Comprehensive Genome-wide Identification, Characterization, and Expression Analysis of CCHC Zinc Finger Gene Family in Wheat (*Triticum aestivum* L.)

Aolong Sun
Hunan University
Yongliang Li
Hunan University
Xiaoxiao Zou
Hunan University
Fenglin Chen
Hunan University
Ruqiong Cai
Hunan University
Hai Xie
Hunan University
Wenjun Xiao
Hunan University
Xinhong Guo (gxh@hnu.edu.cn)
Hunan University

Research Article

Keywords: Wheat, CCHC-ZFP genes, Evolution of gene structure, Abiotic stress, Expression analyses

Posted Date: January 3rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-900125/v2

Abstract

Background: The CCHC zinc finger proteins (CCHC-ZFPs) are transcription factors that play versatile roles in plant growth, development, and responses to biotic/abiotic stress. However, little is known about the *CCHC-ZF* genes in bread wheat (*Triticum aestivum*), an important food crop.

Results: In this study, 50 *TaCCHC-ZF* genes were identified and distributed unevenly on 21 wheat chromosomes. According to the phylogenetic features, the 50 *TaCCHC-ZF* genes were classified into eight groups with specific motifs and gene structures. 43 *TaCCHC-ZF* genes were identified as segmentally duplicated genes that formed 36 segmental duplication gene pairs. Additionally, the collinearity analyses between wheat and eight other representative plant species showed that wheat had closer phylogenetic relationships with monocots compared to dicots. A total of 636 cis-elements related to environmental stress and phytohormone responsiveness were identified in the promoter of *TaCCHC-ZF* genes. Moreover, GO enrichment results revealed that all 50 *TaCCHC-ZF* genes were annotated under metal ion binding and nucleic acid binding. 91 miRNA binding sites within the 34 *TaCCHC-ZF* genes were identified by miRNA targets analyses, indicating that the expression of *TaCCHC-ZF* genes could be regulated by the miRNAs. Based on published transcriptome data, 38 *TaCCHC-ZF* genes were identified as DEGs, and 15 *TaCCHC-ZF* genes among them were verified by qRT-PCR assays, which showed response to drought, heat, or simultaneous response of them.

Conclusions: This study systematically explored the gene structures, evolutionary characteristics, and potential roles during environmental responses of *TaCCHC-ZF* genes, providing a foundation for further investigation and application of *TaCCHC-ZF* genes in the molecular breeding of *T. aestivum*.

Background

The CCHC-ZFPs are one of the most diverse and largest transcription factor families in plants, which play important roles in multiple physiological processes. The CCHC-ZFPs regulate the expression of their target genes by directly or indirectly recognizing and binding the promoters [1, 2]. As a member of zinc finger proteins, the CCHC-ZFPs contain at least one CCHC motif, which is also called zinc knuckle, sharing the consensus sequence $CX_2CX_4HX_4C$ (X for any amino acid, numbers for the number of residues, C and H for cysteine and histidine, respectively) [3]. The CCHC motifs with high affinity for single-stranded DNA or RNA usually consist of a short helix and two short β-strands linked by a zinc knuckle, which function during transcriptional activation, DNA recognition, RNA packaging, regulation of apoptosis [4, 5]. The first CCHC-ZFP was identified in the murine leukemia virus and Rous avian sarcoma virus, and subsequently in many group antigen proteins of retroviral nucleocapsids and eukaryotic retrotransposons [3]. Genome-wide investigations of *CCHC-ZF* gene family were carried out in humans (34), *Arabidopsis* (69), and yeast (7), while comprehensive studies on the CCHC-ZFPs in wheat have not been reported so far [6].

The *CCHC-ZF* genes were extensively founded in the plant genomes, which played central roles in seed development, and plant growth mediated by phytohormones, such as indole-3-acetic acid (IAA),

gibberellins (GA), methyl jasmonate (MeJA), and abscisic acid (ABA). In rice, *OsZFP* regulates lateral root development through IAA signaling pathways [7]. *AtCSP2* negatively regulates seed germination by adjusting GA and ABA levels [8]. In addition, a great number of *CCHC-ZF* genes are also involved in regulating plant tolerance to abiotic stress. For instance, overexpressing *OsZFP6*, a NaCl, H_2O_2, and $NaHCO_3$ responsive gene in rice, is able to increase the tolerance to H_2O_2 and $NaHCO_3$ in *Arabidopsis* [9]. Similarly, *BrCSDP3* positively regulates seed germination and seedling growth under dehydration and salt treatment [10]. The transcription of *OsRZ1, OsRZ2, OsRZ3* is up-regulated by low temperature treatment, but they show no response to high salt and drought stress [11]. In Pak-choi, *BcCSP1* plays important roles in responses to ABA and cold treatments [12]. Besides, some *CCHC-ZF* genes also involve in the regulation of plant defense to biotic stress [13]. Ectopic expression of *TaRZ1* in *Arabidopsis* confers the transgenic plants increased resistance against bacterial, indicating the role of *TaRZ1* in plant immune response [14]. The up-regulation of *AdRSZ21* in response to MeJA and pathogen infection indicates that *CCHC-ZF* genes might play a role in plant defense [15].

Bread wheat (*Triticum aestivum* L., A, B, and D sub-genome) was obtained by natural hybridization between *Triticum dicoccoides* (A and B sub-genome) and *Aegilops tauschii* (D sub-genome), which was a valuable material for evolutionary research due to the specificity of heterohexaploid [16, 17]. As one of the most important food crops in the world, the growth and development of wheat are susceptible to complex and variable environments, resulting in the reduction of yield [18]. Considering *CCHC-ZF* genes play important roles in plant growth and development, as well as responses to biotic and abiotic stresses, a comprehensive investigation of *TaCCHC-ZF* gene family will contribute to wheat stress resistance breeding and gene function research. In this study, we identified the *CCHC-ZF* genes from wheat genome by bioinformatic methods and analyzed the chromosomal location, subcellular localization, phylogenetic relationships, gene structures, proteins interaction network, and expression patterns of them. The promoter cis-elements and the miRNA potentially targeting *TaCCHC-ZF* genes were predicted to explore the transcriptional regulatory network of them. These works will lay the foundation for further analyses and application of *CCHC-ZF* genes in wheat and other plant species.

Results

Identification and characterization of the *CCHC* gene family

In this study, a total of 50 putative genes in wheat were retrieved based on the Hidden Markov Model (HMM) search. After Simple Modular Architecture Research Tool (SMART) searched, 50 wheat proteins sharing the CCHC conserved motifs were obtained, which were consistent to the predictions. Meanwhile, several important dicotyledonous and monocotyledonous plants were selected for reference analyses. We identified 38 *CCHC-ZF* genes in *T. dicoccoides*, 46 *CCHC-ZF* genes in *Ae. tauschii*, 17 *CCHC-ZF* genes in *Hordeum vulgare*, 26 *CCHC-ZF* genes in *Oryza sativa*, 33 *CCHC-ZF* genes in *Zea mays*, 22 *CCHC-ZF* genes in *Arabidopsis thaliana*, 95 *CCHC-ZF* genes in *Glycine max*, and 67 *CCHC-ZF* genes in *Solanum tuberosum* in the same method (Additional file 1: Table S1).

Subsequently, physicochemical properties of TaCCHC-ZFPs were analyzed, including the length of proteins, molecular weight (MW), instability index, aliphatic index (AI), isoelectric point (pI), grand average of hydropathicity (GRAVY), and the subcellular localization (Additional file 1: Table S2). Among the 50 TaCCHC-ZFPs, TaCCHC14 was identified to be the smallest protein with 162 residues of amino acids (aa), while TaCCHC31 with 1149 residues of amino acids is the largest one. The pI ranges from 5.31 (TaCCHC40) to 11.63 (TaCCHC41), and AI fluctuates from 20.06 (TaCCHC25) to 74.88 (TaCCHC5), and instability index varies from 25.67 (TaCCHC22) to 117.51 (TaCCHC41). Besides, the GRAVY values of all TaCCHC-ZFPs are negative, indicating they are hydrophilic proteins. The subcellular localization analyses showed that 28 TaCCHC-ZFPs were predicted to be located both in the cell nucleus and chloroplast, while 20 and 2 TaCCHC-ZFPs were only located in the nucleus or chloroplast, respectively. Additionally, all TaCCHC-ZFPs are composed of four secondary structure, including alpha helix (0.62%-40.61%), extended strand (5.99%-26.53%), beta turn (3.05%-24.62%), and random coil (43.32%-83.6%), of which random coil accounts for the main part of protein secondary structure (Additional file 1: Table S3).

Subsequently, we extracted the amino acid sequences of the conserved motif CCHCs using the MEME tool. As shown in Fig. 1, the CCHC conserved motif from wheat has the consensus sequence $CX_2CX_4HX_4C$, which has high affinity to single-stranded nucleic acids. Except for the completely conserved cysteine (C) and histidine (H) residues in the positions 6, 9, 14, 19, the conserved substituted glycine (G) residue occurs in the positions 10, 13, and hydrophobic or aromatic residues are found in the positions 7, 15 (Fig. 1).

Phylogenetic tree and sequence structure analysis

To explore the evolutionary relationship of the *CCHC-ZF* gene family, a phylogenetic tree was constructed using the full-length protein sequences of CCHC-ZFPs from both wheat and rice. These *CCHC-ZF* genes are classified into nine groups, named as groups to , which are distributed unevenly in each group (Fig. 2). Except for the group , the others all possess *CCHC-ZF* genes from both wheat and rice. The groups and both contain the most members of 13, while the group is also the group with most members of 11 *CCHC-ZF* genes in wheat. In addition, the group possesses the fewest members, two from wheat and one from rice. Based on the phylogenetic analysis, *TaCCHC-ZF* genes are classified into eight groups (groups to) for further analyses (Fig. 3a).

A schematic diagram displaying the motifs of TaCCHC-ZFPs was constructed using the MEME tool [19]. Through the annotations of the Pfam (PF00098) database, we found that the motif 1 was the CCHC domain, and the motif 3 and 4 both were RRM domains, and the motif 5, 6, and 7 were CSD domain, REPA OB domain, and Rep Fac-A C domain, respectively (Additional file 1: Table S4) [20]. As shown in Fig. 3b, the motif 1 was widely distributed in TaCCHC-ZFPs. Moreover, TaCCHC-ZFPs in one group generally tend to share a similar motif composition. For instance, the motif 5 only exists in the group , while the motifs 6,7, and 10 are specific to the group . Similarly, the motifs 8 is unique to the group and , and the motif 3 only occurs in the groups , , and . As a result, the motif patterns of TaCCHC-ZFPs in a

group are similar, suggesting that the protein architecture is conserved within a specific group. The functions of these conserved motifs remain to be elucidated, which may be relevant to specific biological functions.

