Comprehensive genomic survey, structural classification and expression analysis of C2H2-type zinc finger factor in wheat (*Triticum aestivum* L.)

Yongliang Li

Hunan University

Aolong Sun

Hunan University

Qun Wu

Hunan University

Xiaoxiao Zou

Hunan University

Fenglin Chen

Hunan University

Ruqiong Cai

Hunan University

Hai Xie

Hunan University

Meng Zhang

Hunan University

Xinhong Guo (✉ gxh@hnu.edu.cn)

Hunan University

Research Article

Keywords: *Triticum aestivum*, C2H2-ZFP, Molecular structural analysis, Subet- specific motifs, Expression patterns

Posted Date: February 11th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-215409/v1>

License: (cc) (i) This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Plant Biology on August 18th, 2021.
See the published version at <https://doi.org/10.1186/s12870-021-03016-3>.

1 **Comprehensive genomic survey, structural classification and**
2 **expression analysis of C₂H₂-type zinc finger factor in wheat**
3 **(*Triticum aestivum* L.)**

4 Yongliang Li^{1,2}, Aolong Sun^{1,2}, Qun Wu¹, Xiaoxiao Zou¹, Fenglin Chen¹, Ruqiong Cai¹, Hai Xie¹,
5 Meng Zhang^{1,*}, Xinhong Guo^{1,*}

6

7 ¹College of Biology, Hunan University, Changsha 410082, China

8 ²These authors contributed equally: Yongliang Li and Aolong Sun

9 *Correspondence and requests for materials should be addressed to Xinhong Guo

10 (gxh@hnu.edu.cn), Meng Zhang (zhangmeng2019@hnu.edu.cn).

Abstract

Background: The C₂H₂-type zinc finger proteins (C₂H₂-ZFPs) are one of major classes of transcription factors that play important roles during plant growth, development and stress responses. Limited information about the C₂H₂-ZF genes hinders the molecular breeding in bread wheat (*Triticum aestivum*).

Results: In this study, 457 C₂H₂-ZFP proteins (including 253 splice variants), which contain four types of conserved domain (named Q, M, Z, and D), could be further classified into ten subsets. They were identified to be distributed on 21 chromosomes in *T. aestivum*. Subset-specific motifs, like NPL-, SFP1-, DL- (EAR-like-motif), R-, PL-, L- and EK-, might make C₂H₂-ZFP diverse multifunction. Interestingly, NPL- and SFP1-box were firstly found to be located in C₂H₂-ZFP proteins. Synteny analyses showed that only 4 pairs of C₂H₂ family genes in *T. aestivum*, 65 genes in *B. distachyon*, 66 genes in *A. tauschii*, 68 genes in rice, 9 genes in *A. thaliana*, were syntenic relationships respectively. It indicated that *TaZFPs* were closely related to genes in *Poaceae*. From the published transcriptome data, totally 198 of 204 *TaC₂H₂-ZF* genes have expression data. Among them, 25 *TaC₂H₂-ZF* genes were certificated to be significantly differentially expressed in 5 different organs and 15 different development stages by quantitative RT-PCR. The 18 *TaC₂H₂-ZF* genes were verified in response to heat, drought, and heat & drought stresses. According to expression pattern analysis, several *TaZFPs*, like *Traes_5BL_D53A846BE.1*, were not only highly expressed in L2DAAs, RTLS, RMS, but also endowed tolerance to drought and heat stresses, making them good candidates for molecular breeding.

Conclusions: This study systematically characterized the *TaC₂H₂-ZFPs* and their potential roles in *T. aestivum*. Our findings provide new insights into the C₂H₂-ZF genes in *T. aestivum* as well as a foundation for further studies on the roles of *TaC₂H₂-ZF* genes in *T. aestivum* molecular breeding.

