Genetics and Molecular Research

Genome-wide analysis of the maize (*Zea may* L.) *CPP-like* gene family and expression profiling under abiotic stress

X.Y. Song, Y.Y. Zhang, F.C. Wu and L. Zhang

Agro-Biotechnology Research Institute, Jilin Academy of Agricultural Sciences, Changchun, Jilin, China

Corresponding author: L. Zhang
E-mail: zly_jaas@126.com

Genet. Mol. Res. 15 (3): gmr.15038023
Received November 6, 2015
Accepted January 22, 2016
Published July 29, 2016
DOI http://dx.doi.org/10.4238/gmr.15038023

ABSTRACT. Cysteine-rich polycomb-like (CPP) proteins are members of a small family of transcription factors, which have been identified and characterized in *Arabidopsis*, rice, and soybean. In this study, we investigated *CPP-like* genes in the maize genome. The results revealed 13 putative *CPP-like* genes, which were found to encode 17 distinct transcripts and were distributed unequally on 7 of 10 maize chromosomes. Analysis of phylogenetic relationships showed that *Arabidopsis*, rice, and maize *CPP-like* transcription factors can be grouped into two subfamilies. We also used real-time RT-PCR to evaluate changes in the transcript levels of *ZmCPP* genes in response to abiotic stresses (heat, cold, salt, and drought stresses). These findings provide an overview of the evolution of the *ZmCPP* gene family, which will aid in the functional characterization of *CPP-like* genes in maize growth and development.

Key words: *CPP-like* gene family; CXC domain; Phylogenetic tree; Expression analysis; Maize

INTRODUCTION

Cysteine-rich polycomb-like (CPP) proteins are members of a small family of transcription factors that have important roles in the development of reproductive tissues and in the control of cell division in plants (Yang et al., 2008). Members of this family are widely present in plants and animals, but are absent in yeast (Andersen et al., 2007). Based on the function of the CPPs, the structure of genes in this family comprises one or two similar Cys-rich domains, termed the CXC domain, which is highly conserved across species, from amoeba through plants to mammals (Riechmann et al., 2000).

Although, CPP transcription factors are prevalent in a broad spectrum of species, they have received little attention compared with other transcription factor families. One study reported that 111 CPP transcription factors can be found in 16 species of plants (Lu et al., 2013), but only two of these transcription factors have been studied so far. The first CPP transcription factor described, *TSO1*, has been studied in *Arabidopsis thaliana*. *TSO1* is highly expressed in flowers, where it accumulates to the highest level in developing ovules and microspores. The Δ *tso1* mutation results in the loss of control of directional cellular expansion and the coordination of adjacent cell growth, and was also found to cause defects in karyokinesis and cytokinesis (Hauser et al., 1998, 2000; Song et al., 2000). In soybean, *CPP1*, a DNA-binding protein involved in the expression of the soybean *Gmlbc3* gene, was reported to be involved in the regulation of the leghemoglobin genes in the symbiotic root nodule (Cvitanich et al., 2000). In addition, *CPP-like* family genes have been identified and characterized in *Arabidopsis*, rice (Yang et al., 2008), and soybean (Zhang et al., 2015). These findings suggest that differences in transcriptional regulation of each *CPP-like* isoform may play distinct roles in plant growth, development, and stress responses.

Maize is one of the most important cereal crops in the world, which provides a stable food source for many populations. Nevertheless, maize growth and yield are strongly influenced by abiotic stresses such as drought, salt, and cold. Transcription factors regulate gene expression, thus providing plants with control mechanisms that enable them to respond to abiotic and biotic stresses, and to modulate developmental processes (Mitsuda and Ohme-Takagi, 2009). The *CPP-like* genes encode transcription factors that have diverse functions and may have been important in key events of plant diversification. Hence, comparative phylogenetic and molecular evolutionary analyses of the *CPP-like* gene family in maize are very useful for studying their roles in plant evolution.

In our study, we performed a genome-wide identification of members in the *CPP-like* gene family within maize. Further, the levels of maize *CPP-like* gene expression were determined by real-time quantitative reverse transcription-polymerase chain reaction (qRT-PCR) in response to abiotic stresses. Our findings will provide a useful reference for the selection of candidate *CPP-like* genes for cloning and further functional analyses in maize.