In addition, the exon-intron structures of *TaCCHC-ZF* genes were further analyzed to further explore the evolution of the *TaCCHC-ZF* family. The gene structures of *TaCCHC-ZFs* in different groups are changeable in the number of exons (ranging from 2 to 15), and no genes with one exon was found (Fig. 3c). However, the *TaCCHC-ZF* genes in the same group usually share similar numbers of exons as expected, suggesting that they are evolutionarily conserved. For instance, all members of the group Ⅱ contain two or three exons, while seven *TaCCHC-ZF* genes of the group Ⅰ possess four exons. In contrast, some of the more closely related members were also observed to share similar length of exons. In general, the diverse gene structures of *TaCCHC-ZF* genes may be related to the involvement of *TaCCHC-ZF* gene family in many plant biological processes.

Chromosomal location and collinearity analysis of the *TaCCHC* gene family

MapGene2Chrom V2 was used to create the chromosome map of the *TaCCHC-ZF* genes based on the physical location information from the GFF3 file of wheat (Fig. 4a) [21]. The *TaCCHC-ZF* genes are unevenly distributed on wheat chromosomes, with the number of the genes on each chromosome varying from one (2A, 2B, 2D, 3A, 3B, 3D, 4A, 4B, 4D, 7D) to eight (1A) (Fig. 4b). Interestingly, we also found that the numbers of the *TaCCHC-ZF* genes on each chromosome were irrelevant to chromosome size. For instance, the smallest chromosome (6D, 473.6 Mb) encodes three *TaCCHC-ZF* genes, while the largest chromosome (3B, 830.8 Mb) contains only one *TaCCHC-ZF* genes. The *TaCCHC-ZF* genes spread roughly equally across the three sub-genomes of wheat (sub-genome A, 18; sub-genome B, 16; and sub-genome D, 16), which may cause redundant functions with genes on sub-genome A, indicating some *TaCCHC-ZF* genes may experience gene loss event during the evolution with low purifying selection. Meanwhile, the different chromosome position of several homologous gene pairs (e.g., *TaCCHC28*, *TaCCHC34*, and *TaCCHC36*) suggested that the chromosome rearrangement occurred during the evolution of wheat *TaCCHC-ZF* gene family [22].

Next, the synteny analyses were performed to evaluate the gene duplication events in *T. aestivum.* Interestingly, we didn't identify any tandem duplication events in these *TaCCHC-ZF* genes. However, a total of 36 segmental duplication events with 43 *TaCCHC-ZF* genes were identified, indicating that segmental duplication events were the main driving force for the evolution of *CCHC-ZF* gene family in wheat (Fig. 5). All duplicated genes in a pair belong to the same *TaCCHC-ZF* genes group. Furthermore, the Ka/Ks values of the *TaCCHC-ZF* gene pairs were calculated to explore the evolutionary constraints. The Ka/Ks values of the 36 gene pairs in wheat are generally less than 1, implying that the replicated *TaCCHC-ZF* genes could experience strong purification selection pressure (Additional file 1: Table S5). The Ks values were used to assess the divergence time (T) based on the formula $T = Ks/2\lambda \times 10^{-6}$ Mya (λ

= 6.5 × 10^{-9}). The divergence time of these genes diverged between 0.994 and 19.055 Mya (average 6.735, 34 values in 36 earlier than 2.26), mostly before the early Gramineae whole-genome duplication event.

Synteny analyses of CCHC members from wheat and eight other plant species

To further explore the evolutionary mechanisms and homologous genes of *TaCCHC-ZFs*, comparative syntenic maps were constructed by comparing eight representative species with wheat, including five monocots (*T. dicoccoides*, *Ae. tauschii*, *H. vulgare*, *O. sativa*, and *Z. mays*) and three dicots (*A. thaliana*, *G. max*, and *S. tuberosum*) (Fig. 6a). A total of 46 *TaCCHC-ZF* genes show collinearity relationships with 29 *CCHC-ZF* genes in *T. dicoccoides*, 15 in *Ae. tauschii*, 12 in *O. sativa*, 8 in *H. vulgare*, and 8 in *Z. mays*, respectively, while no this relationship between wheat and the three dicots analyzed were found, suggesting the closer phylogenetic relationships with the monocots than the dicots (Fig. 6b). Therefore, 76, 31, 28, 17, 16 orthologous gene pairs among wheat and *T. dicoccoides*, *Ae. tauschii*, *O. sativa*, *H. vulgare*, and *Z. mays* were identified, respectively (Additional file 1: Table S6). Hexaploid wheat (A, B, and D sub-genome) was obtained by natural hybridization between *T. dicoccoides* (A and B sub-genome) and *Ae. tauschii* (D sub-genome). Compared to *T. dicoccoides* and *Ae. tauschii*, more wheat *CCHC-ZF* genes are derived from *T. dicoccoides* based on the number of orthologous *CCHC-ZF* gene pairs. Among the three sub-genomes of wheat, 36 gene pairs (14 between the A and B sub-genomes, 11 between the A and D sub-genomes, 11 between the B and D sub-genomes) were identified, which were less than that between wheat and its sub-genome donors (Fig. 5 and Fig. 6a). This might be related to either the gene lost or chromosomal recombination during the polyploidization and evolution. Additionally, three *TaCCHC-ZF* genes (*TaCCHC37*, *TaCCHC46*, and *TaCCHC48*) were observed in all of five syntenic maps, indicating that these genes are relatively conserved in the evolution. However, some wheat *TaCCHC-ZF* genes identified are collinear with genes from only one species. For instance, *TaCCHC35* was identified to have a collinearity relationship with *Os12t0564600-01*, while there was no collinearity relationship with the *TaCCHC-ZF* genes from the other four species, implying that *TaCCHC35* may have been lost in the rest four plants and remained in wheat and rice.

To further investigate the evolutionary constraints of the *TaCCHC-ZF* gene family, the Ka/Ks ratios of the *CCHC-ZF* gene pairs were calculated. The Ka/Ks ratios of nearly all orthologous *CCHC-ZF* gene pairs were less than 1, suggesting that the *TaCCHC-ZF* genes might undergone purifying selection during evolution to eliminate harmful mutations at the protein level (Additional files 1: Table S6). The divergence time of these duplicated orthologous *TaCCHC-ZF* gene pairs were approximately 4.872 Mya (*T. dicoccoides*), 5.143 Mya (*Ae. tauschii*), 11.074 Mya (*H. vulgare*), 44.765 Mya (*O. sativa*) and 60.761 Mya (*Z. mays*) respectively, which were close to the result of the species evolution tree (Fig. 6b).

Cis-acting elements and GO enrichment analyses of *TaCCHC-ZF* genes

Transcription factors bind the cis-acting elements of the gene promoter regions to regulate transcription. Thus, the 1.5-kb upstream promoter regions of all *TaCCHC-ZF* genes were submitted to the PlantCARE to further investigate the potential biological functions of *TaCCHC-ZF* genes. A total of 636 cis-acting elements related to environmental stress signal and phytohormone responsiveness were found in the promoter regions of *TaCCHC-ZF* genes (Fig. 7a and Additional file 1: Table S7). Among them, 152 MeJA-responsive elements (CGTCA-motif and TGAGG-motif) and 128 ABA (abscisic acid)-responsive elements (ABRE) were found, respectively, which were the two most cis-acting elements in *TaCCHC-ZF* genes. The result suggested that MeJA and ABA might play a key role in the transcriptional regulation of *TaCCHC-ZF* genes. Moreover, 40 auxin-responsive cis-acting elements (TGA-element, AuxRR-core), 27 gibberellin-responsive elements (P-box, TATC-box, and GARE-motif), 40 ethylene-responsive elements (ERE), 16 salicylic acid-responsive elements (SARE and TCA-element) were identified in 40, 27, 16, 14 *TaCCHC-ZF* genes, respectively. Meanwhile, four types of cis-elements related to biotic or abiotic stress responsiveness were identified, such as 45 drought responsive elements (MBS), 60 low-temperature responsive elements (LTR), 9 defense and stress responsive elements (TC-rich repeats), 8 wound responsive elements (WUN-motif). Additionally, except for *TaCCHC31*, *TaCCHC44*, *TaCCHC49*, and *TaCCHC50*, the anaerobic induction (ARE) or anoxic specific inducibility element (GC-motif) were found in the rest 46 *TaCCHC-ZF* genes. In brief, the various types of cis-elements identified in the promoter regions indicate that *TaCCHC-ZF* genes may participate in transcriptional regulation of phytohormone signaling and biotic/abiotic stress responses.

Additionally, the GO enrichment of all *TaCCHC-ZF* genes was constructed to futher explore the gene functions. The GO terms consist of three categories: molecular function (MF), biological process (BP), and cellular component (CC). The enrichment results of the molecular function category revealed that all 50 *TaCCHC-ZF* genes were annotated under nine GO terms, including zinc ion binding, transition metal ion binding, metal ion binding, cation binding, nucleic acid binding, ion binding, heterocyclic compound binding, organic cyclic compound binding, and binding, all of which belonged to the molecular function category, suggesting that they might act as zinc finger transcription factors to regulate gene expression through DNA or RNA binding (Fig. 7b). The enrichment results of the biological process category showed that 12 genes were involved in four kinds of metabolic processes, such as nucleic acid metabolic process and cellular aromatic compound metabolic process. Moreover, seven *TaCCHC-ZF* genes shared five GO terms, which were DNA recombination, DNA replication, DNA repair, cellular response to DNA damage stimulus, and cellular response to stress, implying the potential roles of them during stress responses.

Protein interaction network and miRNA targets analysis

Proteins that perform similar functions or participate in the same pathway are more likely to exhibit interaction networks, forming gene modules or clusters in proteins interaction networks. To further

understand the interaction relationships and biological functions among TaCCHC-ZFPs, the STRING database was used to construct the protein-protein networks within TaCCHC-ZFP family. As shown in Fig. 8, 24 TaCCHC-ZFPs were found to be involved in the protein interaction networks with 202 branches, suggesting that they might perform similar function.

MicroRNAs (miRNAs) are small noncoding RNAs that function in RNA silencing and post-transcriptional regulation of gene expression. Thus, the potential miRNA targets of *TaCCHC-ZF* genes were predicted to provide support information about the regulatory mechanism of the *TaCCHC-ZF* genes. The results revealed that a total of 91 miRNA target sites were identified in 34 *TaCCHC-ZF* genes, with each gene corresponding to one to seven miRNAs (Additional file 1: Table S8). Among the 34 wheat *CCHC-ZF* genes, *TaCCHC36* and *TaCCHC50* had the most targets with seven miRNA target sites, followed by *TaCCHC21* and *TaCCHC26* with six miRNA target sites, indicating that the expression of *TaCCHC-ZF* genes may be regulated by multiple miRNAs. At the same time, tae-miR9652-5p has the most target sites in nine *TaCCHC-ZF* genes among the 47 wheat miRNAs, followed by tae-miR9782 targeting six genes.

