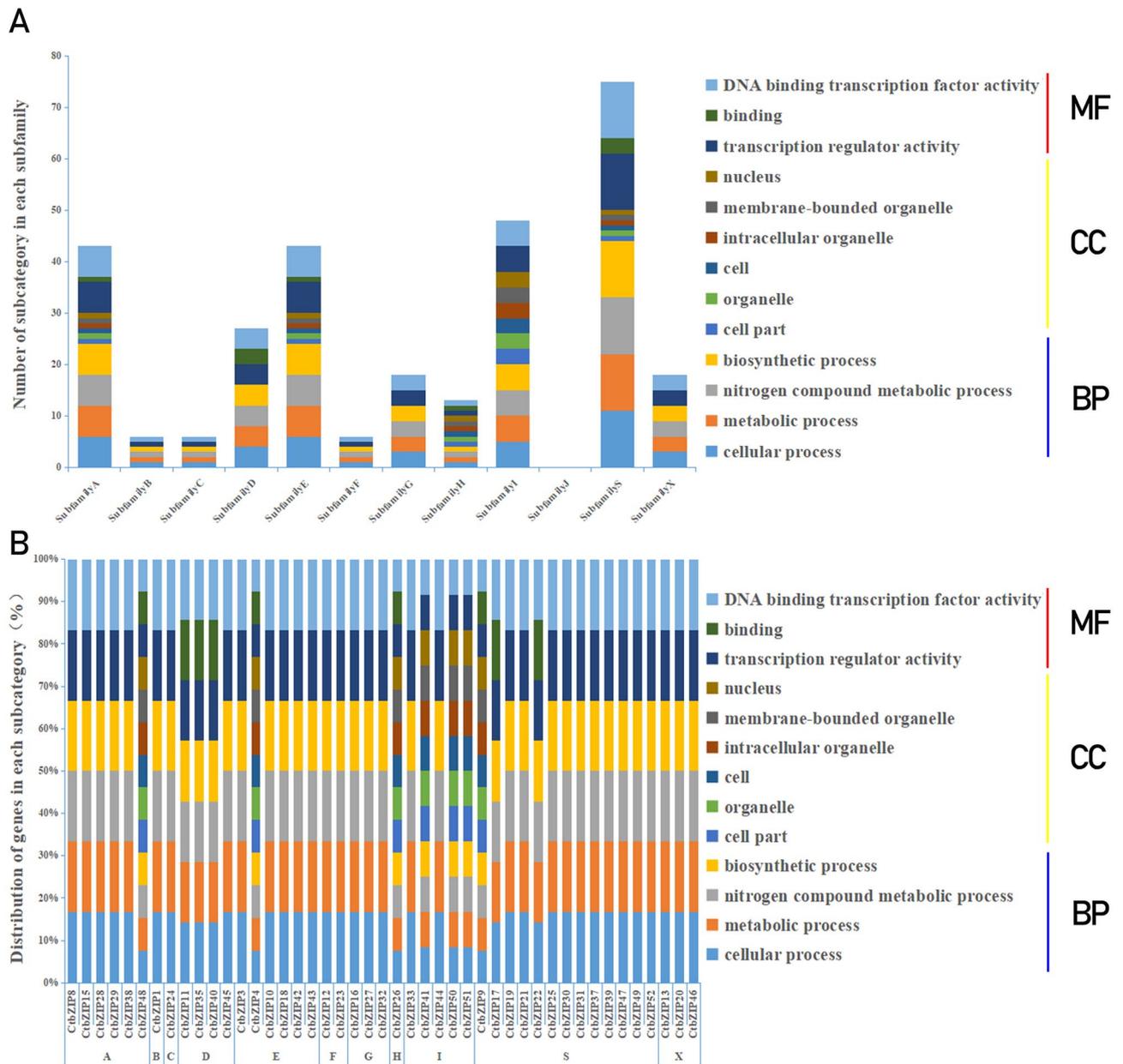


Figure 5. Variation of transcript function class in the CtbZIP family. (A) Distribution of each Subclass in 13 subspecies of Safflower CtbZIP TFs. (B) Variation of the functional categories of the 52 CtbZIP TFs.

and expressions of CtbZIP family members have dramatically diverged, they retain to some extent, the tendency of correlated expression and functional cooperation.

Expression analysis of CtbZIP TFs in various tissues. To further verify the authenticity of the expression pattern, we detected the expression level of 52 *CtbZIP* genes in different tissues of safflower including roots, stems, leaves, flowers, seeds, cotyledons and hypocotyls using RT-qPCR (Fig. 8). The results showed that the *CtbZIP25* gene is highly expressed in all tissues and we speculated that it may be involved in various stages of plant growth and development. The *CtbZIP13* is highly expressed in root and might play a role in root growth. In seeds, *CtbZIP52* has the highest expression and might regulate the development of seeds. Likewise, *CtbZIP25* and *CtbZIP30* have higher expression in hypocotyls. The expression level of *CtbZIP6* and *CtbZIP25* peak in stem and they may affect the growth of the stem. Conversely, the expression level of *CtbZIP2*, *CtbZIP23*, *CtbZIP31* and *CtbZIP34* is relatively low in all tissues, among which *CtbZIP34* is the lowest in roots while *CtbZIP2*, *CtbZIP22*, *CtbZIP31* and *CtbZIP47* are the lowest in stems. Similarly, *CtbZIP23* and *CtbZIP47* are the lowest in leaves, *CtbZIP23* in flowers and *CtbZIP19*, *CtbZIP20* and *CtbZIP23* in seeds have the lowest expression. However, *CtbZIP22* gene expresses in cotyledon and hypocotyl after seed germination. This indicates that the *CtbZIP22* gene is specifically involved in seed germination. In short, the results of RT-qPCR show that the expression pat-