33 **Keywords:** *Triticum aestivum*, C₂H₂-ZFP, Molecular structural analysis, Subet-specific motifs,

34 Expression patterns

Background

The C₂H₂-type zinc finger proteins constitute a large family of transcription factors. The C₂H₂-ZF domain was defined as about 30 amino acids with two conserved Cys and two His residues which bound to one Zn²⁺ atom tetrahedrally and form a structure as C-X₂₋₄-C-X₃-F-X₅-L-X₂-H-X₃₋₅-H [1]. According to the defined C₂H₂-ZF types, C₂H₂-ZFPs were classified into two subsets X-tandem-SF and X-isolated-SF. The X-tandem-SF contains two clearly subsets (X-t1-SF and X-t2-SF), and the X-isolated-SF contains ten clearly subsets (X-1i-Q-SF, X-1i-M-SF, X-1i-Z-SF, X-1i-D-SF, X-2i-Q-SF, X-2i-M-S and X-2i-Z-SF, X-2i-Mix-SF, X-3i-SF, and X-4i-SF), where X represents any species [2, 3]. Previous studies have shown that C₂H₂-ZF genes represent only 0.8%, 2.3%, 3% of all genes in *Saccharomyces cerevisiae*, *Diptera*, *mammals*, respectively [4].

The C₂H₂-ZFP family, which have been reported in numerous plant genomes, significantly contribute to the plant growth development, and also participate in multiple biological processes by transcriptional regulation. For instance, genome-wide analyses of ZFP proteins in plants have been carried out in *Arabidopsis* (179), durum wheat (122), rice (189), maize (211), soybean (321), foxtail millet (124) and other species based on genome sequences [2]. In *T. aestivum*, the C1-2i Q-type and C₃H zinc finger protein (ZFP) transcription factor subclass has been reported to play important roles in plant stress responses and physiological stages [5, 6]. In *Arabidopsis*, the *AZF1*, *AZF2*, *AZF3*, *ZAT6*, *ZAT7*, *ZAT8*, *ZAT10*, *ZAT12* and *ZAT18* have been shown to function in multiple cellular processes, including seed germination, drought, cold, high-salinity, phosphate and oxidative stress responses [7, 8, 9, 10, 11, 12, 13, 14]. In rice, several members of the C₂H₂-type ZFPs, such as *ZFP36*, *ZFP39*, *ZFP182*, *ZFP245*, *ZFP252*, and *ZFP179*, have also been shown to be involved in the responses of drought, salinity, and oxidative stress [14, 15, 16, 17, 18, 19, 20]. In Poplar,

PtrZFP2/19/95 showed high expression levels in leaves and/or roots under environmental stresses by genome-wide analysis, which provided a solid foundation for studying the biological roles of C_2H_2 -ZF genes in *Populus* growth and development [20]. According to genome-wide identification in silkworm *Bombyx mori*, five genes including *CTCF*, *fru*, *wor*, *lola*, and *BIGBMGA004569*, had different expression patterns in ovaries, among which *CTCF* gene expression level increased rapidly in late pupae and adult stages. The results provide a new resource for studying growth and development of the silkworm [21]. To date, extensive studies showed that great majority of C_2H_2 were involved in the growth, development, hormone signaling biotic and/or abiotic stress responses of plants, animals, and fungi as a transcription factor [4, 7, 22, 23].

The conserved amino acid residues of these C_2H_2 -ZF genes were consistent with previous studies of the C_2H_2 domain in other plant species [10]. These C_2H_2 -ZF genes studies include main structural features in C_2H_2 -ZFPs, such as the arrangement of C_2H_2 -ZF domains (tandem or dispersed), the length of spacer between the ZFs, the number of C_2H_2 -ZF domains and the “QALGGH” sequence, whereas yeast and animal C_2H_2 -ZFPs do not have this motif [3]. Compared with other eukaryote C_2H_2 -ZFPs, in multiple-fingered C_2H_2 -ZFPs, the plant zinc-finger domains are separated by long spacers that vary in length and sequence (such as, the double zinc finger protein ZPT2-7 and ZPT2-11 in petunia, the distance between its adjacent zinc finger structures is 19 and 65 amino acids, respectively), whereas the C_2H_2 -ZFPs of yeast and animals are mostly clustered and separated by only six to eight amino acids [4, 24].