MATERIAL AND METHODS

Genome-wide identification of CPP-like proteins in maize

CPP-like family genes in maize, *A. thaliana*, and rice were identified from online databases (http://planttfdb.cbi.pku.edu.cn/) (Jin et al., 2014). Information on members of the maize *CPP-like* family of genes, including chromosomal locations, genomic sequences,

full-length coding sequences (CDS), and protein sequences were also downloaded from the website Phytozome (www.phytozome.net). The molecular weight (kDa) and isoelectric point (pI) of each gene were calculated using the compute pI/Mw tool from ExPASy (http://www.expasy.org/tools/) (Gasteiger et al., 2003). Subcellular localization was predicted using WoLF PSORT (http://wolfpsort.org/) (Horton et al., 2007) and TargetP 1.1 (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson et al., 2007).

Multiple alignment and phylogenetic analysis

Two phylogenetic trees were constructed to classify the maize *ZmCPP* genes. One phylogenic tree only contained maize ZmCPP protein sequences; the other also included *A. thaliana* and ZmCPP protein sequences. Multiple-sequence alignments of all predicted ZmCPP protein sequences were performed with the Clustal X2.0 software using default parameters. Phylogenetic trees were constructed using MEGA 5.0 with the neighbor-joining (NJ) method, and bootstrap analysis was conducted using 1000 replicates with a pairwise gap deletion mode (Tamura et al., 2011).

Chromosomal location of *ZmCPP*, structural analysis, and identification of conserved motifs

An image showing the chromosomal location of the *ZmCPP* genes was generated by the MapChart software, based on information provided in the Phytozome database. The structure of *ZmCPP* was analyzed using Gene Structure Display Server (GSDS; http://gsds.cbi.pku.edu.cn/) (Guo et al., 2007). In addition, protein sequences were used to predict domain and motifs online using MEME (http://meme-suite.org) (Bailey et al., 2006). The parameters were as follows: number of repetitions: any; maximum number of motifs: 10; and optimum motif widths: 6-200-amino acid residues.

Plant growth and stress treatments

Maize (*Zea mays* L. inbred line B73) plants were grown in a greenhouse at 28° ± 2°C with a photoperiod of 14-h light and 10-h dark. When the seedlings developed three fully opened trifoliate leaves (approximately 3 weeks after sowing), we performed abiotic stress experiments. For heat and cold stress, the seedlings were transferred to a growth chamber at 42° ± 1°C and 4° ± 1°C, respectively. For drought stress, 15% (w/v) PEG3000 was used. In addition, 300 mM NaCl was used to generate salt stress. After the leaves were subjected to these stresses, they were collected at 0-, 3-, 6-, 12-, and 24-h intervals. After collection, the samples were immediately frozen in liquid N_2, and stored at -80°C for RNA extraction. Three biological replicates were conducted per sample.

RNA extraction and qRT-PCR analysis

Total RNA was isolated from all samples using TRIzol reagent (Invitrogen, USA), followed by DNaseI (Promega, USA) treatment to remove any traces of genomic DNA. First-strand cDNA was synthesized using the TransScript First-Strand cDNA Synthesis SuperMix Kit (TransGen, China). qRT-PCR was conducted using an Applied Biosystems StepOne

Real-Time PCR System (Applied Biosystems, USA). The total reaction volume was 20 μL, which included 10 μL SYBR premix Ex Taq™ (2X) mixture; 1 μL cDNA (diluted 10 times); 0.4 μL upstream primer (10 pm); 0.4 μL downstream primer (10 pM); and 8.2 μL ddH_2O. The reaction was performed using the following cycling profile: 95°C for 30 s, 40 cycles of denaturation at 95°C for 5 s, and 60°C for 30 s. Three technical replicates were performed for each sample. Levels of gene expression were calculated using the $2^{-\Delta\Delta Ct}$ method as previously described (Livak and Schmittgen, 2001). *ZmActin* (NM_001155179.1) was used as a positive internal control and 13 primers specific for *ZmCPP* genes were designed for use in qRT-PCR using conserved sequences as listed in **Table S1**.

RESULTS

Identification and prediction of subcellular localization of CPP

We identified 13 *CPP-like* gene family members (*ZmCPP1-ZmCPP13*), which encoded 17 transcripts. Basic genome information, containing protein sequences (**Figure S1**), CDSs (**Figure S2**), genomic sequences (**Figure S3**), and 2000 bp of nucleotide sequences (**Figure S4**), was downloaded from the Phytozome database (http://www.phytozome.com). Basic information on all of the maize *CPP-like* genes (including gene name, chromosome location, open reading frame length, exon and intron number, protein length, molecular weight, and pI value) is showed in Table 1. Large variations in the length of the amino acid sequence of these genes were found, ranging from 235 (*ZmCPP13.2*) to 800 amino acids (*ZmCPP10*), and the molecular weights were 25.45-84.98 kDa. The predicted pI of the *ZmCPP* candidates were between 5.86 (*ZmCPP1.1* and *ZmCPP1.2*) and 9.84 (*ZmCPP13.2*). WoLF PSORT assessment revealed that within the 17 ZmCPP proteins, 13 maize CPP proteins were localized to the nucleus, three to the chloroplast, and one was extracellular. TargetP analysis revealed that four maize CPP-like proteins were located in the chloroplast, two in the mitochondria, one in the secretory pathway, and 10 in other compartments. These detailed parameters are provided in Table 1.