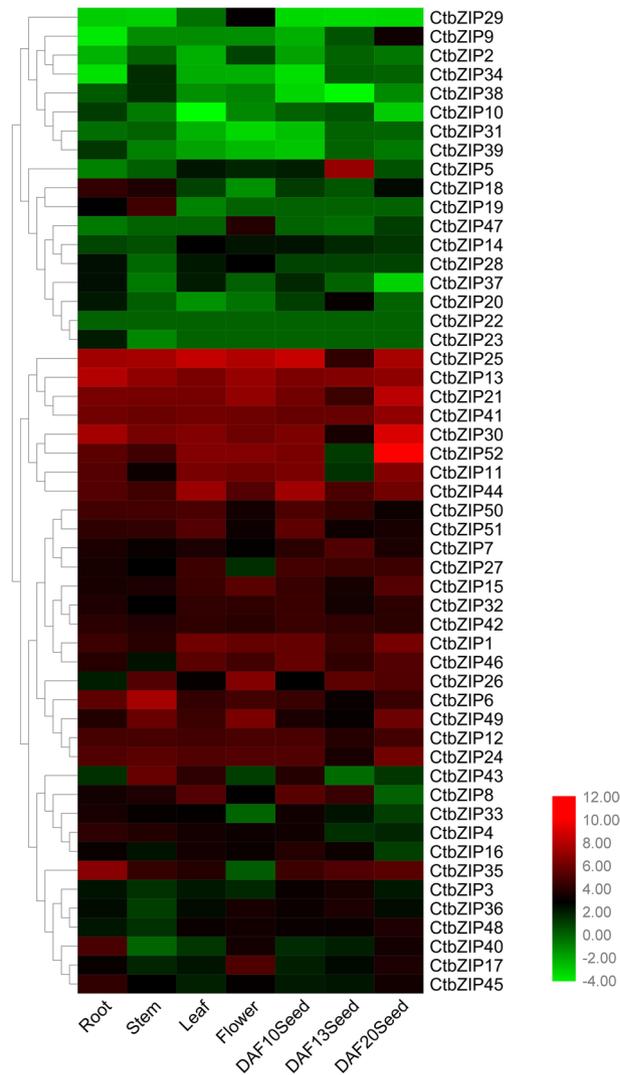


Figure 6. Heatmap of *CtbZIP* genes expressed among 7 tissues based on the fold change (\log_2) in RPKM values. The color scale at the top represents \log_2 -transformed reads per kilobase million (RPKM) values.

tern of safflower is consistent with the predicted expression. According to this expression pattern, the function of *CtbZIP* TFs can be more effectively estimated.

Expression level of *CtbZIP* TFs with GA3 treatment under different light radiation. In order to study the function of *CtbZIP* TFs, we detected the expression level of all 52 *CtbZIP* genes by RT-qPCR after GA3 treatment under different light radiation (no treatment under 16.8 MJ/m² light radiation, no treatment under 5.04 MJ/m² light radiation, GA3 treatment under 5.04 MJ/m² light radiation and GA3 treatment under MJ/m² light radiation) (Fig. 9). Among 52 genes, the expression of *CtbZIP15*, *CtbZIP26* and *CtbZIP28* highly increased in all the seven tissues, however, *CtbZIP28* and *CtbZIP38* upregulated in six tissues (excluding roots) after GA3 treatment under 16.8 MJ/m² light radiation. Similarly, *CtbZIP2*, *CtbZIP33*, *CtbZIP50* and *CtbZIP51* in roots and leaves while *CtbZIP6*, *CtbZIP36*, *CtbZIP49* and *CtbZIP52* in seeds were up-regulated. *CtbZIP8* and *CtbZIP15* were significantly affected by illumination intensity and their expression increased in leaves, flowers, seeds, cotyledon and hypocotyl. Likewise, in seeds, *CtbZIP35*, *CtbZIP40* and *CtbZIP45* up-regulated after GA3 treatment under 5.04 MJ/m² light radiation and *CtbZIP16*, *CtbZIP27* and *CtbZIP32* in cotyledon and hypocotyl were induced by GA3 and light.

Discussion

Safflower is an important plant used for ornamental, food, feed and medicinal purposes. In terms of tolerance for abiotic stresses such as water deficit, it is a tough plant however, for increasing demand of edible oil as well its vast pharmaceutical properties, its improvement seeks comprehensive understanding through omics. Omics by combining genomics, transcriptomics, proteomics and metabolomics (as solving a puzzle) attempts to obtain a clear picture of molecular and biochemical circuitries underlying primary and secondary metabolites/products³³.