Expression patterns of *TaCCHC* genes under different stresses

To further dissect the function under abiotic stresses, the expression patterns of all 50 *TaCCHC-ZF* genes under different treatments (drought, heat, drought & heat) were analyzed in this study (Fig. 9a). Among the 50 genes, 38 DEGs were screened out in these samples by using the edgeR package from five kinds of treatments (drought stress for 1 (6) h: DS-1 (6) h, heat stress for 1 (6) h: HS-1 (6) h, and combined drought and heat stress for 1 (6) h: DHS-1 (6) h), while no DEGs were identified under the DS-1h treatment (Additional file 1: Table S9). As shown in Fig. 9b, 32 *TaCCHC-ZF* genes responded to at least two treatments, while six *TaCCHC-ZF* genes only responded to one treatment. For instance, *TaCCHC34, TaCCHC42,* and *TaCCHC47* showed decreased expression under the HS-1h and DHS-1h treatments, while *TaCCHC11* exhibited increased expression under the HS-1h and DHS-1h treatments, implying that these genes were sensitive to heat and drought. As expected, some genes with close evolutionary relationships showed similar expression patterns. The expression of *TaCCHC22*, *TaCCHC23*, and *TaCCHC24* were down-regulated under the HS-1h and DHS-1h treatments, and the expression of *TaCCHC32* and *TaCCHC33* were up-regulated under the HS-6h and DHS-6h treatments. Additionally, the expression patterns of *TaCCHC-ZF* genes under cold and phosphorous starvation treatments were also analyzed (Additional file 2: Figure S1). Only eight *TaCCHC-ZF* genes responded to cold treatment, while no DEGs were identified under the phosphorous starvation treatment (Additional file 1: Table S9).

To verify the transcription profiles of the *TaCCHC-ZF* genes derived from the transcriptome data, 15 *TaCCHC-ZF* genes from the eight groups were selected to analyze their expression level under different treatments by quantitative real-time PCR (qRT-PCR). As shown in Fig. 9c, the expression patterns of most *TaCCHC-ZF* genes are congruent with the previously published data according to the results of qRT-PCR. Overall, the expression of *TaCCHC-ZF* genes could be influenced by multiple treatments.

Discussion

As one of the most important food crops in the world, wheat is subject to biotic and abiotic stresses, resulting in the reduction of yield. Zinc finger protein transcription factors play a vital role in plant growth and development, and biotic and abiotic stress responses [23]. Some previous studies showed that *TaCCHC-ZF* genes regulated plant growth and stress responses, such as *AtRZ-1a*, *Mt-Zn-CCHC*, and *NTT* [24-27]. Thus, the comprehensive bioinformatic analyses of *TaCCHC-ZF* gene family were performed to better understand the gene functions of *TaCCHC-ZF* genes due to the limited work on *TaCCHC-ZF* gene family.

In this study, we identified 50 TaCCHC-ZFPs from wheat and extracted CCHC motif sequences of these members, in which the conserved sites were consistent with the previous study (Fig. 1) [28]. Studies revealed that CCHC motif was a kind of nucleic acid binding domain, which contributed to RNA binding, DNA regulation, or protein-protein interactions [29-31]. Meanwhile, the results of GO enrichment showed that all 50 *TaCCHC-ZF* genes were annotated under nucleic acid binding term and zinc ion binding term (Fig 7b), implying that TaCCHC-ZFPs might function by binding DNA or RNA. Previous study showed that WCSP1 (TaCCHC7 in this study) was capable of binding dsDNA, ssDNA, and RNA homopolymers, whereas its ability to bind dsDNA was almost eliminated in the absence of C-terminal CCHC motif [32]. In *Arabidopsis*, CSDP1, homologous to TaCCHC14, which possesses seven tandem repeated CCHC motifs in the C-terminal half, acts as an RNA chaperone in the response to cold stress, helping to export mRNA from the nucleus to the cytoplasm [33].

The analyses of phylogenetic relationships, gene structures, and protein motifs showed that the homologous *TaCCHC-ZF* genes in sub-genomes A, B, D shared similar gene structures and conserved motifs, indicating the functions of *TaCCHC-ZF* genes were conservative during the evolution (Fig. 3a-c). The motif 1 (CCHC motif) is conserved in all TaCCHC-ZFPs. It is noteworthy that some motifs are distributed in specific groups, such as RRM, CSD, REPA OB, and Rep Fac-A C, which may play various roles in biological processes based on the different functions of *TaCCHC-ZF* genes (Additional file 1: Table S4) [34-36]. For instance, AtGRP2 containing one CSD and two CCHC zinc fingers motifs may be involved in cold-response and flower development [37]. Besides, RRM exists in groups □, □, and □, which can bind single-strand RNA and participate in the regulation of flowering and adaptation to heat stress [38]. AtSF1, a protein containing RRM, takes part in regulating heat stress response by affecting the alternative splicing of the pre-mRNA of the heat shock transcription factor HsfA2 [39].

Previous researches revealed that gene families usually experienced tandem duplication events or segmental duplication events to expand gene family members in the process of evolution [40]. Subsequently, syntenic analyses were carried out in this study (Fig. 5 and Fig. 6a). Wheat has undergone two major polyploid evolutionary events, accompanied by tandem duplication, segmental duplication, and transposition events [41]. However, the number of *TaCCHC-ZF* genes in a specific sub-genome was severely reduced during the transition from tetraploid to hexaploidy through the identification of *CCHC-ZF* genes in wheat and its sub-genomes donors, *T. dicoccoides* and *Ae. tauschii* (for A sub-genome, from 20

to 18 genes; B sub-genome, from 18 to 16 genes; D sub-genome, from 46 to 16 genes), proving that gene loss during hexaploidy wheat formation occurred extensively [42]. Generally, the Ka/Ks ratios for all the homologous gene pairs are less than 1, implying that *TaCCHC-ZF* genes may have undergone purifying selection pressure and the functions of these gene pairs do not diverge much after the two polyploidization events (Additional file 1: Table S5 and Table S6).

Cis-acting elements and miRNAs are involved in the regulation of gene expression at the transcriptional and post-transcriptional levels, respectively [43, 44]. Therefore, we predicted the cis-acting elements in the promoter regions of wheat *CCHC-ZF* genes and miRNAs targeting *TaCCHC* genes. Plenty of studies showed that cis-elements were essential factors of modulating gene expression under biotic and abiotic stress. For instance, *PbrMYB21* could interact with the MYB-recognizing cis-element in the promoter region of *PbrADC* to modulate polyamine synthesis by regulating ADC expression, improving drought tolerance [45]. In this study, a lot of cis-acting elements associated with environmental stress and phytohormone responsiveness were identified, indicating that *TaCCHC-ZF* genes might take part in multiple signaling pathways (Fig. 7a and Additional file 1: Table S7) [46]. In addition, plant miRNAs are associated with cell biology processes and response to stress, which can regulate gene expression at the post-transcriptional level through splicing mRNA or inhibiting translation. In this study, we found 47 wheat miRNAs target with 34 *TaCCHC-ZF* genes, including tae-miR9652-5p, tae-miR9782, tae-miR156, tae-miR159a/b, tae-miR164, and tae-miR167, etc. (Additional file 1: Table S8). Previous studies reported that some plant miRNAs, such as miR156, miR159a/b, miR164, miR319, and miR399, played a key role in the regulation of plant developmental time, the differentiation of tissues, and response to environmental stresses [47]. The miR156-overexpression alfalfa showed significant improvement in drought tolerance with reduced water loss and higher survival compared with the wild-type control [48]. Moreover, ABA induced the accumulation of miR159 to mediate the cleavage of *MYB33* and *MYB101* transcripts in geminating *Arabidopsis* seeds [49]. In brief, cis-acting elements and miRNAs may be regulators of *TaCCHC-ZF* gene expression.

Previous studies showed that *CCHC-ZF* genes responded to multiple stresses. For example, the cold resistance of atRZ-1a-overexpressing transgenic *Arabidopsis* plants was enhanced compared to wild-type plants, with earlier germination and better seedling growth under cold treatment as well [25]. Drought and heat are the major environmental stresses affecting wheat growth and development, often resulting in the decline of wheat yield. In this study, we investigated the potential functions of *TaCCHC-ZF* genes under drought and heat treatments, and 38 DEGs were screened out (Fig. 9a-c and Additional file 1: Table S9). Previous research revealed that AtCSP3-overexpressing transgenic *Arabidopsis* plants showed higher survival rates under the drought and salt treatment, whereas the *atcsp3* mutant displayed lower survival rate [50]. *TaCCHC14*, homologous to *Arabidopsis At4g36020.1* (*AtCSP3*), was down-regulated under the DHS-6h treatment, suggesting that they might have similar functions under the drought and heat stresses. *TaRZ2* (*TaCCHC49* in this study) can negatively regulate seed germination and seedling growth under the salt or dehydration treatments but contribute to enhancing cold tolerance of transgenic *Arabidopsis* [51]. Meanwhile, *TaCCHC14* and *TaCCHC49* were found to share similar cis-elements, such

as MBS, LTR, and so on. Overall, these results indicated that *TaCCHC-ZF* genes might be involved in the plant responses to drought and heat stresses.

Conclusions

CCHC-ZFPs are involved in multiple physiological processes, such as seed development, plant growth, and responses to biotic and abiotic stresses. In this study, a total of 50 *TaCCHC-ZF* genes were identified from wheat by bioinformatics tools. Subsequently, these *TaCCHC-ZF* genes were classified into eight groups with specific motifs and gene structures. Interestingly, only segmental duplication events were identified in *TaCCHC-ZF* genes, suggesting that the segmental duplication events are the main driving force for *TaCCHC-ZF* genes evolution. In addition, collinearity relationships between wheat and eight other representative organisms were analyzed and no gene pairs were found between wheat and the three dicots. Plenty of cis-acting elements related to environmental stress were found in the promoter regions of *TaCCHC-ZF* genes. GO enrichment results showed that all *TaCCHC-ZF* genes were annotated under metal ion binding and nucleic acid binding. The analyses of miRNA targets suggested that the *TaCCHC-ZF* genes could be regulated by the miRNAs. Furthermore, the expression patterns of *TaCCHC-ZF* genes and qRT-PCR verification showed that some *TaCCHC-ZF* genes were involved in the responses to drought and heat stresses.

Methods

Plant materials and abiotic stress treatments

Bread wheat cultivar Fielder was used throughout this study. Seeds were transferred into Petri dishes with wet filter paper and cultured at 4℃ for 5 days. Then, the germinated seedlings were grown at 22℃ in a greenhouse (16 h light and 8 h dark period, about 60-70 μmol/(m^2·s), and 50% relative humidity), and cultured with half strength Murashige and Skoog liquid medium. Two-week seedlings of Fielder were treated by drought stress (1/2 MS liquid medium with 20% (m/V) PEG-6000), heat stress (1/2 MS liquid medium and 40℃), or combined drought & heat stress (1/2 MS liquid medium with 20% PEG-6000 and 40℃) for 1 h or 6 h, respectively, while the seedlings under normal growth conditions (22℃, watered) were used as control. All the experiments were carried out in parallel, and three biological replicates were performed for each time point. Leaves were collected at 1h and 6h after treatments and frozen in the liquid nitrogen immediately and stored at -80℃ for further analysis.