Moreover, the structures of C_2H_2 -ZFPs also contain several non-zinc-finger motifs, such as some known EAR-box and unknown R-box, PL-box, L-box, and EK-box in *T. aestivum* [25]. The EAR motif (DLNxxP or LxLxL) was showed to be a predominant transcriptional repressor motif in some plant

79 ZFP proteins [25]. Overexpression of most DREB/ERF proteins with an EAR motif led to reduced
80 expression levels of stress-related genes and decreased abiotic stress tolerance, and fusion of the EAR
81 motif to a number of transcriptional activators was found to convert them to dominant repressors [26].
82 Thus, the different classifications or different types of C₂H₂-ZF domains provides some extrinsic of
83 function information for the C₂H₂-ZF genes, might make C₂H₂-ZFP diverse multifunction in
84 comprehensive physiological stages.

85 Bread wheat (*Triticum aestivum*) as a staple crop of the world plays essential roles to sustain food
86 security [27, 28]. However, dissecting gene function in *T. aestivum* is not easy due to its hexaploid
87 genome and highly redundancy of genes. In this study, the comprehensive analyses were carried out
88 based on phylogenetic relationships, physicochemical properties, subcellular localization prediction,
89 chromosomal locations, and conserved protein domain in bread wheat C₂H₂-ZF genes family. The
90 results of this study will provide the foundation for further functional analyses of the C₂H₂-ZF genes
91 families, important theoretical significance and application value for our understanding of the genetics
92 and the evolution of *T. aestivum*.

Results

Classification and distribution analysis of *Triticum aestivum* C₂H₂-ZFPs

The 457 putative C₂H₂-type Zinc Finger proteins (C₂H₂-ZFP) (including 253 splice variants), were identified from the Plant transcription factor database (PlantTFDB) and the iTAK database. Further identification were conducted through the HMM profile HMMER 3.0 program and SMART database search. After removing incomplete and duplicate sequences by using the CD-HIT software, a total of 204 non-redundant candidate genes were obtained to encode 457 proteins, of which 316 proteins in full length and 141 cases as partial in N- and/or C-terminal regions but containing complete C₂H₂-ZF domain. In addition, according to the retrieved of the PlantTFDB database, iTAK database and ExPASy server, the 204 *T. aestivum* C₂H₂-ZFPs sequences showed that their amino acid length, molecular weight (Mw), isoelectric points (pIs), subcellular localization predictions, for details about TaC₂H₂-ZFP, please refer to Additional files 1: Table S2.

We identified 204 *TaC₂H₂-ZF* genes in *T. aestivum*, and revealed the distribution of the 204 *TaC₂H₂-ZF* genes on all the seven chromosomes in both A, B, and D sub-genomes (Additional files 1: Table S2; Fig 2a). A maximum of 54 (26.5%) *TaC₂H₂-ZF* genes were located in chromosome 5, 48 on long arms and 6 on short arms, and 5B sub-genomes (26.47%) encoded a maximum number of *TaC₂H₂-TaZF* genes in the among of all sub-genomes. 38 (18.6%) *TaC₂H₂-ZF* genes were located in chromosome 2, 14 on long arms and 24 on short arms. The chromosome 7 encoded only 13 genes, 7 out of 13 were distributed on long arms. 85 (41.67%) *TaC₂H₂-ZF* genes were encoded in the B sub-genome and 115 (56.37%) *TaC₂H₂-ZF* genes were distributed on long arms of the chromosomes, suggesting the members of each subsets were not distributed on each chromosome averagely.

TaC₂H₂-ZF genes were also more identified on chromosomes 3 and 4, with 27 and 33 genes, respectively (Fig 2b).