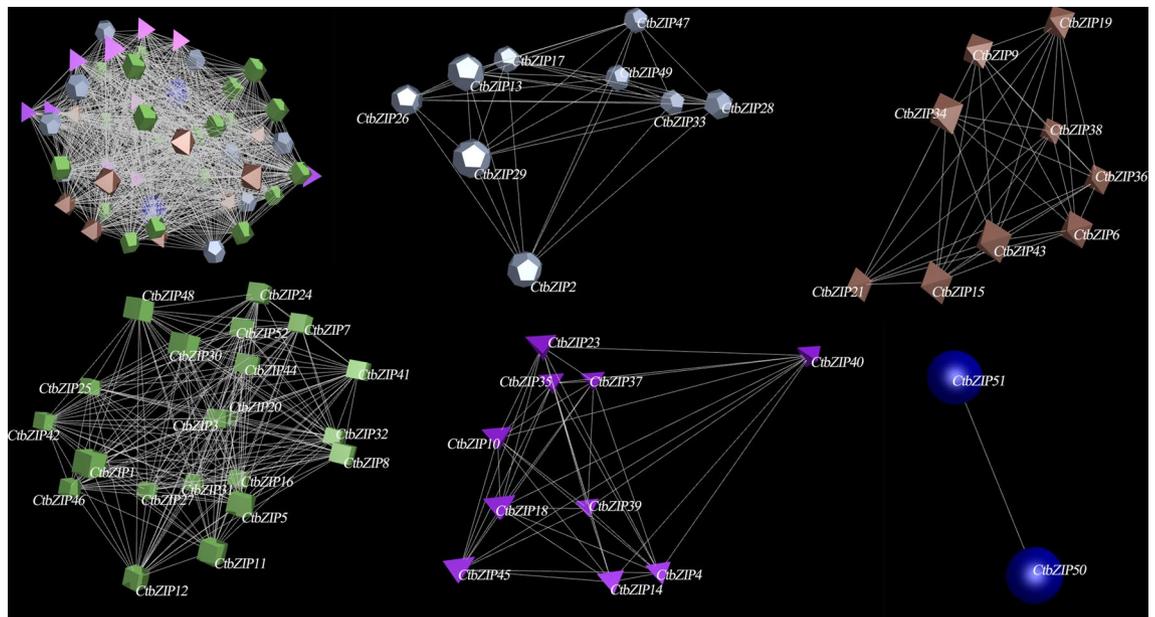


Figure 7. Network analysis of the CtbZIP TFs in seven different tissues of safflower. The co-expression network was constructed from the expression of 52 CtbZIP genes. The network was constructed at $P \leq 5.0 \times e^{-02}$. It consists of 51 gene nodes and 1,199 edges. Five clusters in the network are shown separately.

In the same race, we tried to unravel an important class of transcription factors in *Carthamus tinctorius* L. Transcription factors express genes thus are essentially present in all organisms. They comprise several classes holding fundamental role in various growth and developmental processes. The bZIP gene family plays role in plant growth and seed and fruit development³⁴. Plant bZIP TFs preferentially bind promoters containing ACGT core sequence including A- (TACGTA), C- (GACGTC) and G- (CAGTG) boxes³⁵, however nonpalindromic targets have also been reported^{36,37}. The genome-wide analysis of CtbZIPs would aid in their further functional analyses as well as safflower breeding research.

Our genomic survey identified 52 members in *Carthamus tinctorius* bZIP TF family. These TFs constitute a large families in all organisms reported to date. CtbZIPs also look a big gene family however, as compared to Arabidopsis (78 TFs), rice (89), maize (125), *Brassica napus* (247) and soybean (131), safflower got a relatively small bZIP family. Based on phylogenetic reconstruction (Fig. 1), we categorized CtbZIPs into 13 subfamilies (A-J, S and X) according to their relevance in Arabidopsis⁴, rice⁵, *Ricinus communis*²⁶ and *Camellia sinensis*²⁷. This categorization was further supported protein structure analyses. None of CtbZIP proteins clustered into subfamily K and M indicating loss of these proteins throughout safflower evolution.

CtbZIPs protein structure analyses revealed 20 motifs in total, same as reported in *Manihot esculanta*³⁸, which were named sequentially from motif1 to motif20 (Figs. 2, 3). Relating their motifs to some known motifs revealed some functions of CtbZIP TFs. The motif2 was further identified as the extension of the leu zipper region, closely related to motif1. The motif4 was a new highly conserved cysteine-rich sequence which might be involved in protein-protein or protein-DNA interactions. In most of the cases, motif1 and motif2 conserved domains are located next to each other, however, some motifs are located far from each other. The maximum distance between two motifs is found in CtbZIP45 of subfamily D. In addition, there are three motifs (motif4, 5, 13) between bZIP domains in subfamily E of CtbZIP TFs, and motif1 and motif2 together with three motifs form a conserved structural group, as the subfamily E of OsbZIPs⁵. The same situation exists in the subfamily I of CtbZIP TFs, motif1 and motif2 together with motif4, 5, 9, 17 form a conserved structural group, but motif9 is not between motif1 and motif2. The conserved groups of E and I subfamilies exist near the C-terminus which predicts that the functions of subfamily E and subfamily I could make a significant difference with other subfamilies. CtbZIP26 only contains the bZIP domain (motif1 and motif2) in the subfamily H, which confirms that the function of CtbZIP26 is more conservative. The motif11 in subfamily D is a conserved structure of Dog1 (PF14144) also found in *Arabidopsis* bZIP⁴. This family appears to be a highly specific controller of seed dormancy. On one hand, MEME results further prove that outcomes of Hidden Markov Model (HMM) have high reliability. On the other hand, they also reveal the functional diversity of CtbZIP family. These analyses are an important starting point for further functional verification.

The genome-wide expression prediction of CtbZIPs genes flaunted their differential transcript level in various developmental stages and tissues. As shown in Fig. 6, there seems a vast level of divergence in expression pattern with respect to tissue type and seed stage. The varied expression pattern indicates functional divergence of different groups of CtbZIP TFs, which predicts that the functions of CtbZIP family vary with variation in their expression. We quantified the expressions of all 52 CtbZIP TFs in different tissues and seeds (of various developmental stages). The network is composed of 5 clusters as shown in Fig. 7. There is a certain degree of related expression trend between these clusters and this tendency was statistically significant. These results indicate that although the functions and expressions of CtbZIP family members have dramatically diverged, they retain to some extent,