Data retrieval and identification of *CCHC* genes

The reference genome and protein sequences of all the species in this study were downloaded from the Ensemble Plants database (http://plant.ensembl.org/index.html). To identify the *CCHC-ZF* family members, the HMM profile of CCHC conserved motif (PF00098) was retrieved from Pfam database (http://pfam.xfam.org) and used to search against all of the protein sequences through the HMMER

search tool with an E-value cut-off < $1e^{-4}$ [20, 52]. After removing the redundant sequences, candidate genes were submitted to SMART (http://smart.emblheidelberg.de/) to further confirm CCHC-ZFP members [53]. The theoretical isoelectric point (pI), molecular weight (MW), instability index, and grand average of hydropathicity (GRAVY) were calculated using the ExPasy site (http://web.expasy.org/protparam/) [54]. The subcellular localization of each CCHC-ZFP protein was predicted using the Cell-PLoc 2.0 (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/) [55]. The secondary structure of TaCCHC-ZFPs was predicted using SOPMA secondary structure prediction (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html) [56].

Sequence analysis and structural characterization of the CCHC proteins in wheat

Multiple protein sequence alignment of the characterized CCHC-ZFPs was performed via ClustalX2 [57]. Then, depending on the full-length protein sequence alignment, the phylogenetic tree was constructed using MEGA 7.0 with the neighbor-joining (NJ) method based on Poisson model, 1000 bootstrap replications and pairwise deletion [58]. To analyze the conserved motifs of CCHC-ZFPs in wheat, the MEME online program (https://meme-suite.org/meme/tools/meme) was used with the parameters as follow: a maximum number of 10 motifs and optimal motif width of 6 to 50 amino acid residues [19]. Then, the conserved motif of wheat CCHC was extracted and visualized by WebLogo (http://weblogo.threeplusone.com/) [59]. The genome annotation file (GFF3 file) of wheat was obtained from the Ensemble Plants database (http://plant.ensembl.org/index.html) for analyzing the exon-intron structures of *TaCCHC-ZF* genes. Finally, the prepared files were imported into TBtools for visualizing the protein conserved motifs and gene structures [60].

Chromosome distribution, collinearity analysis, and Ka/Ks analysis

According to the chromosome location information obtained from the Ensemble Plants database, the *TaCCHC-ZF* genes were mapped to the wheat chromosome using MapGene2Chrom V2 (http://mg2c.iask.in/mg2c_v2.0/) [21]. Subsequently, MCScanX and DIAMOND were used to analyze the gene duplication events and synteny of *CCHC-ZF* genes of wheat with the default parameters, and the figure was displayed by Circos [61-63]. In addition, the collinearity relationships and segmental duplication events of *TaCCHC-ZF* gene pairs from other species were also performed similarly. Species evolution tree was constructed by using TimeTree online tool (http://www.timetree.org) [64]. Then, TBtools was adopted to calculated Ka (non-synonymous) and Ks (synonymous) of the duplicated gene pairs for further estimating duplication events [60]. The time (T) of duplication in millions of years (Mya) was estimated with the formula $T = Ks / 2\lambda \times 10^{-6}$ Mya ($\lambda = 6.5 \times 10^{-9}$).

Cis-acting element analysis and gene ontology annotation of *TaCCHC* family genes

In order to investigate the putative cis-regulatory element in the promoter regions of *TaCCHC-ZF* genes, the 1.5-kb upstream genomic DNA sequences from the transcription start codon were submitted to the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html) [65]. Then, the Gene Structure Display Server (GSDS, http://gsds.gao-lab.org) was adopted to visualized the cis- element distribution [66]. The gene ontology (GO) analysis of wheat *CCHC* genes was predicted for functional annotation using Omicshare Tools (https://www.omicshare.com/).

Prediction of protein interaction network and miRNA targets

The wheat CCHC-ZFPs were submitted to the STRING database (https://string-db.org/) to assemble protein-protein interaction networks with high confidence (0.700) [67]. Then, The Cytoscape was used to visualize the interaction network with default parameters [68]. To predict the miRNAs targeting *TaCCHC-ZF* genes, mature miRNA sequences and *TaCCHC-ZF* gene sequences of wheat were submitted to the psRNATarget tool (https://www.zhaolab.org/psRNATarget/), filtered at an expectation level ≤ 5.0 [69].

Expression analysis of *CCHC* genes in *Triticum aestivum*

The expression profiles of *TaCCHC-ZF* genes under different abiotic stresses were available from Wheat Expression Browser powered by the expVIP (http://www.wheat-expression.com/) [70, 71]. Subsequently, the differentially expressed genes (DEGs) under different abiotic stresses (drought stress, heat stress and combined drought and heat stress) were identified by using the edgeR package (fold change ≥ 2 and q-value ≤ 0.5) [72]. TBtools was used to generate the gene expression heatmap [60]. Finally, EVenn (http://ehbio.com/test/venn/#/) was adapted to construct Venn diagrams.

RNA extraction and qRT-PCR analyses

The total RNA from wheat leaves was extracted using TRIzol reagent (Vazyme Biotech Co., Ltd), following the manufacturer's instructions. For qRT-PCR analyses, RNA concentration was assessed by the NanoDrop 2000 spectrophotometer (ND-2000, Thermo Fisher Scientific, Inc.). Total RNAs were reverse transcribed with the HiScript II 1st Strand cDNA Synthesis Kit (+gDNA wiper) (Vazyme Biotech Co., Ltd). 15 different expression genes of the *TaCCHC-ZF* family in response to stress were detected by qRT-PCR analyses, while *TaRP15* was used as the internal reference gene. The reaction system consisted of 5 μL of 2 × ChamQ Universal SYBR qPCR Master Mix (Vazyme Biotech Co., Ltd), 2 μL of template, 0.2 μL of each prime, and 2.6 μL of ddH_2O. The reaction was carried out as follows: pre-denaturation at 95℃ for

30 s (step 1), denaturation at 95℃ for 10 s (step 2), primer annealing/extension and collection of fluorescence signal at 60℃ for 30 s (step 3). The next 40 cycles started at step 2. Each sample was performed in three biological replicates and three technical replicates. Subsequently, the data from qRT-PCR analyses was analyzed with the $2^{-\Delta\Delta CT}$ method. Primer sequences used in this study were listed in detail in Additional file 1: Table S10.

Abbreviations

CCHC-ZFP: CCHC zinc finger protein; IAA: indole-3-acetic acid; GA: gibberellins; ABA: abscisic acid; MeJA: methyl jasmonate; MW: molecular weight; AI: aliphatic index; pI: isoelectric point; GRAVY: grand average of hydropathicity; aa: amino acids; REPA OB: Replication protein A OB; Rep Fac-A C: Replication factor-A C terminal domain; Ka: Non-synonymous; Ks: Synonymous; Mya: Mya millions of years; SA: salicylic acid; MF: molecular function; BP: biological process; CC: cellular component; miRNA: MicroRNA; DS-1 (6) h: drought stress for 1 (6) h; HS-1 (6) h: heat stress for 1 (6) h; DHS-1 (6) h: combined drought and heat stress for 1 (6) h; DEG: differentially expressed gene; qRT-PCR: quantitative real-time PCR; HMM: Hidden Markov Model; NJ: neighbor-joining; SMART: Simple Modular Architecture Research Tool; GSDS: Gene Structure Display Server; GO: gene ontology; UTR: untranslated regions

Declarations

Ethics approval and consent to participate

The wheat seedlings used in this study were grown in the greenhouse in Hunan University, Changsha, China. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and analyzed during the current study are available in the Ensembl Plants (http://plant.ensembl.org/index.html), MaizeGDB (https://maizegdb.org/genome/assembly/Zm-B73-REFERENCE-GRAMENE-4.0), Pfam (http://pfam.xfam.org), SMART (http://smart.emblheidelberg.de/), STRING (https://string-db.org/), and expVIP (http://www.wheat-expression.com/) repository.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by grants from National Natural Science Foundation of China (31872866), the Science and Technology Innovation Program of Hunan Province (2019NK2081), and National Key Research and Development Program of China (2017YFF0210301).

Authors' contributions

X.G. and W.X. designed the experiments. A.S. and X.G. wrote the main manuscript text. Y.L. and A.S. conducted the experiments. A.S., X.Z., F.C., R.C. and H.X. collected and analyzed phenotype data. A.S., Y.L., W.X. prepared Figures 1-9. All authors read and approved the manuscript.

Acknowledgements

The authors are grateful Dr. Shuai Hu, School of life sciences, Tsinghua University for critical review of this manuscript.

Author information

[1]College of Biology, Hunan University, Changsha 410082, P.R. China

References

1. Takatsuji H. Zinc-finger proteins: the classical zinc finger emerges in contemporary plant science. Plant Mol Biol. 1999;39(6):1073-78.
2. Kiełbowicz-Matuk A. Involvement of plant C2H2-type zinc finger transcription factors in stress responses. Plant Sci. 2012;185-6:78-85.
3. Summers MF. Zinc finger motif for single-stranded nucleic acids? Investigations by nuclear magnetic resonance. J Cell Biochem. 1991;45(1):41-8.
4. Laity JH, Lee BM, Wright PE. Zinc finger proteins: new insights into structural and functional diversity. Curr Opin Struct Biol. 2001;11(1):39-46.
5. Wang Y, Yu Y, Pang Y, Yu H, Zhang W, Zhao X, et al. The distinct roles of zinc finger CCHC-type (ZCCHC) superfamily proteins in the regulation of RNA metabolism. RNA Biol. 2021;1-20.
6. Aceituno-Valenzuela U, Micol-Ponce R, Ponce MR. Genome-wide analysis of CCHC-type zinc finger (ZCCHC) proteins in yeast, *Arabidopsis*, and humans. Cell Mol Life Sci. 2020;**77**(20):3991-4014.
7. Cui P, Liu H, Ruan S,Ali B, Gill RA, Ma H, et al. A zinc finger protein, interacted with cyclophilin, affects root development via IAA pathway in rice. J Integr Plant Biol. 2017;59(7):496-505.