Phylogeny analysis of C₂H₂-ZF domains

Furthermore, according to the variation of the plant-specific conserved amino acid sequence “QALGGH” and distances between metal ligands, the 457 C₂H₂-ZF domains of the *T. aestivum* were also classified into four categories (named Q, M, Z and D) [2]. According to the classification results of the C₂H₂-ZF domain, 214 (46.8%) were classified as Ta-tandem-SF and 243 (53.2%) were classified as Ta-isolated-SF. 253 (55%) *TaC₂H₂-ZF* genes were different splice variants, while contain the same number C₂H₂-ZF domains for most *TaC₂H₂-ZF* genes, only the length of the amino acid sequence is different (Additional files 1: Table S1, S2). However, there are some *TaC₂H₂-ZF* genes splice variants were exceptions, which number and types of their domains have been reduced and changed. For instance, *Traes_6DL_F47CFC6D9.1/2/3/4/5/6* have five C₂H₂ (M3;M4;M4;M4;Z1) domains, four splice variants *Traes_6DL_F47CFC6D9.7/8/9/10* have three (M4;M4;Z1), the other three *Traes_6DL_F47CFC6D9.11/12/13* only have two (M4;Z1) domains. And the C₂H₂ domains of *Traes_3AS_BE15533CD.1/2* (Z1;D)-C₂H₂ domains are completely different from its variant *Traes_3AS_BE15533CD.3* (M3-C₂H₂ domain). Undoubtedly, different splice variants of a *TaC₂H₂-ZF* gene have different number or type C₂H₂ domains, may perform different functions [42]. For the detailed information of the classified C₂H₂-ZFPs in *T. aestivum*, see Additional files 1: Table S1.

To evaluate the evolutionary relationships of the C₂H₂ gene family in *T. aestivum*, an unrooted phylogenetic tree was constructed with the neighbor-joining method based on the domain sequence

alignment of 527 conserved ZFP domains (Fig 3a). Through the analysis of domain sequence alignments, our analysis showed that most members were well separated. The majority of C₂H₂-ZF proteins in the same clade, particularly the most closely related proteins, typically shared common types (for example, Q-type, M1-type), indicating a similar potential function between these C₂H₂-ZF proteins. Within each subfamily member in a branch, the strong amino acid sequence conservation is evident from the short branch lengths at the tips of the trees, suggestion of strong evolutionary relationships among subfamily members (Fig 3c). However, there are clear exceptions, and the different type subfamilies of C₂H₂-ZF proteins are found in the same clade (Fig 3b, c). Our results cast a new light on classification of C₂H₂-ZF domain.

Gene structure and conserved motif analysis of C₂H₂-ZFPs

To further explore the evolutionary relationship among *TaC₂H₂-ZF* genes, an unrooted neighbor-joining phylogenetic tree of 204 *TaC₂H₂-ZFPs* full-length amino acid sequence was constructed. The gene structures and motif characteristics of *TaC₂H₂-ZFPs* were analyzed (Fig 4a). The results of Gene Structure Display Server (GSDS) analysis showed that the number of introns in the *TaC₂H₂-ZF* genes contained from zero to ten introns, most of the genes had one to three introns (104 genes have no introns) (Fig. 4b), and most members contained typical C₂H₂ domains. The lengths of individual *TaC₂H₂-ZFP* were variable in intron length and it could partly reflect the length of different genes. For instance, the longest gene, *Traes_5DS_976CBF9EB.1*, with a size of 29.1kb, was due mainly to the fact that it contained a longest intron with length of 21.9kb.

To identify common motifs among different groups of *TaZFPs*, we employed the MEME Version 5.1.1 [43] software to analyze the conserved motifs in amino acid sequences of 204 *TaC₂H₂-ZFP* genes

in *T. aestivum*. Figure 4c shows the symbols of the motifs and the composition of the 204 *TaC₂H₂-ZFP* in the C₂H₂ family. Protein domain analysis showed that most members contained typical C₂H₂ domains. The C₂H₂ domains were annotated as the ZFP specific domains, in their N-terminal and C-terminal regions. Fifteen specific motifs were defined, with the ZFPs of subset Ta-t1-SF (*Traes_6DL_3B9E63F04.2*) containing the largest number of motifs. The *TaC₂H₂-ZFPs* belonging to the same subset have a similar motif composition (e.g. *Traes_7AL_0BC1C3E70.1* and *7DL_38929DD1E.1*, *IAS_21ACC7F67.1* and *6DL_F47CFC6D9.1*). Also, some motifs appeared in only certain specific subsets. For example, motif 5 was unique to subset Ta-t1-SF and Ta-1i-M-SF, while L-box was specific to subset Ta-2i-Q-SF and Ta-3i-SF. However, it may be due to the divergence of the C₂H₂ motif, especially the diverse motif numbers and spacing in the amino acid sequences between Cys and Cys or Cys and His in each protein. Overall, aboved analysis suggested that the C₂H₂-ZFPs with only one and two Q-type C₂H₂ domains were most conservative in *T. aestivum*.