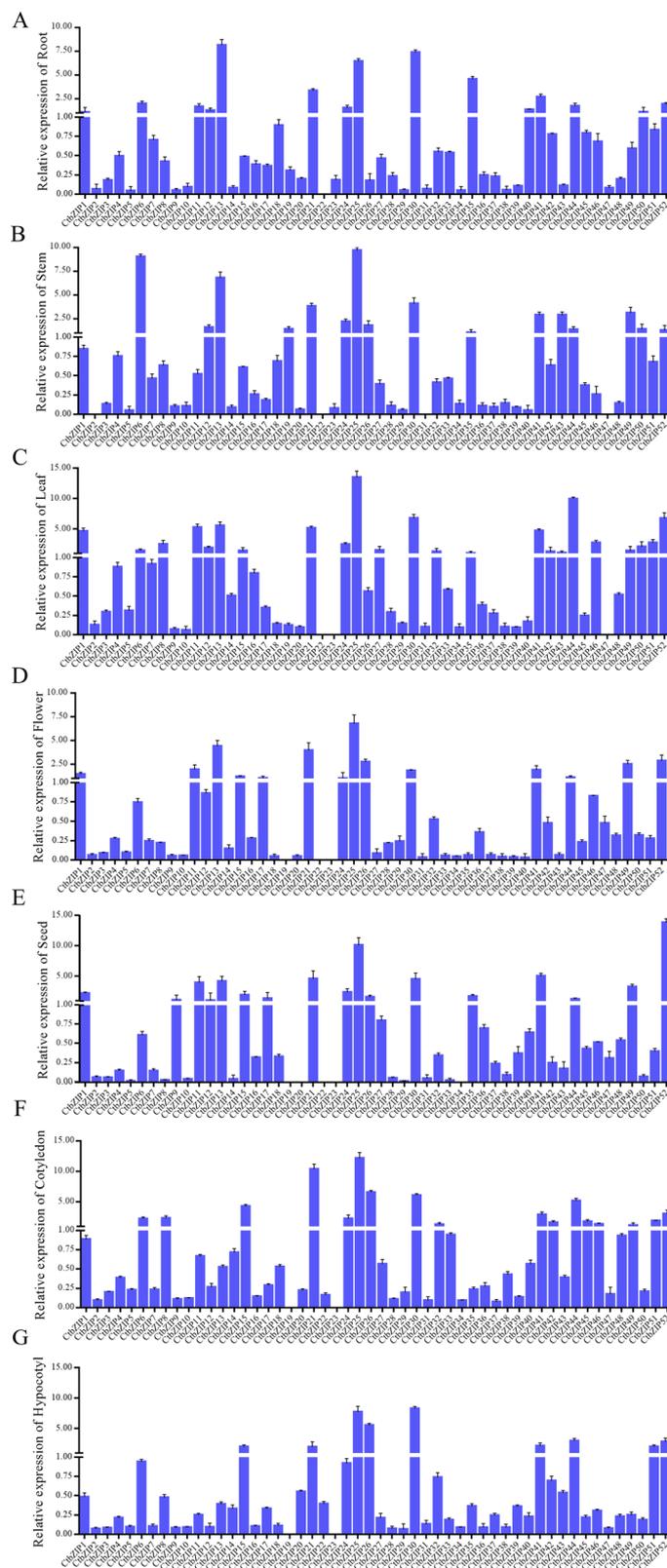


Figure 8. Relative expression profiles of the *CtbZIP* genes in various tissues. (A) Root, (B) stem, (C) leaf, (D) flower, (E) seed, (F) cotyledon, (G) hypocotyl. The reference gene used in RT-qPCR is EF1 α . Values are average of three replicates \pm SD. Asterisks indicate significant difference applying ANOVA ($p < 0.05$, $p < 0.01$ and $p < 0.001$).