Table 1. List of 13 *CPP-like* genes identified in maize, their sequence characteristics, and subcellular localization.

No.	Name	Feature name	Chr.	Location coordinates (5'-3')	ORF length (bp)	Protein length (aa.)	Molecular weight (kDa)	Theoretical pI	Subcellular location	
									WoLF PSORT	TargetP
1	ZmCPP1.1	GRMZM2G322090_P01	1	171,296,402-171,301,898	1948	587	63.521	5.86	nucl: 14	-
2	ZmCPP1.2	GRMZM2G322090_P02	1	171,296,402-171,301,738	1898	587	63.52	5.86	nucl: 14	-
3	ZmCPP2	GRMZM2G366584_P01	1	252,868,475-252,877,327	2474	662	71.464	6.18	nucl: 14	C
4	ZmCPP3	GRMZM2G015097_P01	1	171,249,195-171,254,764	1843	394	42.985	7.76	nucl: 13	C
5	ZmCPP4	AC203865.3_FGP001	3	119,995,077-119,998,350	1185	394	42.460	6.61	nucl: 13	-
6	ZmCPP5	GRMZM2G701689_P01	3	119,999,294-120,014,210	3066	772	83.066	7.19	nucl: 14	-
7	ZmCPP6	GRMZM2G059678_P01	3	195,307,506-195,312,044	1307	356	39.109	9.14	chlo: 10, nucl: 4	-
8	ZmCPP7.1	GRMZM2G388148_P01	5	134,229,441-134,245,208	2247	530	57.878	6.08	nucl: 10, chlo: 4	M
9	ZmCPP7.2	GRMZM2G388148_P02	5	134,229,441-134,244,350	1860	485	53.255	6.69	nucl: 9, chlo: 5	M
10	ZmCPP8	GRMZM2G342588_P01	6	122,278,172-122,285,658	2005	499	55.124	7.45	nucl: 14	-
11	ZmCPP9	GRMZM2G060170_P04	6	158,013,048-158,016,012	1133	359	39.940	8.29	nucl: 14	C
12	ZmCPP10	GRMZM2G153754_P01	7	9,332,269-9,337,505	2706	800	84.981	6.09	nucl: 14	C
13	ZmCPP11.1	GRMZM2G096600_P01	8	121,748,701-121,752,582	1330	341	38.067	8.69	nucl: 14	-
14	ZmCPP11.2	GRMZM2G096600_P02	8	121,748,701-121,752,565	1358	356	39.672	8.41	nucl: 14	-
15	ZmCPP12	GRMZM2G173198_P01	8	121,766,217-121,767,657	1159	298	33.509	8.53	extr: 6, chlo: 3, nucl: 3, cyto: 2	S
16	ZmCPP13.1	GRMZM2G104246_P01	8	172,942,841-172,947,390	1546	412	44.688	9.32	chlo: 7, nucl: 5, cyto: 2	-
17	ZmCPP13.2	GRMZM2G104246_P02	8	172,942,841-172,944,976	1044	235	25.454	9.84	chlo: 10, nucl: 2, cyto: 1	-

WoLF PSORT predictions: chlo (chloroplast), cyto (cytosol), nucl (nucleus), extr (extracellular). TargetP predictions: C (chloroplast), M (mitochondrion), S (secretory pathway), - (any other location); values indicate score (0.00-1.00) and reliability class (1-5; best class is 1).