8. Sasaki K, Kim M, Kanno Y, Seo M, Kamiya Y, Imai R. *Arabidopsis* cold shock domain protein 2 influences ABA accumulation in seed and negatively regulates germination. Biochem Biophys Res Commun. 2015;456(1):380-4.
9. Guan Q, Wang L, Bu Q, Wang Z. The rice gene OsZFP6 functions in multiple stress tolerance responses in yeast and *Arabidopsis*. Plant Physiol Biochem. 2014;82:1-8.
10. Choi MJ, Park YR, Park SJ, Kang H. Stress-responsive expression patterns and functional characterization of cold shock domain proteins in cabbage (*Brassica rapa*) under abiotic stress conditions. Plant Physiol Biochem. 2015;96:132-40.
11. Kim JY, Kim WY, Kwak KJ,Oh SH, Han YS, Kang H, et al. Zinc finger-containing glycine-rich RNA-binding protein in *Oryza sativa* has an RNA chaperone activity under cold stress conditions. Plant Cell Environ. 2010;33(5):759-68.
12. Huang F, Tang J, Hou X. Molecular cloning and characterization of *BcCSP1*, a Pak-choi (*Brassica rapa ssp. chinensis*) cold shock protein gene highly co-expressed under ABA and cold stimulation. Acta Physiol Plant. 2016;38(2):47.
13. Amorim L, Santos R, Neto J,Guida-Santos M, Crovella S, Benko-Iseppon. Transcription factors involved in plant resistance to pathogens. Curr Protein Pept Sci. 2017;8(4):335-51.
14. Xu T, Lee HJ, Sy ND,Kang H. Wheat (*Triticum aestivum*) zinc finger-containing glycine-rich RNA-binding protein TaRZ1 affects plant growth and defense response in *Arabidopsis thaliana*. Plant Growth Regul. 2015;76(3):243-50.
15. Kumar KRR, Kirti PB. Novel role for a serine/arginine-rich splicing factor, AdRSZ21 in plant defense and HR-like cell death. Plant Mol Biol. 2012;80(4-5):461-76.
16. Petersen G, Seberg O, Yde M,Berthelsen K. Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). Mol Phylogenet Evol. 2006;39(1):70-82.
17. Ozkan H, Levy AA, Feldman M. Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. Plant Cell. 2001;13(8):1735-47.
18. Fujita M, Fujita Y, Noutoshi Y,Takahashi F, Narusaka Y, Yamaguchi-Shinozaki Ket al. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Curr Opin Plant Biol. 2006;9(4):436-42.
19. Bailey TL, Boden M, Buske FA,Frith M, Grant CE, Clementi Let al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37:W202-8.
20. Mistry J, Chuguransky S, Williams L,Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: The protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412-9.
21. Chao JT, Kong YZ, Wang Q,Sun Y, Gong D, Lv J, et al. MapGene2Chrom, a tool to draw gene physical map based on Perl and SVG languages. Yi Chuan. 2015;37(1):91-7.
22. Chen Y, Song W, Xie X,Wang Z, Guan P, Peng H, et al. A collinearity-incorporating homology inference strategy for connecting emerging assemblies in the triticeae tribe as a pilot practice in the plant pangenomic era. Mol Plant. 2020;13(12):1694-1708.

23. Li WT, He M, Wang J, Wang Y. Zinc finger protein (ZFP) in plants-A review. Plant Omics. 2013;6(6):474-80.
24. Lee K, Kang H. Emerging roles of RNA-binding proteins in plant growth, development, and stress responses. Mol Cells. 2016;39(3):179-85.
25. Kim Y, Kim JS, Kang H. Cold-inducible zinc finger-containing glycine-rich RNA-binding protein contributes to the enhancement of freezing tolerance in *Arabidopsis thaliana*. Plant J. 2005;42(6):890-900.
26. Radkova M, Revalska M, Kertikova D,Iantcheva A. Zinc finger CCHC-type protein related with seed size in model legume species *Medicago truncatula*. Biotechnol. Biotechnol. Equip. 2019;33(1):278-85.
27. Crawford BCW, Ditta G, Yanofsky MF. The *NTT* gene is required for transmitting-tract development in carpels of *Arabidopsis thaliana*. Curr Biol. 2007;17(13):1101-8.
28. Armas P, Calcaterra NB. Retroviral zinc knuckles in eukaryotic cellular proteins. Nova Science Publishers. 2013;1(2):51-80.
29. Espinosa J. Trypanosoma cruzi poly-zinc finger protein: a novel DNA/RNA-binding CCHC-zinc finger protein. Mol Biochem Parasitol. 2003;131(1):35-44.
30. Matthews JM, Kowalski K, Liew CK,Sharpe BK, Fox AH, Crossley M, et al. A class of zinc fingers involved in protein-protein interactions. Eur. J. Biochem. 2000;267(4):1030-8.
31. Ganie SA. RNA chaperones: potential candidates for engineering salt tolerance in rice. Crop Sci. 2020;60(2):530-40.
32. Karlson D, Nakaminami K, Toyomasu T,Imai R. A cold-regulated nucleic acid-binding protein of winter wheat shares a domain with bacterial cold shock proteins. J Biol Chem.2002;277(38):35248-56.
33. Park SJ, Kwak KJ, Oh TR,Kim YO, Kang H. Cold shock domain proteins affect seed germination and growth of *Arabidopsis thaliana* under abiotic stress conditions. Plant Cell Physiol. 2009;50(4):869-78.
34. Budkina KS, Zlobin NE, Kononova SV,Ovchinnikov LP, Babakov AV. Cold shock domain proteins: structure and interaction with nucleic acids. Biochemistry (Mosc). 2020;85(Suppl 1):S1-19.
35. Daubner GM, Clery A, Allain FH. RRM-RNA recognition: NMR or crystallography...and new findings.Curr Opin Struct Biol. 2013;23(1):100-08.
36. Bochkarev A, Pfuetzner RA, Edwards AM,Frappier L. Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. Nature. 1997;385(6612):176-81.
37. Fusaro AF, Bocca SN, Ramos RL, Barrôco RM, Magioli C, Jorge VC,, et al. *AtGRP2*, a cold-induced nucleo-cytoplasmic RNA-binding protein, has a role in flower and seed development. Planta. 2007;225(6):1339-51.
38. Pi B, He X, Ruan Y, Jang J, Huang Y. Genome-wide analysis and stress-responsive expression of CCCH zinc finger family genes in *Brassica rapa*. BMC Plant Biol. 2018;18(1).

39. Lee KC, Jang YH, Kim S,Park H, Thu MP, Lee JH,et al. RRM domain of *Arabidopsis* splicing factor SF1 is important for pre-mRNA splicing of a specific set of genes. Plant Cell Rep. 2017;36(7):1083-95.
40. Cannon SB, Mitra A, Baumgarten A,Young ND, May G. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biol. 2004;4:10.
41. Peng JH, Sun D, Nevo E. Domestication evolution, genetics and genomics in wheat. Mol Breeding. 2011;28(3):281-301.
42. Berkman PJ, Visendi P, Lee HC,Stiller J, Manoli S, Lorenc MT, et al. Dispersion and domestication shaped the genome of bread wheat. Plant Biotechnol J. 2013;11(5):564-71.
43. Hernandez-Garcia CM, Finer JJ. Identification and validation of promoters and cis-acting regulatory elements. Plant Sci. 2014;217-8:109-19.
44. Hausser J, Zavolan M. Identification and consequences of miRNA-target interactions-beyond repression of gene expression. Nat Rev Genet. 2014;15(9):599-612.
45. Li K, Xing C, Yao Z,Huang X. PbrMYB21, a novel MYB protein of*Pyrus betulaefolia*, functions in drought tolerance and modulates polyamine levels by regulating arginine decarboxylase gene. Plant Biotechnol J. 2017;15(9):1186-203.
46. Liu J, Peng T, Dai W. Critical cis-acting elements and interacting transcription factors: key players associated with abiotic stress responses in plants. Plant Mol Biol Rep. 2014;32(2):303-17.
47. Willmann MR, Poethig RS. Conservation and evolution of miRNA regulatory programs in plant development. Curr Opin Plant Biol. 2007;10(5):503-11.
48. Arshad M, Feyissa BA, Amyot L,Aung B, Hannoufa A. MicroRNA156 improves drought stress tolerance in alfalfa (*Medicago sativa*) by silencing *SPL13*. Plant Sci. 2017;258:122-36.
49. Reyes JL, Chua N. ABA induction of miR159 controls transcript levels of two MYB factors during *Arabidopsis* seed germination. Plant J. 2007;49(4):592-606.
50. Kim M, Sato S, Sasaki K,Saburi W, Matsui H, Imai R. COLD SHOCK DOMAIN PROTEIN 3 is involved in salt and drought stress tolerance in *Arabidopsis*. FEBS Open Bio. 2013;3(1):438-42.
51. Xu T, Gu L, Choi MJ,Kim RJ, Suh MC, Kang H, et al. Comparative functional analysis of wheat (*Triticum aestivum*) zinc finger-containing glycine-rich RNA-binding proteins in response to abiotic stresses. PLoS One. 2014;9(5):e96877.
52. Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. Bioinformatics. 2013;29(19):2487-9.
53. Letunic I, Khedkar S, Bork P. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res. 2021;49(D1):D458-60.
54. Gasteiger E. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31(13):3784-8.
55. Chou K, Shen H. Cell-PLoc 2.0: an improved package of web-servers for predicting subcellular localization of proteins in various organisms. Natural Science. 2010;02(10):1090-103.

56. Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci. 1995;11(6):681-4.
57. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947-48.
58. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870-74.
59. Crooks GE, Hon G, Chandonia JM,Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14(6):1188-90.
60. Chen C, Chen H, Zhang Y,Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13(8):1194-202.
61. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.
62. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015; 12(1):59-60.
63. Krzywinski M, Schein J, Birol I,Connors J, Gascoyne R, Horsman D,et al. Circos: An information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639-45.
64. Kumar S, Stecher G, Suleski M,Hedges SB. TimeTree:a resource for timelines, timetrees, and divergence times. Mol Biol Evol. 2017;34(7):1812-9.
65. Lescot M, Déhais P, Thijs G,Marchal K, Moreau Y, Van de Peer Y, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 2002;30(1):325-7.
66. Hu B, Jin J, Guo A,Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31(8):1296-7.
67. Szklarczyk D, Gable AL, Lyon D,Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607-13.
68. Shannon P, Markiel A, Ozier O,Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.
69. Dai X, Zhuang Z, Zhao PX. psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic Acids Res. 2018;46(W1):W49-54.
70. Ramírez-González RH, Borrill P, Lang D,Harrington SA, Brinton J, Venturini L,et al. The transcriptional landscape of polyploid wheat. Science. 2018;361(6403):r6089.
71. Liu Z, Xin M, Qin J,Peng H, Ni Z, Yao Y,et al. Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.). BMC Plant Biol. 2015;15(1):152.

72. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26(1):139-40.

Figures

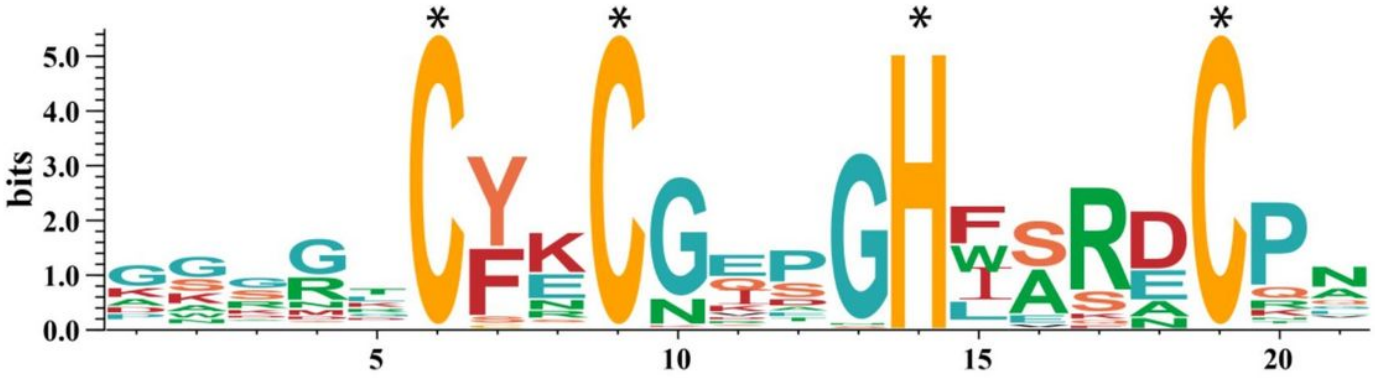

Figure 1

Sequence of the CCHC motifs in wheat. The height of the letter at each location (in bits) represents the conservation of the sequences, and the height of every single letter in the letter means the relative frequency of the corresponding amino acid of that position.