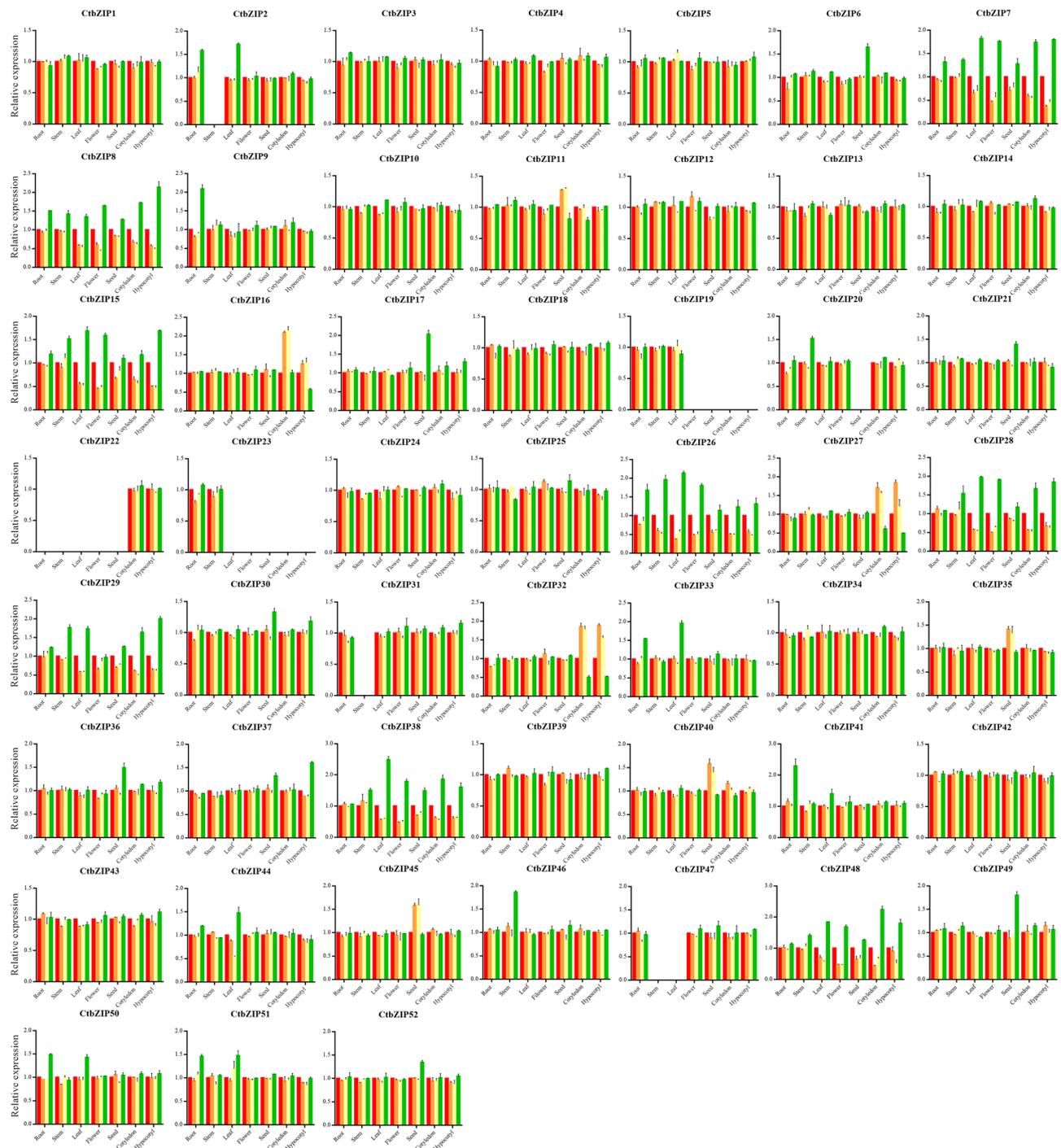


Figure 9. Expression profiles of *CtbZIP* genes after GA_3 treatment under different light radiation. The red bars indicate no treatment under 16.8 MJ/m^2 light radiation (group 1). The orange indicates no treatment under 5.04 MJ/m^2 light radiation (group 2). Yellow indicates GA_3 treatment under 5.04 MJ/m^2 light radiation (group 3). Green indicates GA_3 treatment under 16.8 MJ/m^2 light radiation (group 4). The reference gene was *EF1 α* . Values are average of three replicates \pm SD. Asterisks indicate significant difference applying ANOVA ($p < 0.05$, $p < 0.01$ and $p < 0.001$).

the tendency of correlated expression and functional cooperation. To verify the transcript abundance of *CtbZIPs* genes, we used RT-qPCR and evaluated their expression in root, stem, leaf, flower, seed, cotyledon and hypocotyl (Fig. 8). The results of RT-qPCR showed that the expression pattern of safflower is consistent with the predicted expression. According to this expression pattern, the function of *CtbZIP* TFs can be more effectively estimated.

In the process of plant growth and development, light and hormone are the key factors that directly affect these two processes. At present, it has been confirmed that the A subfamily bZIP members of *Arabidopsis thaliana* are mainly involved in ABA signaling³⁹ whereas H and G subfamilies regulate photoresponse^{14,40}. In rice,

OsZIP12 has been reported as a positive regulator of ABA signalling⁴¹ while in *Medicago esculenta*, *bZIP11*, 27, 52 and 64 were upregulated at time points of ABA treatment³⁸. In *Ipomoea trifida*, eight *bZIP* genes were upregulated at least in one tissue type as well as one time point, in response to ABA treatment⁴². *AtbZIP16* has been reported to regulate early development of seedling by integrating hormone and light signalling pathways thereby promoting germination as well elongation of hypocotyl⁴³. Under RL (Red Light) treatment, *ClabZIP6* and *ClabZIP56* were significantly induced while *ClabZIP37* and *ClabZIP22* were repressed in leaves of *Citrullus lanatus*⁴⁴. Figure 9 depicts that changes in expression of CtbZIPs under GA3 and light reveal that some of CtbZIP genes might be directly or indirectly affected by light intensity and hormones. These results provide a basis for further exploration of the function of CtbZIP TFs.

In summary, our study provides genome-wide analysis of the safflower *bZIP* family. We accurately screened 52 CtbZIP TFs, and divided them into 12 subclasses by identifying the conserved homology between them. Their basic physical and chemical properties were analyzed including ORF, number of amino acids and conserved structural positions. A total of 20 conserved structures are found in CtbZIP TFs family. All CtbZIP TFs contain a typical conserved *bZIP_1* domain. For the enrichment analysis of the CtbZIP TFs, we found that 45 of the 52 CtbZIPs were enriched, and among the 45, none of the genes were individually enriched into a certain GO functional category. Six CtbZIP TFs were enriched in three major categories CC, BP and MF, and 39 CtbZIP TFs are enriched in BP and MF. A total of four clusters within the CtbZIP TFs were discovered, which constitute a complex interplay network. The expression patterns of the CtbZIP family were predicted and verified by heat map and qRT-PCR. This study improves our understanding of safflower *bZIP* TFs and lays the foundation of cultivating new cultivars of safflower through molecular breeding methods.

Methods

Plant materials and treatments. The JiHong No. 1 safflower seeds purchased from safflower edge Co. Ltd. in Xinjiang of China, were cultivated in experimental field of Jilin Agricultural University for multiplication. The collected seeds of safflower were germinated in soil and allowed to grow at 23 ± 2 °C in growth room. It takes about 7 days to sprout cotyledons and hypocotyls, flowers in approximately 100 days while seeds in about 135 days. For light treatment, some safflower plants were grown under normal light radiation (16.8 MJ/m²) while another set of plants under weak light radiation (5.04 MJ/m²). For GA3 treatment, the plants that grew after flowering were sprayed with 50 mg/L GA3 once daily for 5 days. Each experimental group was sprayed simultaneously at 10 am. We collected various tissues, such as leaf, stem, root, flower, cotyledon, hypocotyl and seeds, immediately froze in liquid nitrogen and stored at -80 °C for further use.