Chromosomal location, gene structure, and motif analysis of CPPs in maize

The 13 putative *ZmCPP* gene candidates were distributed across 6 of the 10 chromosomes in the maize genome. Among them, chromosomes 1, 3, and 8 each had three *ZmCPP* genes. Two *ZmCPP* genes were located on each of chromosomes 5 and 6; one *ZmCPP* gene was situated on chromosome 7; and no *ZmCPP* genes were detected on chromosome 2, 4, 9, or 10 (Figure 1). We also investigated the structure of genes within the *ZmCPP* family, and the results showed that the 17 *ZmCPP* transcripts were categorized into four classes based on the unrooted phylogenetic tree. Class I was the largest group, which contained six ZmCPP proteins. The second was class II (containing five members), followed by classes III and IV, which both contained three proteins (Figure 2A). Figure 2B provides a detailed illustration of the relative lengths of the introns, and the conservation of the corresponding exon sequences within each *CPP* gene of maize by alignment of the cDNA to genomic sequences. This sequence analysis revealed introns within the coding sequences of all *CPP* genes, with the number of exons varying from 3 to 10 (Figure 2B). Conserved motifs were predicted using the MEME motif detection software to reveal the diversification of the ZmCPP proteins. The details of the 10 putative motifs are shown in Table 2, and the motif matches shown have a positional P value of less than 0.0001 (Figure 3). All ZmCPP proteins contained motif 1, and the type, order, and number of motifs were similar in proteins from the same subfamily, but differed from those in other subfamilies.

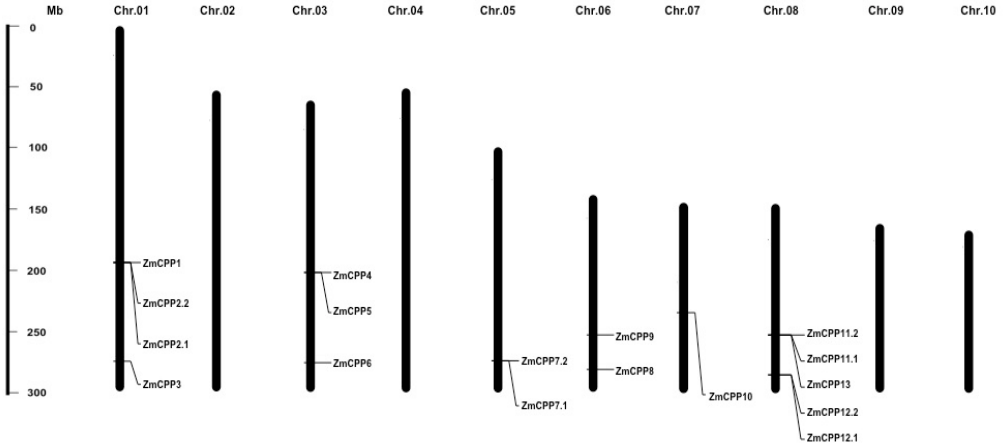

Figure 1. Chromosomal location of 17 *ZmCPP* transcripts on the 10 maize chromosomes. The scale represents Mb. The chromosome numbers are indicated at the top of each bar.

Numbers correspond to the motifs described in Figure 3. Sequences obtained from the analysis of three groups of maize CPP complete proteins using the MEME tools.

Comparative analysis of the *ZmCPP* genes in maize, *Arabidopsis*, and rice

An NJ phylogenetic tree was constructed using 46 full-length protein sequences in order to reveal the evolutionary relationships among maize (17), *Arabidopsis* (9), and rice (20)

CPP-like proteins (Figure 4 and **Table S2**). All CPP proteins fell broadly into two major classes: classes A and B, with well-supported bootstrap values, which included representative genes of maize, rice and *Arabidopsis*. Classes A and B were further subdivided into three subclasses according to their bootstrap values and phylogenetic relationships, designated as A1, A2, A3, B1, B2, and B3, respectively. The combined phylogenetic tree also revealed paralogous and orthologous relationships among the CPP-like family members. Seventeen pairs of paralog proteins were found between species, five pairs of paralogous genes in maize, eight pairs in rice, and four pairs in *Arabidopsis* (**Table S3**). Furthermore, three pairs of orthologous genes from maize and rice were identified, which were *ZmCPP10* and *OsCPP8*, *ZmCPP2* and *OsCPP2* in subfamily A3, and *ZmCPP8* and *OsCPP7* in subfamily B3.