Discovery of conserved motifs specific of the C₂H₂-type zinc finger domain: SFP1-, DL-, R-, PL-, L-, EK-, and NPL-box

Several other conserved motifs except C₂H₂ subfamily members were detected by MEME algorithm and multiple sequence alignments (Fig 5). The C₂H₂ subset Ta-1i-M-SF is composed of ZF proteins that contain not only the C₂H₂-ZF domain but also a Nucleoplasmin-like (NPL) domain [44, 45]. Ta-2i-Mix-SF contains *Schizosaccharomyces pombe* finger protein 1 (SFP1) domain [46, 47]. Ta-1i-Q-SF contains DL-box. Ta-2i-Q-SF contain R-box, PL-box, L-box, and EK-box, respectively [25] (Fig 5). NPL-box (Ta-1i-M-SF) and L-box (Ta-2i-Q-SF) are located in the N-terminal region of C₂H₂-ZF domains. DL-box and R-box (Ta-1i-Q-SF) are located in the C-terminal region of C₂H₂-ZF

domains. SFP1-box and R-box (Ta-2i-Q-SF) are located in the C-terminal region of the first C₂H₂-ZF domains. PL-box (Ta-2i-Q-SF) and EK-box are located in the C-terminal region of the second C₂H₂-ZF domains. Conservation motif of L-box, NPL-box and SFP1-box in some C₂H₂-type zinc fingered proteins have been showed in the previous studies [45, 47].

The DL-box is similar to the previously reported EAR motif (LDLSL) and may make the proteins contain transcriptional repressor functions [48]. Another promising finding was that different splice variants of many genes contained different motifs. For instance, two splice variant of Traes_5DL_50E35B4F3 contain R-box and PL-box, respectively. Traes_4DS_676CACEAC contain R-box, L-box, and EK-box in three splice variants. It indicated that the different subset members maybe contain a novel motif and execute different biological functions.

The C₂H₂ family genes in *T. aestivum* were more closely evolutionary relationship with *Poaceae*

To further infer the phylogenetic mechanisms of the *T. aestivum* C₂H₂-ZF gene family and to better understand the origin, we constructed five comparative syntenic maps of *T. aestivum* associated with four representative species, including three monocots (*A. tauschii*, *B. distachyon*, and rice) and one dicot (*A. thaliana*) (Fig 6). Only four pairs of C₂H₂ family genes in *T. aestivum* showed syntenic relationships. In four representative species, there was a syntenic relationship between C₂H₂ family genes and 65 genes in *B. distachyon*, 66 genes in *A. tauschii*, 68 genes in rice, 9 genes in *A. thaliana*, which indicated that C₂H₂ family genes in *T. aestivum* were closely related to genes in the grass family. More importantly, we found that some C₂H₂-ZF genes were associated with more gene pairs, such as Traes_2DS_2CE71025C.1 of the C₂H₂ gene family in *T. aestivum* was the syntenic relationship with KQK10684 and KQK06003. The Os05t0114400-01 is syntenic relationship

with *Traes_4AL_E7FE8D6BE.3*, *7BL_83946C3EC.1* and *1BS_527ED00B9.1*. The results indicated that these genes might have similar functions. Furthermore, there was a high degree of syntenic blocks among the *T. aestivum*, *A. tauschii*, *B. distachyon*, and rice as shown in Fig 6. In contrast, fewer gene pairs were located in syntenic blocks between *T. aestivum* and *A. thaliana*, which may be related to the phylogenetic relationship between *T. aestivum* and three plant species.

Interestingly, 33 *TaC₂H₂-ZF* genes in *T. aestivum* were found to be collinear with *ZFP* gene in *A. tauschii*, as well as with *ZFP* gene in *B. distachyon* and rice, but not found between *T. aestivum* and *A. thaliana*, such as *Traes_6DL_88DFCF552.1/KQK08112* and *4AL_24A54C8BA.2/Os05t0106000-01*, which indicated that these orthologous pairs formed after the divergence of dicotyledonous and monocotyledonous plants. Additionally, one *TaC₂H₂-ZF* gene in *T. aestivum* was identified to be collinear with *ZFP* genes in *A. tauschii*, *B. distachyon*, rice, and *A. thaliana*, indicating that these orthologous pairs may already exist before the ancestral divergence (Additional files 1: Table S3; Fig 6).