Identification and characterization of CtbZIP TFs. The sequences of *CtbZIP* were obtained from the safflower genome database (Accession: PRJNA399628 ID: 399628). We downloaded HMM profile of *bZIP_1* (PF00170) from Pfam database²⁸ (<https://pfam.xfam.org/>) and the similar sequence of *bZIP_1* was searched using Hidden Markov Model (HMM) as the query ($P < 0.001$). To avoid missing possible *bZIP* members, NCBI BLAST was performed using the known *Arabidopsis bZIP* sequences (downloaded from the TAIR, <https://www.arabidopsis.org/>), as queries against the safflower genome database²⁶. All of the possible *bZIP* TFs were screened according to the significant e-value $< 1 \times 10^{-5}$ in our data. In addition, the conserved *bZIP* domains were predicted using SMART⁴⁵ (<https://smart.embl-heidelberg.de/>) and Search Pfam²⁸ (<https://pfam.xfam.org/search/sequence>) in all of the possible *bZIP* TFs. Therefore, the high-confidence *bZIP* TFs were screened, which were named as CtbZIP. Afterwards, we analyzed the physical and chemical properties of the predicted high-confidence CtbZIP TFs by ProtParam online tool⁴⁶ (<https://www.expasy.org/>).

Phylogenetic analysis of the CtbZIP proteins. The *bZIP* protein sequences of *Arabidopsis* and *Ricinus communis* were downloaded from database of PlantTFDB (<https://planttfdb.cbi.pku.edu.cn>) and that of rice were downloaded from the Rice Genome Annotation Project⁴⁷ (<https://rice.plantbiology.msu.edu/index.shtml>). Multiple alignment of the full-length *bZIP* sequences of safflower, *Arabidopsis*, rice and *Ricinus communis* was executed using Clustal X 2.0 program⁴⁸ and saved in the Clustal X file format. Using MEGA 7.0 program⁴⁹, we constructed a cladogram tree with 1,000 bootstrap replications and Neighbor-joining algorithm. The phylogenetic tree was modified using the iTOL online software⁵⁰ (<https://itol.embl.de/login.cgi>).

Motifs analysis of CtbZIP proteins. We searched the open reading frames of *CtbZIP* genes through the ORF finder at NCBI (<https://www.ncbi.nlm.nih.gov/gorf/gorf.html>). *CtbZIP* transcripts were analyzed in the Pfam²⁸ (<https://pfam.sanger.ac.uk/>) protein database. Analysis of the conserved motifs in safflower CtbZIP TFs were further carried out by multiple EM for motif elicitation software (MEME²⁹) (<https://meme.sdsc.edu/meme/cgi-bin/meme.cgi>) with default parameters. The maximum number of motifs was set to 20 and motif width to 6-50aa. Whereafter a conservative structure was generated using TBtools³⁰ (<https://www.tbtools.com/>). The related motif information used is listed in Table S2.

Gene ontology annotations of CtbZIP TFs. The functions of the CtbZIP TFs were categorized in silico using Blast2GO software³¹ (<https://www.blast2go.com/>). The GO functional categorization of 52 CtbZIP TFs was used into each subcategory for enrichment analysis. The enrichment of the number of CtbZIP transcripts categorized into each subcategory was determined by Chi-square test.

Network analysis of the CtbZIP TFs. The construction of the co-expression network is conceptually simple and intuitive. Through the similarity of gene expression, the possible interactions of gene products can

be analyzed to understand the intergenic interaction. The various traits are the result of genetic interactions. In order to excavate the network of interactions during *CtbZIP* genes family, we used the R programming language and software³¹ to calculate Pearson correlation coefficient. A gene co-expression network was constructed using BioLayout Express 3D Version 3.2 software³².

Gene expression patterns analysis. To investigate the *CtbZIP* gene family expression patterns, the high-throughput safflower transcriptome sequencing data were used to analyze the *CtbZIP* gene expression patterns in various tissues for roots, stems, leaves, flowers and DAF10, 13 and 20 seeds. The expression estimations of *CtbZIP* genes were normalized and represented in the form of RPKM (reads per kilo base per million mapped reads), and fold change (\log_2) values were calculated through the ratio of gene expression to draw heatmaps with R⁵¹ and TBtools³⁰ software.

RNA extraction and cDNA synthesis. The experimental materials (various tissues: root, stem, leaf, flower, seed, cotyledon, hypocotyl) were pulverized adequately and put into centrifuge tubes. Total RNA of various tissues was isolated using Trizol (Invitrogen, Carlsbad, CA, USA), according to the instructions of the manufacturer. The extracted total RNA was treated with RNase-free DNase (Promega, USA) to remove the genomic DNA contamination. RNA quality was checked on OD260/280 values by Nano Drop 2000 (ThermoFisher Scientific, Beijing, China) and 1.2% agarose gel electrophoresis. The cDNA was synthesized from total RNA isolated from various tissues using the PrimeScript RT reagent kit with gDNA Eraser (Takara, Japan), according to the manufacturer's protocols. First, 2 μ L 5 \times DNA Eraser buffer, 1 μ L gDNA Eraser, 2 μ L total RNA (about 1,000 ng) and 5 μ L RNase free ddH₂O were mixed in tube and incubated at 42 °C for 2 min to remove DNA. The purified RNA was reverse-transcribed into cDNA by adding 4 μ L 5 \times PrimeScript buffer, 1 μ L PrimeScript enzyme mix I, 1 μ L RT primer mix and 4 μ L RNase free ddH₂O into the above-mentioned reaction and incubated at 37 °C for 15 min followed by 85 °C for 15 s. The cDNA was stored at – 20 °C.