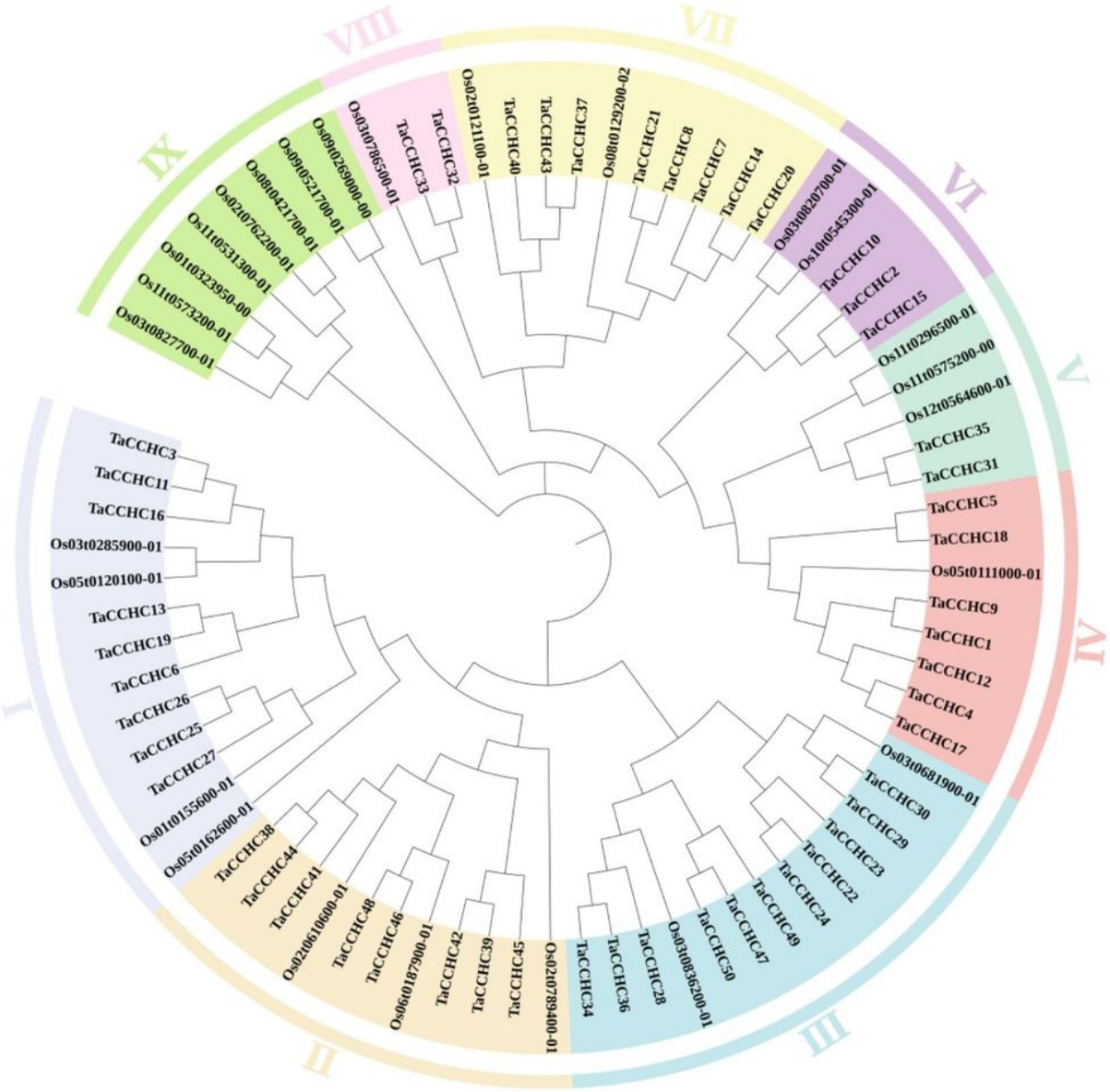

Figure 2

Phylogenetic tree of CCHC-ZFPs in wheat and rice. Protein sequences were aligned using the ClustaX2, and the Neighbor-joining (NJ) phylogenetic tree was constructed using the MEGA 7.0 based on Poisson model. Different groups are indicated by different colors, respectively.

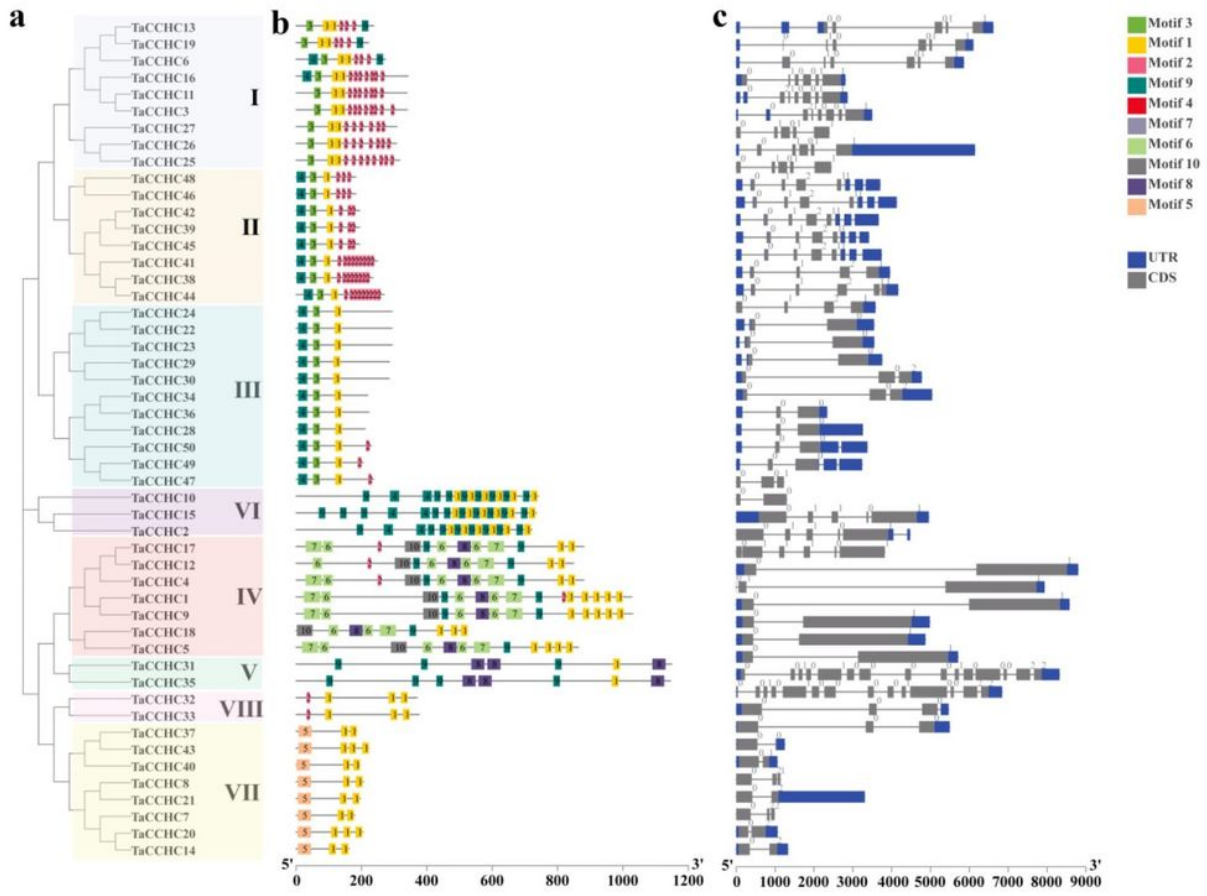

Figure 3

Comparative analyses of the phylogenetic relationships, protein conserved motifs, and gene structures of *CCHC-ZF* family in wheat. **a** Phylogenetic tree of 50 TaCCHC-ZFPs was constructed by using MEGA 7.0. Each group was marked by a different color. **b** Motif composition of wheat CCHC-ZFPs. MEME online tool was used to identify the conserved motifs of the TaCCHC-ZFPs. Each motif is represented by different colored boxes with the corresponding number in the center of the motifs. **c** Gene structures of *TaCCHC-ZF* genes. The black lines represent the introns, while the blue and grey boxes represent the untranslated regions (UTRs) and exons, respectively. The numbers represent the phases of corresponding introns.