To better understand the evolutionary constraints acting on this gene family, we calculated the Ka/Ks ratios for the *TaC₂H₂-ZF* gene pairs. Generally, Ka/Ks>1 indicates positive selection, Ka/Ks = 1 indicates neutral selection, and Ka/Ks<1 indicates purification [49]. Additional files 1: Table S3 showed that the all Ka/Ks ratios were less than 1, suggesting that *TaC₂H₂-ZF* genes were undergone purifying selection, and it was speculated that their gene functions did not differentiate, which largely maintained the functional similarity of the members of the *TaC₂H₂-ZF* family. The divergence time of duplication events were inferred by Ks (Additional files 1: Table S3). We predicted that the divergence time of *C₂H₂-ZF* gene pairs should be approximately 286 Mya (*T. aestivum*), 20 Mya (*A. tauschii*), 37 Mya (*B. distachyon*), 39 Mya (rice), 64 Mya (*A. thaliana*) (Additional files 1: Table S3). The result

indicated that *TaC₂H₂-ZF* gene family shared an intimate correlation with those in *A. tauschii*, *B. distachyon*, rice, and *A. thaliana*.

Expression profiles of *T. aestivum* *TaC₂H₂-ZF* genes

To explore the expression patterns of *T. aestivum* *C₂H₂-ZF* genes during development and in response to abiotic stresses, we used publicly available RNA-seq data and mapped to the bread wheat genome by Borrill [50]. These samples included diverse developmental stages, tissues, and abiotic stress conditions. The RNA-seq data corresponded to 5 organs at 15 developmental stages [51].

To detect preferentially expressed *TaC₂H₂-ZF* genes in certain *T. aestivum* tissues and at certain stages, we found that totally 198 of 204 *TaC₂H₂-ZF* genes were expressed in all organs and development stages, and 37 of 204 *TaC₂H₂-ZF* genes were responded to heat, drought, and heat & drought abiotic stresses according to public transcriptome data.

A heat map showed the relative expression profile of 198 *TaC₂H₂-ZF* genes in 5 different organs and 15 different development stages (Fig 7a). The 204 *TaC₂H₂-ZF* genes were clustered into five groups. The 44 genes (22.2%) of group I were lowly transcript accumulated in most organs and development stages. The 35 genes (17.7%) of group II were all preferentially moderately expression in the root. Among 35 genes, the 9 genes (25.7%) were moderately expression in stem, leaf, spike and grain, and 26 genes (74.3%) were lowly transcript accumulation in stem, leaf, spike and grain. The 8 genes (4.1%) of group III were highly transcript accumulation in all organs and development stages. The 43 genes (21.7%) of group IV were moderately expressed in most organs and development stages, but in which 12 genes (27.9%) were lowly transcript accumulation in root. The 74 genes (37.4%) of group V were moderately and consecutively expressed in all organs and development stages (Fig 7a).

We found that the Ta-t1-SF members were mainly distributed into the expression group V (44.6%), while the Ta-1i-Q-SF into group I (54.5%), the Ta-2i-Q-SF into group II (60%), the Ta-1i-M-SF into group III (50%) and IV (27.9 %).

To verify the expression profiles of *TaC₂H₂-ZF* genes obtained by the RNA-seq data analysis, qRT-PCR was performed for 43 selected genes (25 *TaC₂H₂-ZF* genes that showed a large change in different tissues expression, 18 *TaC₂H₂-ZF* genes is very significantly in response to drought and heat stresses). The gene-specific primers are listed in Additional files 1: Table S4. As shown in Fig 7b and Fig 7c, among the 25 genes, 3 genes showed the highest transcript accumulation in all development stages. 4 genes in the root (seedling, three leaves, and meiosis), stem (Spike at 1 cm and two nodes), leaf (seedling, Three tillers), grain (2DAAs and 30 DAAs). The one genes showed moderately transcript accumulation in the root (three leaves and meiosis), stem (two nodes), leaf (Three tillers and 2 DAAs). The one genes showed moderately transcript accumulation in the root (three leaves and meiosis), leaf (2 DAAs). The 15 genes showed moderately transcript accumulation in the root (three leaves and meiosis), leaf (2 DAAs), spike (two nodes, meiosis, and anthesis), grain (2 DAAs and 30DAAs), respectively. These results were almost consistent with those observed in the RNA-seq data analysis, while the levels of upregulation were slightly different (Fig 7a, b). The expression of *Traes_5DL_AA1443700.1* has almost no change at all stages, which is slightly different from the results of the above-mentioned RNA-seq data analysis (Fig 7b).