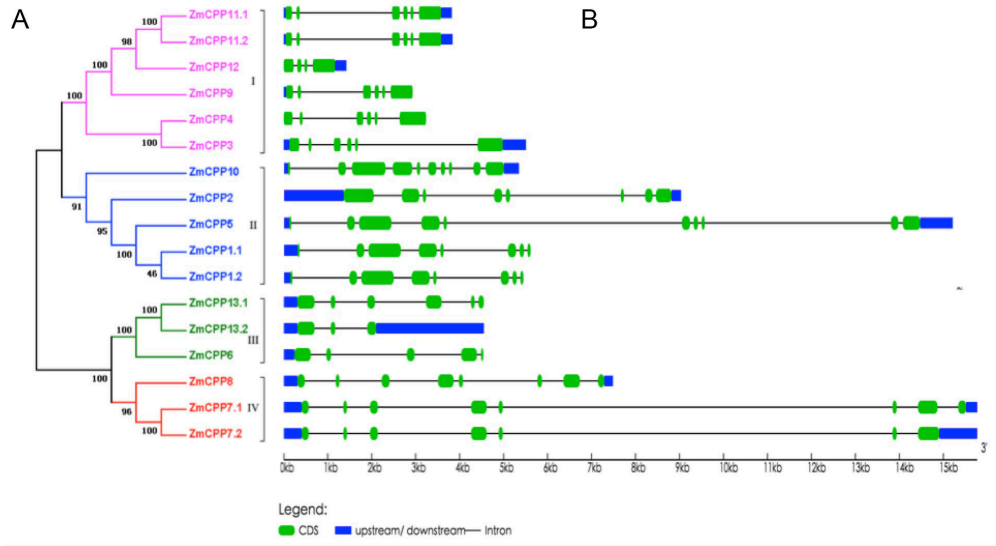

Figure 2. Phylogenetic relationships and gene structure of maize *CPP-like* genes. **A.** Unrooted tree generated with the MEGA5.0 program. Subfamilies of maize *CPP-like* genes (I-IV) are highlighted with different colored backgrounds and vertical bars next to the gene names of the tree. **B.** Exon/intron organization of maize *CPP-like* genes.

Table 2. Motif sequences of *CPP-like* genes identified in maize using MEME tools.

Motif	Width	Best possible match
1	29	CKHCNCKKSRCLKKYCECFQGGVYCSENC
2	50	CVQILNGMVELSQVEKSVAPDVFLLPGNREIFVSLGGDVRAMWLKRKIQH
3	50	ESSFHQTPPHLRASSRDAHVFPQAVSQWQALPRSWHCSNKRNGNDRAMDD
4	27	GCQGCFNKEAHMETVQTTRKQIESRNP
5	41	CTNCFNNVENEVARREAIEAIRERNPDAFRPKIGNDPHTNR
6	21	RCMDCKNFFGKKEGIILDQVD
7	50	MHNDESGPDRMDTSHSFIMIHENQIVEQNNDEPAMYNEQYIITHHHTSDM
8	29	YLYTGADLDHSEGEHDFVVERSPRLQSPM
9	29	RMSMKINRRPEANAEPMEDAHHSSSSTPP
10	50	SGYCTQNSVHEPHLYWTGAVEGSAVSYTPQTLPGALQSQLMPCNKLSEPK

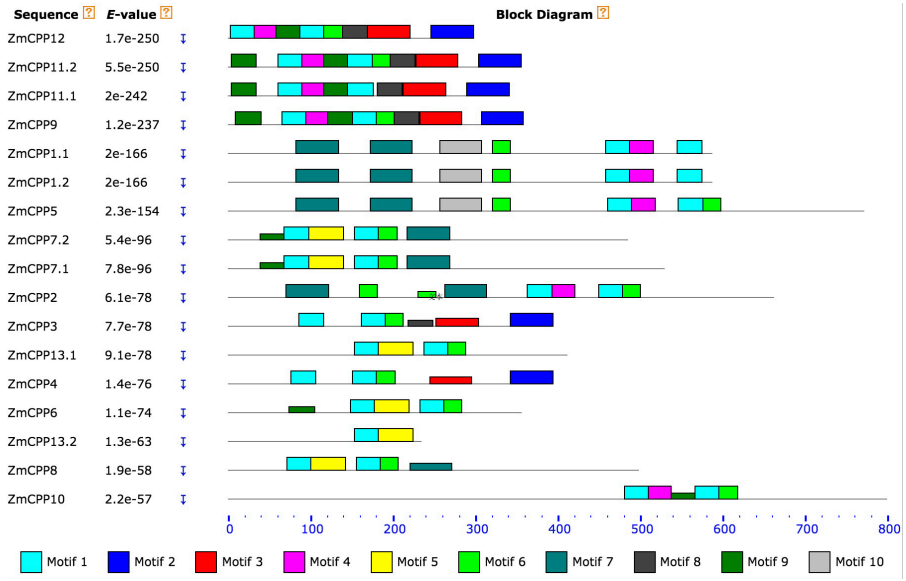

Figure 3. Schematic representation of the conserved motifs in maize CPP-*like* proteins elucidated from publicly available data. Each colored box represents a motif in the protein, with the motif name indicated in the box along the bottom.