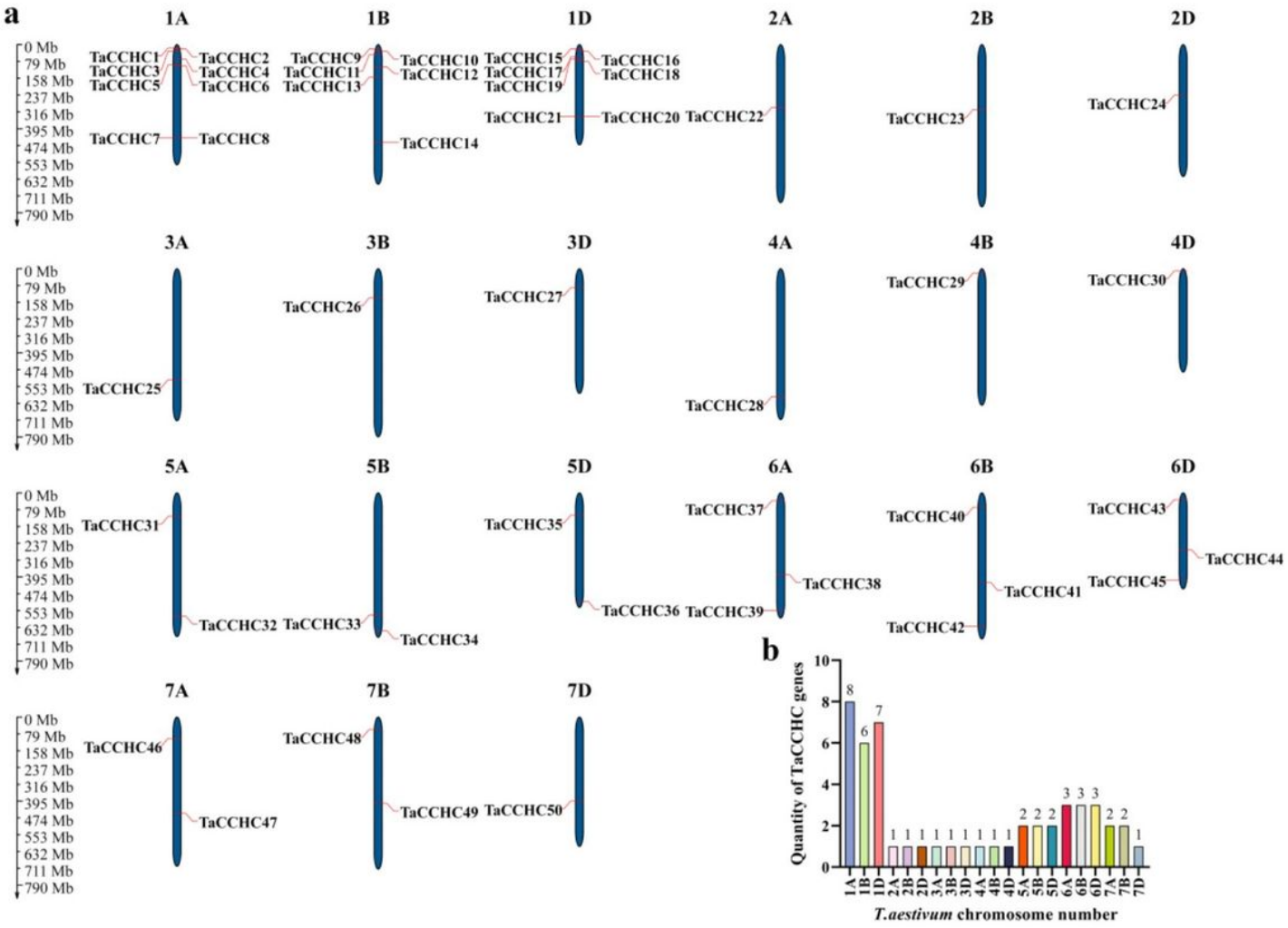

Figure 4

Chromosome distribution of the wheat *CCHC-ZF* genes. **a** Chromosomal localization of the *TaCCHC-ZF* genes. The dark blue columns indicate wheat chromosomes with the scale in megabases (Mb). The chromosome numbers are displayed at the top of each chromosome. **b** Numbers of *TaCCHC-ZF* genes on each *T. aestivum* chromosome.

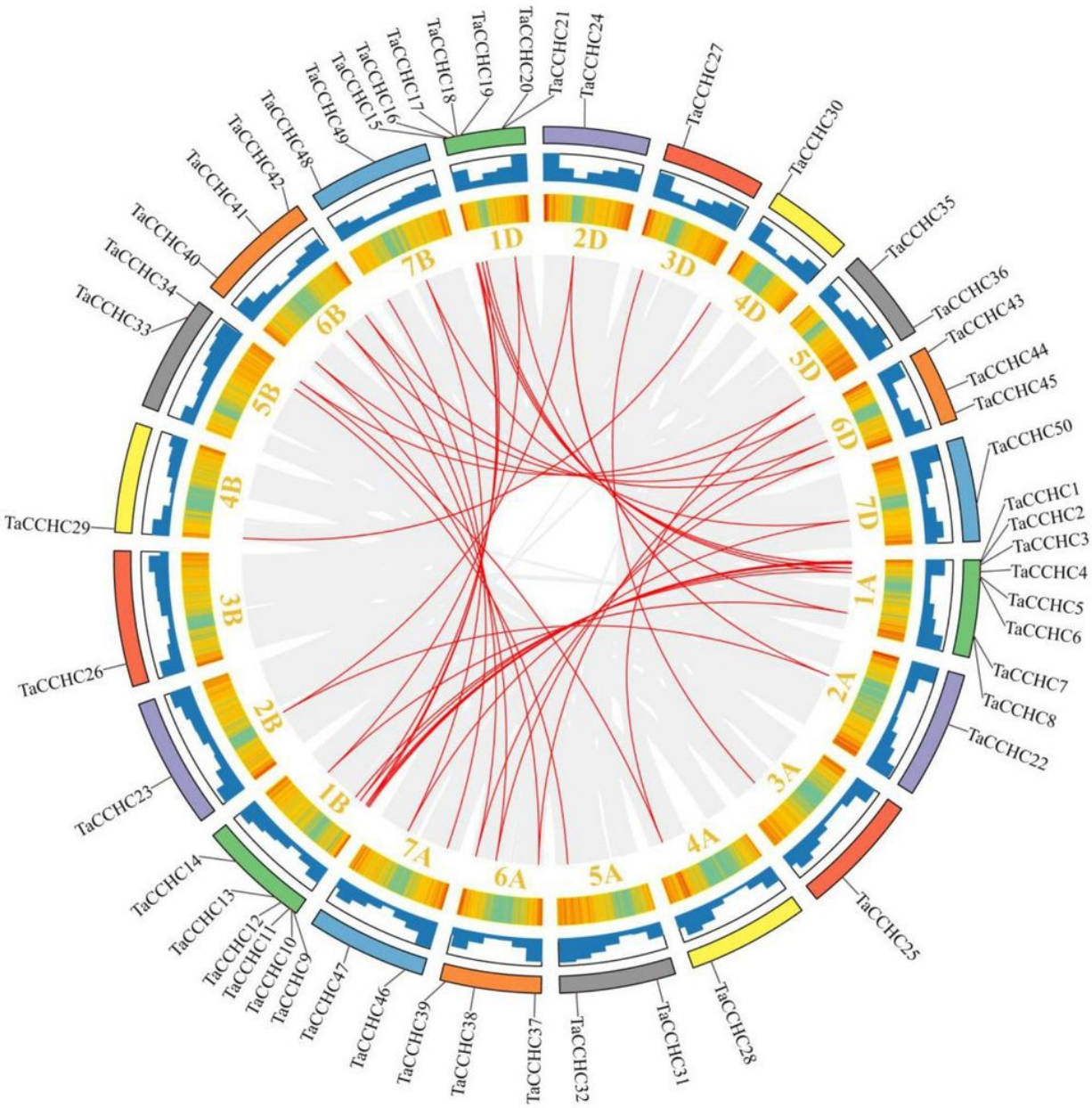

Figure 5

Collinearity analysis of the *CCHC-ZF* family in wheat. The gray lines represent all synteny blocks in the wheat genome, while the red lines represent duplicated *CCHC-ZF* gene pairs. Wheat chromosomes are displayed by rectangles with different colors, and the heatmaps and histograms along the rectangles indicate the gene density of each chromosome.

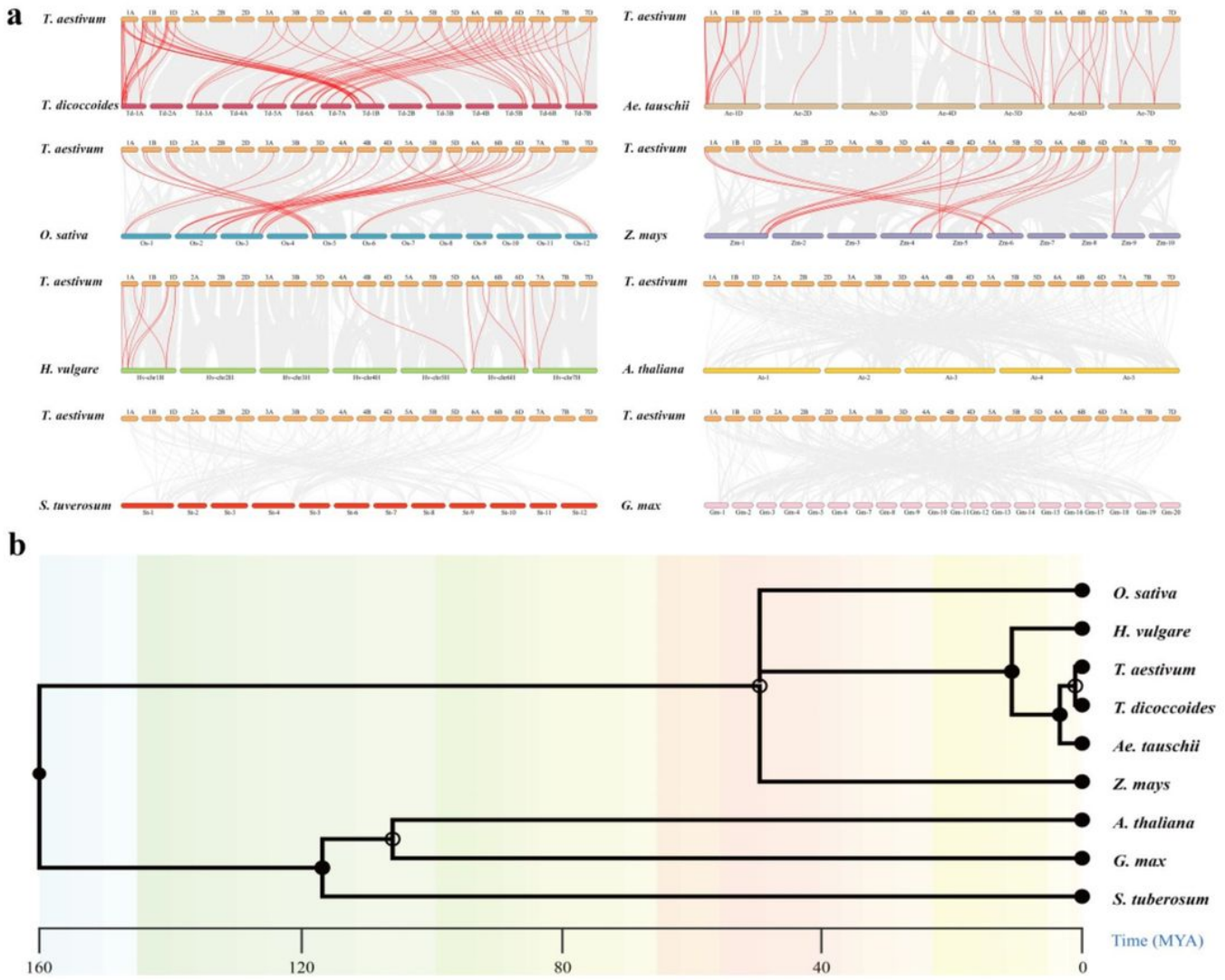

Figure 6

Synteny analysis of the *CCHC-ZF* family between wheat and other species. **a** Collinearity analysis of the *TaCCHC-ZF* family with other eight representative species. The grey lines in the background represent the collinear blocks in the genome of wheat and other species, while the red lines indicate the syntenic *CCHC-ZF* gene pairs. **b** Species evolution tree of wheat and other eight species.