Real-time fluorogenic quantitative PCR. Real-time fluorogenic quantitative PCR (RT-qPCR) was performed using SYBR Premix Ex Taq II kit (Takara, Japan) and Stratagene Mx3000P thermocycler (Agilent) to monitor DNA products. The most stable housekeeping reference gene (*EF1 α*) was selected for the expression analysis in various tissues. The relative expression of *CtbZIP* was normalized to the expression of *EF1 α* and expressed relative to the level in various treatment. Gene-specific primers designed for the *CtbZIP* genes are listed in Table S3. RT-qPCR amplification was performed in 15 μ L reaction volume containing 500 ng template cDNA (1 μ L), 0.3 μ L primer (10 m), 7.5 μ L SYBR Premix Ex Taq (2 \times), 0.3 μ L ROX Reference Dye (10 m), and 5.6 μ L DEPC ddH₂O. RT-qPCR profile was set as an initial denaturation at 95 °C for 5 min, followed by 40 cycles of 95 °C for 5 s and annealing at 60 °C for 30 s. The fold change in relative expression level was calculated using 2^{– $\Delta\Delta$ CT} method.

Statistical analysis. The experiment was designated for three random replications. All data were analyzed by one-way analysis of variance (ANOVA) and all means were separated at the P < 0.05 level. The different tissues and GA3 treatment by the biological significance of the differential expression were analyzed.

Received: 24 December 2019; Accepted: 31 August 2020

Published online: 23 September 2020

References

- Meshi, T. & Iwabuchi, M. Plant transcription factors. *Plant Cell Physiol.* **36**, 1405–1420 (1995).
- Lee, B.-J., Park, C.-J., Kim, S.-K., Kim, K.-J. & Paek, K.-H. In vivo binding of hot pepper bZIP transcription factor CabZIP1 to the G-box region of pathogenesis-related protein 1 promoter. *Biochem. Biophys. Res. Commun.* **344**, 55–62 (2006).
- Cao, X. *et al.* Isolation and functional analysis of the bZIP transcription factor gene TaABP1 from a Chinese wheat landrace. *J. Integr. Agric.* **11**, 1580–1591 (2012).
- Droge-Laser, W., Snoek, B. L., Snel, B. & Weiste, C. The Arabidopsis bZIP transcription factor family—an update. *Curr. Opin. Plant Biol.* **45**, 36–49 (2018).
- Nijhawan, A., Jain, M., Tyagi, A. K. & Khurana, J. P. Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. *Plant Physiol.* **146**, 333–350 (2008).
- Liu, J. *et al.* Genome-wide analysis and expression profile of the bZIP transcription factor gene family in grapevine (*Vitis vinifera*). *BMC Genomics* **15**, 281 (2014).
- Wei, K. *et al.* Genome-wide analysis of bZIP-encoding genes in maize. *DNA Res.* **19**, 463–476 (2012).
- Zhou, Y. *et al.* Genome-wide identification and structural analysis of bZIP transcription factor genes in *Brassica napus*. *Genes (Basel)* **8**(10), 288–312 (2017).
- Wang, J. *et al.* Genome-wide expansion and expression divergence of the basic leucine zipper transcription factors in higher plants with an emphasis on SorghumF. *J. Integr. Plant Biol.* **53**, 212–231 (2011).
- Liao, Y. *et al.* Soybean GmbZIP44, GmbZIP62 and GmbZIP78 genes function as negative regulator of ABA signaling and confer salt and freezing tolerance in transgenic Arabidopsis. *Planta* **228**, 225–240 (2008).
- Izawa, T., Foster, R., Nakajima, M., Shimamoto, K. & Chua, N. H. The rice bZIP transcriptional activator RITA-1 is highly expressed during seed development. *Plant Cell* **6**, 1277–1287 (1994).
- Huang, X. *et al.* Arabidopsis FHY3 and HY5 positively mediate induction of COP1 transcription in response to photomorphogenic UV-B light. *Plant Cell* **24**, 4590–4606 (2012).
- Jakoby, M. *et al.* bZIP transcription factors in Arabidopsis. *Trends Plant Sci.* **7**, 106–111 (2002).
- Lindemose, S., O'Shea, C., Jensen, M. K. & Skriver, K. Structure, function and networks of transcription factors involved in abiotic stress responses. *Int. J. Mol. Sci.* **14**, 5842–5878 (2013).