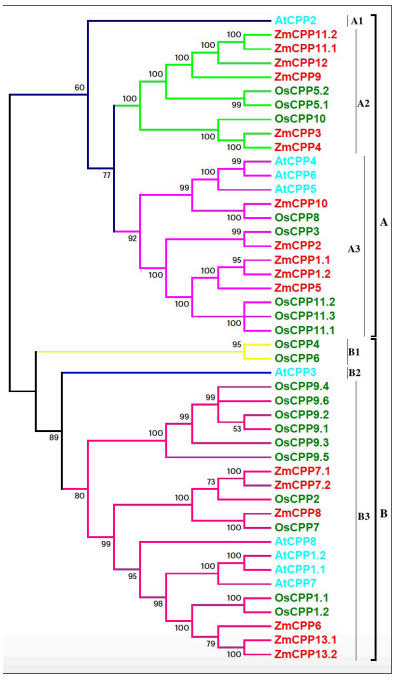

Figure 4. Phylogenetic tree of full-length CPP-like proteins from maize, *Arabidopsis*, and rice. The 17 maize, 9 *Arabidopsis*, and 20 rice CPP-like protein sequences were aligned in Clustal X 1.83 and the phylogenetic tree was constructed using MEGA5.0 by the neighbor-joining (NJ) method. The bootstrap value was 1000 replicates. Each Hsp70 subfamily is indicated by a specific color.

qRT-PCR analysis of maize *CPP-like* gene expression under different abiotic stresses

To gain further insight into the roles of maize CPP in abiotic tolerance, the expression profiles of 13 maize *CPP-like* genes in response to drought, heat, cold, and salt stresses were analyzed by qRT-PCR. The analysis revealed that these genes are differentially expressed in the leaves under different stress conditions (Figure 5). Under heat stress (42°C), the expression of most genes was dramatically upregulated (>7-fold) after a 12-h treatment, with the exception of *ZmCPP5* (Figure 5A). Drought stress (15% PEG) resulted in the upregulation (>3-fold) of four genes (*ZmCPP1.1*, *ZmCPP7.1*, *ZmCPP9*, *ZmCPP12*) after a 12-h treatment (Figure 5B). In response to salt stress, there were no obvious changes in gene expression, with only four genes (*ZmCPP5*, *ZmCPP71*, *ZmCPP8*, *ZmCPP10*) showing weak upregulation (<4.5-fold) at some stages (Figure 5C). Under cold stress (4°C), most of the *ZmCPP* genes were significantly upregulated (>10-fold) after a 12-h treatment (Figure 5D), with the exception of *ZmCPP5*.

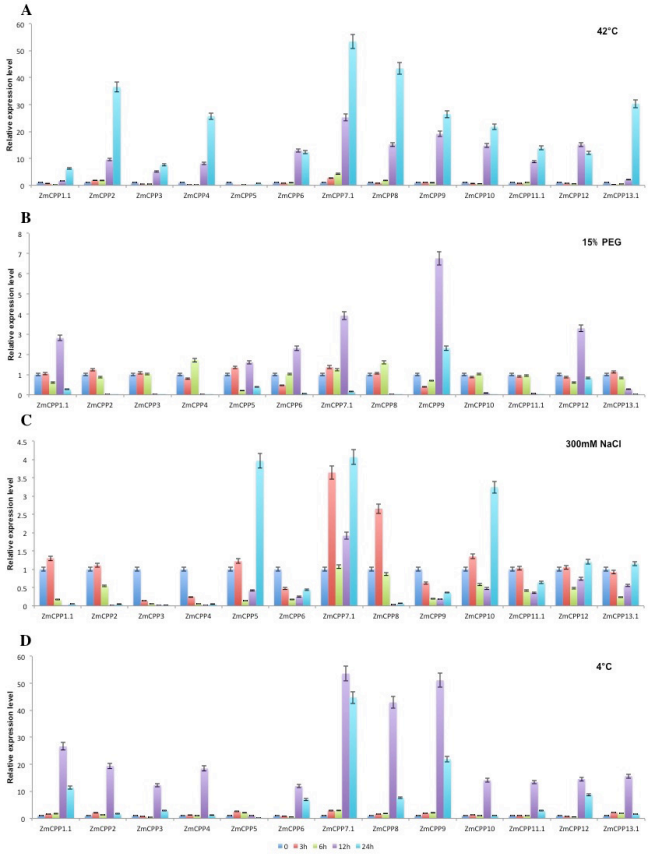

Figure 5. Expression patterns of 13 maize *CPP-like* genes under abiotic stresses determined using qRT-PCR. Relative transcript levels were calculated by real-time PCR using *Actin* as an internal standard. Bars represent standard deviations (SD) of three biological replicates. Y-axes indicate the scale of the relative expression levels. X-axes show the time courses of abiotic stress treatments for each gene.