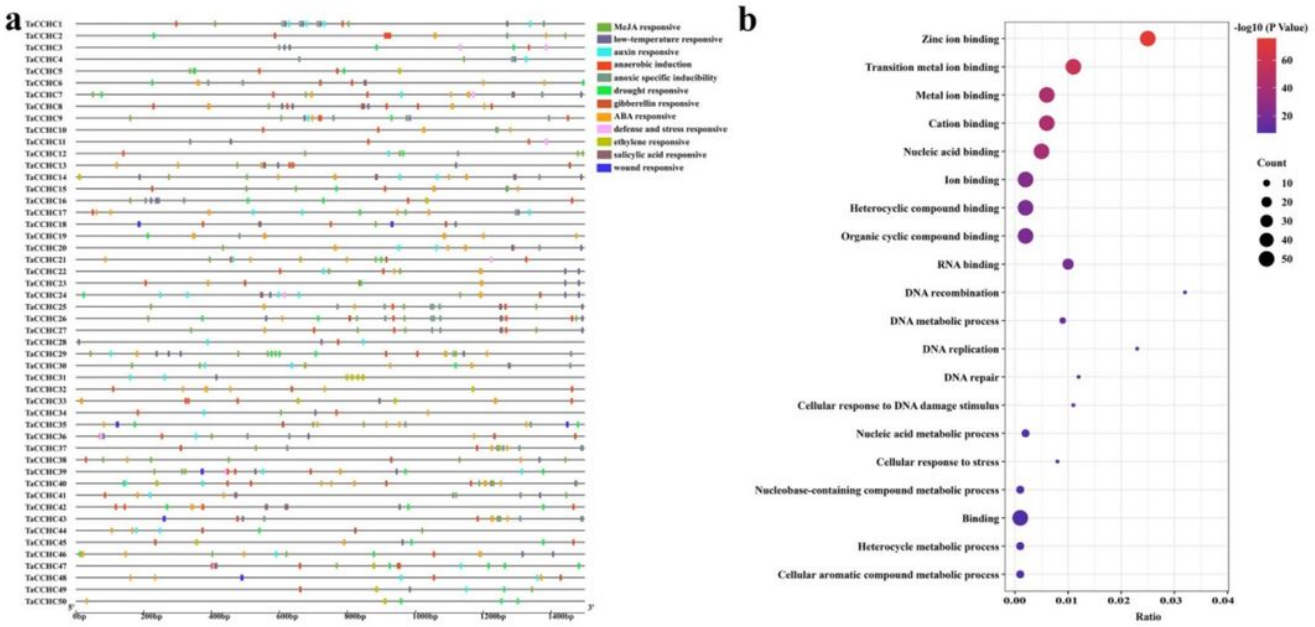

Figure 7

Analyses of cis-acting elements and functional annotation of *TaCCHC-ZF* genes. **a** Distribution of predicted cis-acting elements in the promoter regions of *TaCCHC-ZF* genes. The color blockers indicate different cis-acting elements and their locations in these *TaCCHC-ZF* genes. **b** GO enrichment analysis of the *TaCCHC-ZF* genes.

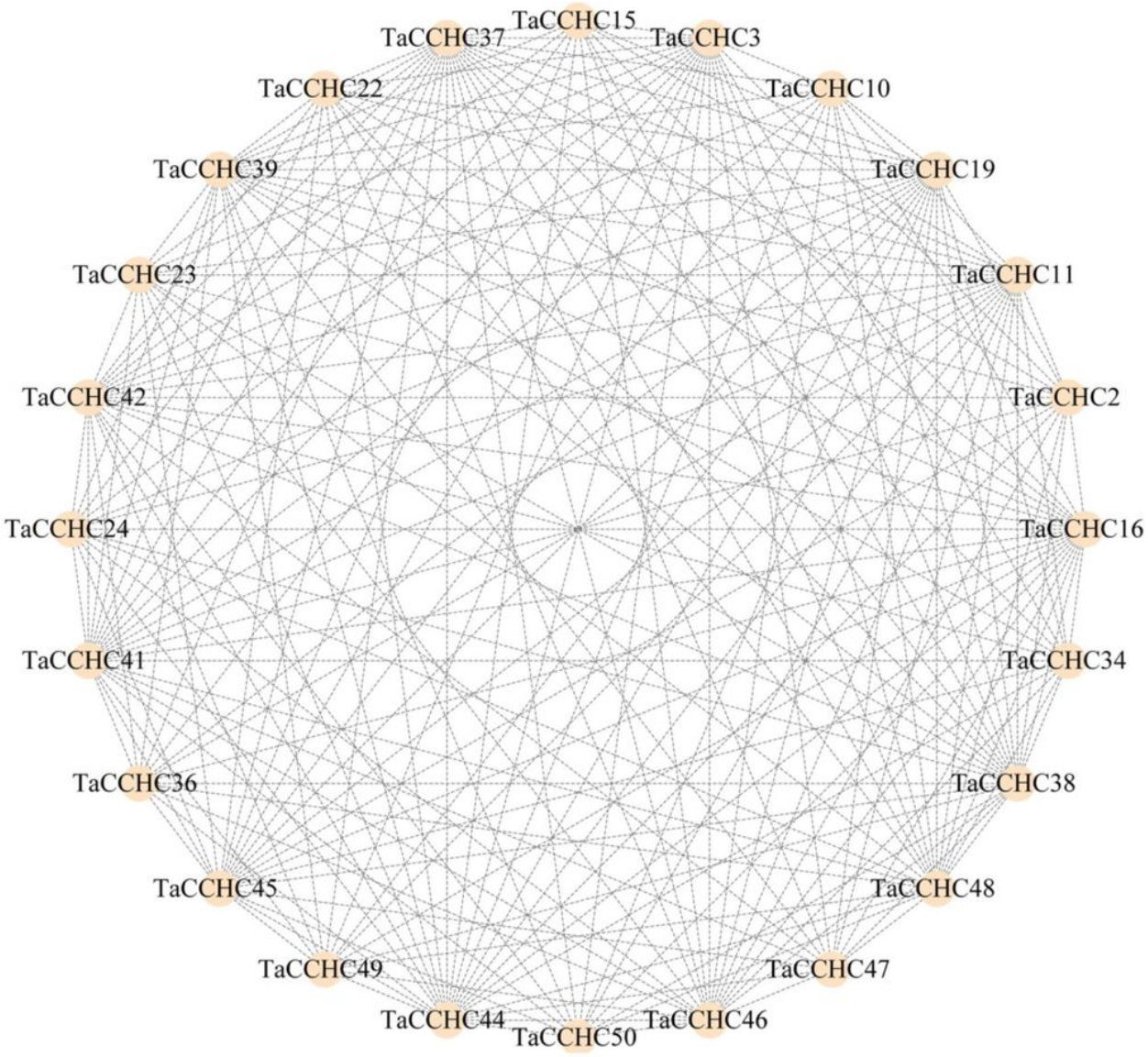

Figure 8

Interaction network of TaCCHC-ZFPs. A total of 202 interactions are displayed among 24 TaCCHC-ZFPs. The protein-protein interaction networks of wheat CCHC-ZFPs were predicted using the STRING tools with high confidence (0.700) [67], and was used to visualized by the the Cytoscape with default parameters [68].

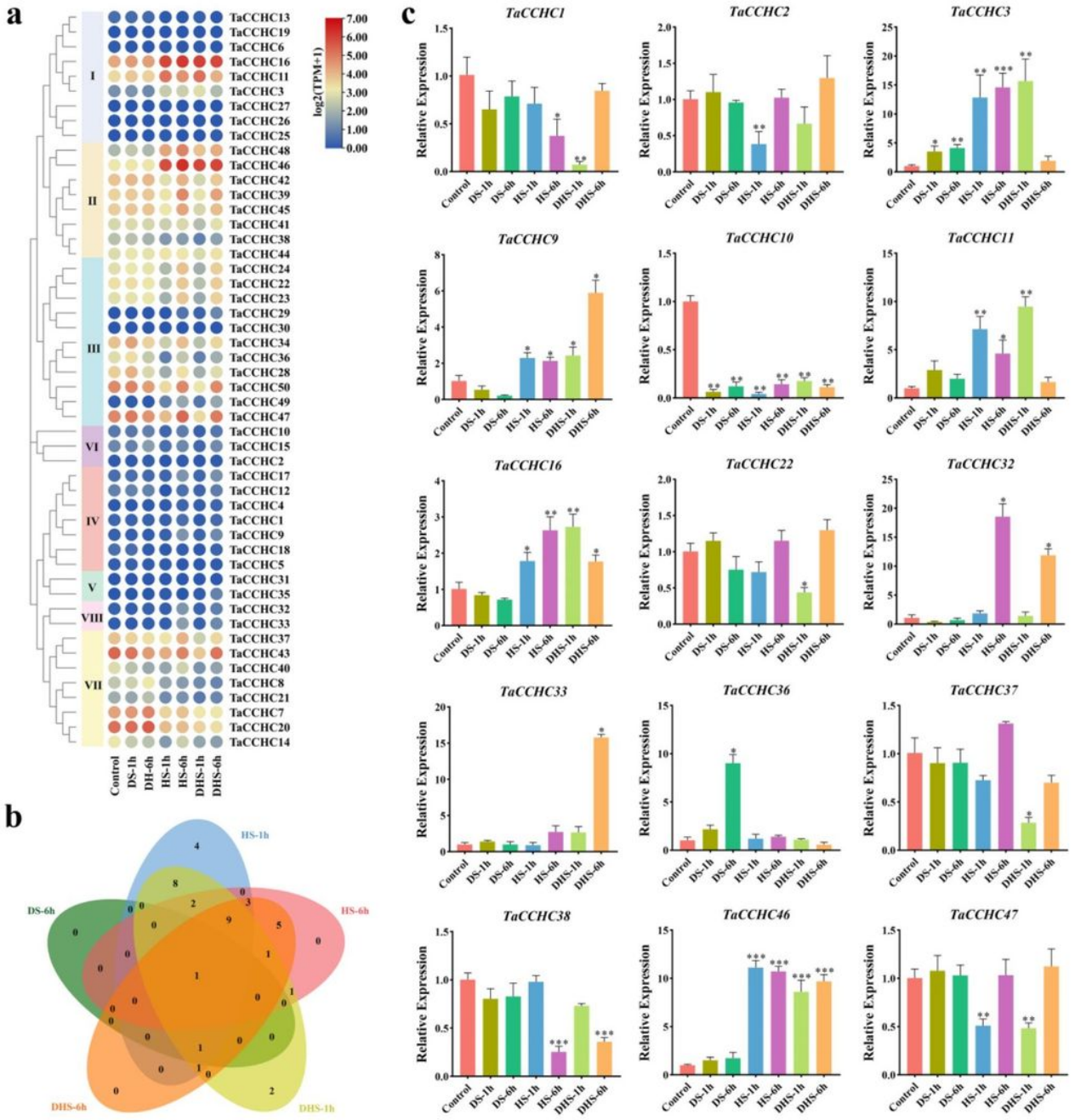

Figure 9

Expression patterns of wheat *CCHC-ZF* genes under different conditions. **a** Expression profiles of 50 *TaCCHC-ZF* genes under different stress treatments. HS-1 (6) h: heat stress for 1 (6) h; DS-1 (6) h: drought stress for 1 (6) h; DHS-1 (6) h: combined drought and heat stress for 1 (6) h. The color in the heat map reflect *TaCCHC-ZF* genes expression level. **b** Venn diagrams of DEGs under different treatments. **c** Expression analyses of 15 *TaCCHC-ZF* genes in response to different treatments by qRT-PCR. Data were normalized to *actin-TaRP15* and error bars represent standard deviation among three independent replicates (* P < 0.05, ** P < 0.01, *** P < 0.001, Student's *t*-test).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.doc
- Additionalfile2.doc