15. Sornaraj, P., Luang, S., Lopato, S. & Hrmova, M. Basic leucine zipper (bZIP) transcription factors involved in abiotic stresses: a molecular model of a wheat bZIP factor and implications of its structure in function. *Biochim. Biophys. Acta* **1860**, 46–56 (2016).
16. Zander, M., La Camera, S., Lamotte, O., Metraux, J.-P. & Gatz, C. Arabidopsis thaliana class-II TGA transcription factors are essential activators of jasmonic acid/ethylene-induced defense responses. *Plant J.* **61**, 200–210 (2010).
17. Lee, J. *et al.* Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell* **19**, 731–749 (2007).
18. Mukherjee, K., Choudhury, A. R., Gupta, B., Gupta, S. & Sengupta, D. N. An ABRE-binding factor, OSBZ8, is highly expressed in salt tolerant cultivars than in salt sensitive cultivars of indica rice. *BMC Plant Biol.* **6**, 18 (2006).
19. Shimizu, H. *et al.* LIP19, a basic region leucine zipper protein, is a Fos-like molecular switch in the cold signaling of rice plants. *Plant Cell Physiol.* **46**, 1623–1634 (2005).
20. Bowers, J. E., Pearl, S. A. & Burke, J. M. Genetic mapping of millions of SNPs in safflower (*Carthamus tinctorius* L.) via whole-genome resequencing. *G3 (Bethesda)* **6**, 2203–2211 (2016).
21. Ghamarnia, H. & Gholamian, M. The effect of saline shallow ground and surface water under deficit irrigation on (*Carthamus tinctorius* L.) in semi arid condition. *Agric. Water Manag.* **118**, 29–37 (2013).
22. Chen, J. *et al.* Full-length transcriptome sequences and the identification of putative genes for flavonoid biosynthesis in safflower. *BMC Genomics* **19**, 548 (2018).
23. Li, H. *et al.* De novo transcriptome of safflower and the identification of putative genes for oleosin and the biosynthesis of flavonoids. *PLoS ONE* **7**, e30987 (2012).
24. Chi, M. *et al.* Genome-wide identification, expression profiling, and functional validation of oleosin gene family in *Carthamus tinctorius* L. *Front. Plant Sci.* **9**, 1–11 (2018).
25. Liu, X. *et al.* De novo sequencing and analysis of the safflower transcriptome to discover putative genes associated with safflower yellow in *Carthamus tinctorius* L. *Int. J. Mol. Sci.* **16**(10), 25657–25677 (2015).
26. Jin, Z., Xu, W. & Liu, A. Genomic surveys and expression analysis of bZIP gene family in castor bean (*Ricinus communis* L.). *Planta* **239**, 299–312 (2014).
27. Cao, H. *et al.* Isolation and expression analysis of 18 CsbZIP genes implicated in abiotic stress responses in the tea plant (*Camellia sinensis*). *Plant Physiol. Biochem. PPB* **97**, 432–442 (2015).
28. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2018).
29. Bailey, T. L., Williams, N., Misleh, C. & Li, W. W. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* **34**, W369–W373 (2006).
30. Chen, C. *et al.* TBtools - an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* <https://doi.org/10.1016/j.molp.2020.06.009> (2020).
31. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* **25**, 25–29 (2000).
32. Theocharidis, A., van Dongen, S., Enright, A. J. & Freeman, T. C. Network visualization and analysis of gene expression data using BioLayout Express (3D). *Nat. Protoc.* **4**, 1535–1550 (2009).
33. Mani, V., Lee, S.-K., Yeo, Y. & Hahn, B.-S. A metabolic perspective and opportunities in pharmacologically important safflower. *Metabolites* **10**(6), 253–271 (2020).
34. Wang, Y. *et al.* Identification and characterization of the bZIP transcription factor family and its expression in response to abiotic stresses in sesame. *PLoS ONE* **13**, e0200850–e0200850 (2018).
35. Izawa, T., Foster, R. & Chua, N.-H. Plant bZIP protein DNA binding specificity. *J. Mol. Biol.* **230**, 1131–1144 (1993).
36. Choi, H., Hong, J., Ha, J., Kang, J. & Kim, S. Y. ABFs, a family of ABA-responsive element binding factors. *J. Biol. Chem.* **275**, 1723–1730 (2000).
37. Fukazawa, J. *et al.* Repression of shoot growth, a bZIP transcriptional activator, regulates cell elongation by controlling the level of gibberellins. *Plant Cell* **12**, 901–915 (2000).
38. Hu, W. *et al.* Genome-wide characterization and analysis of bZIP transcription factor gene family related to abiotic stress in cassava. *Sci. Rep.* **6**, 22783 (2016).
39. Zhang, Z. *et al.* Genome-wide identification, expression profiling, and SSR marker development of the bZIP transcription factor family in *Medicago truncatula*. *Biochem. Syst. Ecol.* **61**, 218–228 (2015).
40. Hsieh, W.-P., Hsieh, H.-L. & Wu, S.-H. Arabidopsis bZIP16 transcription factor integrates light and hormone signaling pathways to regulate early seedling development. *Plant Cell* **24**, 3997–4011 (2012).
41. Joo, J., Lee, Y. H. & Song, S. I. Overexpression of the rice basic leucine zipper transcription factor OsbZIP12 confers drought tolerance to rice and makes seedlings hypersensitive to ABA. *Plant Biotechnol. Rep.* **8**, 431–441 (2014).
42. Yang, Z. *et al.* Genome-wide identification, structural and gene expression analysis of the bZIP transcription factor family in sweet potato wild relative *Ipomoea trifida*. *BMC Genet.* **20**, 41 (2019).
43. Shen, H., Cao, K. & Wang, X. AtbZIP16 and AtbZIP68, two new members of GBFs, can interact with other G group bZIPs in Arabidopsis thaliana. *BMB Rep.* **41**, 132–138 (2008).
44. Yang, Y. *et al.* The bZIP gene family in watermelon: genome-wide identification and expression analysis under cold stress and root-knot nematode infection. *PeerJ* **7**, e7878 (2019).
45. Letunic, I. & Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **46**, D493–D496 (2017).
46. Gasteiger, E. *et al.* ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788 (2003).
47. Kawahara, Y. *et al.* Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 4 (2013).
48. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
49. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
50. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
51. Team, R. C. R: A language and environment for statistical computing. *Computing* **1**, 12–21 (2014).

Acknowledgements

The research was supported by the grants from the National Natural Science Foundation of China. (No. 31801396). The fund of all reagents and materials for this research are provided by the National Natural Science Foundation of China.

Author contributions

H.Y.L. performed the experiments and completed the bioinformatics analysis of the CtbZIP TFs. L.X.L. and C.J. and S.N.D. analyzed the bioinformatics analysis of CtbZIP. G.D.S.G., Y.P.L. and L.H. determined the expression pattern of the CtbZIP family genes. L.N.D. and Y.X.G. and X.M.Z. analyzed the data. H.Y.L. (HaiYan Li) provided

instrument and equipment. J.Y. designed the experiments, acquired funding and wrote the manuscript. M.N. and N.A. revised the manuscript. Y.Y.D. screened gene sequences. All authors have read and approved the final draft.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-72390-z>.

Correspondence and requests for materials should be addressed to L.D., H.L. or J.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020