DISCUSSION

Preliminary analysis of the *CPP-like* gene family has been performed in the model plants *Arabidopsis* and rice. Eight members of this family were found in *Arabidopsis* and 11 in rice (Yang et al., 2008). A recent study identified 20 *CPP-like* genes in soybean (Zhang et al., 2015). The results of the present study expand the number of known *CPP-like* genes in maize to 13, and the 13 putative *ZmCPP* gene candidates were found to be distributed across 6 of the 10 chromosomes in the maize genome. We also performed an analysis of the *CPP-like* gene family in maize, including analysis of their phylogeny, chromosomal location, gene structure, conserved motifs, and expression profiles. Based on these analyses, we discovered that the most closely related members of this family within the same subfamilies share similar exon/intron structures and intron numbers, which are consistent with the characteristics defined in the phylogenetic analysis (Figure 2). By analyzing their subcellular localization, we found that the nucleus might be the primary site of ZmCPP protein activity.

Phylogenetic analysis of CPP proteins in maize, rice, and *Arabidopsis* indicated that these genes are mainly classified into two groups, with 23 genes, respectively. In Group A, there are 4 *Arabidopsis* genes, 8 rice genes, and 11 maize genes, while in Group II there are 4 *Arabidopsis* genes, 12 rice genes, and 6 maize genes. Genes within the same group might share similar functions because of their similar sequence profiles. For example, *AtCPP4* (*SOL1*) and *AtCPP5* (*TSO1*) are both clustered in Class A3, and previous studies have shown that these two genes share the same function that controls flower tissue development (Hauser et al., 1998, 2000; Song et al., 2000). This indicates that proteins from different groups are highly distinguishable by their protein sequences, which reflect distinct functions.

Gene duplications are crucial during plant evolution and are important for driving genome evolution (Kent et al., 2003; Zhang, 2003). Paralogous genes typically exhibit different functions, while orthologous genes may retain the same functions (Conant et al., 2007). Previous research estimated that the fraction of retained paralogs is 72% in maize, having occurred over the course of 11 million years of evolution (Ahn and Tanksley, 1993). In our study, 15 pairs of paralog *CPP* genes were identified and three pairs of orthologous genes were found in the *Arabidopsis*, rice, and maize genomes; these results indicate that these gene families extended in a species-specific manner (**Table S1**). This phenomenon has been widely verified in other gene families in plants (Bai et al., 2002; Zhang et al., 2005; Jain et al., 2006).

Expression analysis of 13 *ZmCPP* genes by qRT-PCR showed that maize *CPP* genes are differentially expressed in response to heat, drought, salt, and cold stress. All of these genes responded significantly to 2-4 kinds of stress, indicating that they would also be involved in the physiological processes associated with multiple-stress responses in maize. It is interesting to note that the *ZmCPP* genes showed similar levels of expression under heat and cold stress, with differences in most of these genes peaking after 24 h of heat stress and 12 h of cold stress. Furthermore, we found no changes in the expression of the *ZmCPP* gene under salt stress, which may illustrate that the *CPP* family of transcription factors is not regulated under these conditions. Our findings indicate that the *CPP-like* genes are not only involved in growth and development, but also in responses to high and cold temperature stress in maize. However, their detailed roles in stress responses require further study.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#31301329, #31501327), the Ministry of Agriculture of China for Transgenic Research (#2016ZX08011-003), and the Youth Scientific Research Foundation of Jilin (#20150520123JH).

REFERENCES

Ahn S and Tanksley SD (1993). Comparative linkage maps of the rice and maize genomes. *Proc. Natl. Acad. Sci. USA* 90: 7980-7984. http://dx.doi.org/10.1073/pnas.90.17.7980

Andersen SU, Algreen-Petersen RG, Hoedl M, Jurkiewicz A, et al. (2007). The conserved cysteine-rich domain of a tesmin/TSO1-like protein binds zinc *in vitro* and TSO1 is required for both male and female fertility in *Arabidopsis thaliana*. *J. Exp. Bot.* 58: 3657-3670. http://dx.doi.org/10.1093/jxb/erm215

Bai J, Pennill LA, Ning J, Lee SW, et al. (2002). Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Res.* 12: 1871-1884. http://dx.doi.org/10.1101/gr.454902

Bailey TL, Williams N, Misleh C and Li WW (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34: W369-73. http://dx.doi.org/10.1093/nar/gkl198

Conant GC, Wagner GP and Stadler PF (2007). Modeling amino acid substitution patterns in orthologous and paralogous genes. *Mol. Phylogenet. Evol.* 42: 298-307. http://dx.doi.org/10.1016/j.ympev.2006.07.006

Cvitanich C, Pallisgaard N, Nielsen KA, Hansen AC, et al. (2000). CPP1, a DNA-binding protein involved in the expression of a soybean leghemoglobin c3 gene. *Proc. Natl. Acad. Sci. USA* 97: 8163-8168. http://dx.doi.org/10.1073/pnas.090468497

Emanuelsson O, Brunak S, von Heijne G and Nielsen H (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2: 953-971. http://dx.doi.org/10.1038/nprot.2007.131

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, et al. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 31: 3784-3788. http://dx.doi.org/10.1093/nar/gkg563

Guo AY, Zhu QH, Chen X and Luo JC (2007). GSDS: a gene structure display server. *Yi Chuan* 29: 1023-1026. http://dx.doi.org/10.1360/yc-007-1023

Hauser BA, Villanueva JM and Gasser CS (1998). *Arabidopsis* TSO1 regulates directional processes in cells during floral organogenesis. *Genetics* 150: 411-423.

Hauser BA, He JQ, Park SO and Gasser CS (2000). TSO1 is a novel protein that modulates cytokinesis and cell expansion in *Arabidopsis*. *Development* 127: 2219-2226.

Horton P, Park KJ, Obayashi T, Fujita N, et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35: W585-7. http://dx.doi.org/10.1093/nar/gkm259

Jain M, Tyagi AK and Khurana JP (2006). Genome-wide analysis, evolutionary expansion, and expression of early auxin-responsive SAUR gene family in rice (Oryza sativa). *Genomics* 88: 360-371. http://dx.doi.org/10.1016/j.ygeno.2006.04.008

Jin J, Zhang H, Kong L, Gao G, et al. (2014). PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* 42: D1182-D1187. http://dx.doi.org/10.1093/nar/gkt1016

Kent WJ, Baertsch R, Hinrichs A, Miller W, et al. (2003). Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl. Acad. Sci. USA* 100: 11484-11489. http://dx.doi.org/10.1073/pnas.1932072100

Livak KJ and Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25: 402-408. http://dx.doi.org/10.1006/meth.2001.1262

Lu T, Dou Y and Zhang C (2013). Fuzzy clustering of CPP family in plants with evolution and interaction analyses. *BMC Bioinformatics* 14 (Suppl 13): S10. http://dx.doi.org/10.1186/1471-2105-14-S13-S10

Mitsuda N and Ohme-Takagi M (2009). Functional analysis of transcription factors in *Arabidopsis*. *Plant Cell Physiol.* 50: 1232-1248. http://dx.doi.org/10.1093/pcp/pcp075

Riechmann JL, Heard J, Martin G, Reuber L, et al. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290: 2105-2110. http://dx.doi.org/10.1126/science.290.5499.2105

Song JY, Leung T, Ehler LK, Wang C, et al. (2000). Regulation of meristem organization and cell division by TSO1, an *Arabidopsis* gene with cysteine-rich repeats. *Development* 127: 2207-2217.

Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.

http://dx.doi.org/10.1093/molbev/msr121
Yang Z, Gu S, Wang X, Li W, et al. (2008). Molecular evolution of the CPP-like gene family in plants: insights from comparative genomics of *Arabidopsis* and rice. *J. Mol. Evol.* 67: 266-277. http://dx.doi.org/10.1007/s00239-008-9143-z
Zhang J (2003). Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18: 292-298. http://dx.doi.org/10.1016/S0169-5347(03)00033-8
Zhang L, Zhao HK, Wang YM, Yuan CP, et al. (2015). Genome-wide identification and expression analysis of the CPP-like gene family in soybean. *Genet. Mol. Res.* 14: 1260-1268. http://dx.doi.org/10.4238/2015.February.13.4
Zhang S, Chen C, Li L, Meng L, et al. (2005). Evolutionary expansion, gene structure, and expression of the rice wall-associated kinase gene family. *Plant Physiol.* 139: 1107-1124. http://dx.doi.org/10.1104/pp.105.069005

Supplementary material

Figure S1. Protein sequence data of the maize *CPP-like* genes.

Figure S2. Coding sequence data of the *CPP-like* gene superfamily in maize.

Figure S3. Genomic sequence data of the *CPP-like* gene superfamily in maize.

Figure S4. Data showing 2000 bp of the nucleotide sequence upstream of the translation initiation codon of the *CPP-like* gene superfamily in maize.

Table S1. List of primer sequences used for qRT-PCR analysis of the *ZmCPP* genes.

Table S2. Accession No. of *CPP-like* genes from *Arabidopsis*, and rice that were used to construct the neighbor-joining phylogenetic tree.

Table S3. Paralogous pairs of *CPP-like* genes in maize, rice, and *Arabidopsis.*