Subsequently, the expression profiles of the 18 *TaC₂H₂-ZF* genes were further analyzed under drought and heat stresses, and most of which were well following microarray profiles. The qRT-PCR results were highly consistent with the RNA-seq data. Among the 18 *TaZFPs*, *Traes_2AL_1FB739291.1* and *5DL_B97546DA11.1* were observed to be increased in response to

drought and heat stress. The *4DS_276818E4F.1* was upregulated significantly at HD-1h. The *4AL_B808CA890.1* was suppressed after DS-1h, DS-6h and HS-1h treatments, whereas it was upregulated after HS-6h, HD-1h, and HD-6h treatment (Fig 7c). Notably, *Traes_5BL_B038BE0A3.1*, *5BL_3A9D3CEEF.1* and *5BL_D53A846BE.1* displayed significantly high expression levels at all treatment time points. The *5BL_5391EC4A0.1* and *5DL_4E890ABF3.1* displayed significantly high expression levels at DS-1h, DS-6h, HS-1h, HD-1h and HD-6h treatment time points.

In summary, our results showed that these selected *TaC₂H₂-ZF* genes responded to one or more stresses. According to expression pattern analysis found, like *Traes_5BL_D53A846BE.1*, not only expressed highly in L2DAAs, RTLS, RMS, but also endowed tolerance to drought and heat stresses, making them good candidates for molecular breeding.

Discussion

The C₂H₂-ZF genes of *T. aestivum* have relatively conservative structure in evolution

The C₂H₂-type zinc finger proteins belongs to one of the largest plant TF families and participates in multiple biological processes, including plant growth, development, and stress responses. Therefore, *T. aestivum* C₂H₂-ZF genes may be promising targets for crop breeding and improvement.

The 457 C₂H₂-ZF genes (including 253 splice variants) of the *T. aestivum* were classified into four great categories (named Q, M, Z, and D) according to the variation of the plant-specific conserved amino acid sequence “QALGGH” and distances between metal ligands [1, 2] (Fig 1). The TaC₂H₂-ZFPs were classified into two subsets Ta-tandem-SF (46.8%) and Ta-isolated-SF (53.2%). The Ta-tandem-SF contained two clearly subsets (Ta-t1-SF and Ta-t2-SF), The Ta-isolated-SF contained ten clearly subsets (Ta-1i-Q-SF, Ta-1i-M-S, Ta-1i-Z-SF, Ta-2i-Q-SF, Ta-2i-M-S, X-2i-Mix-SF, X-3i-SF and X-4i-SF) (Additional files 1: Table S1) [2, 3]. Many previous studies have demonstrated that the molecular functions of the proteins with tandem C₂H₂-ZF domain are usually different from the C₂H₂-ZFPs with dispersed fingers [52, 53, 54]. So, the proteins with single or dispersed C₂H₂-ZFP domains are separated performed category from tandem C₂H₂-ZF domain proteins [54]. According to the statistical results of adjacent lengths in *T. aestivum* C₂H₂-ZFPs (Additional files 1: Table S1), C₂H₂-ZFPs with two or more fingers and every two adjacent domains linked by <12 amino acid residues were defined as tandem ZFs, which was the same as the tandem ZFs in soybean and *Brassica rapa* L [2, 3]. There was no Q-type C₂H₂-ZF domain in these C₂H₂-ZFPs, similar to the yeast and animals [52, 54] (Additional files 1: Table S1). Furthermore, the Q-type and M-type C₂H₂-ZF domains were not the same, as those were defined in previous studies [22, 23]. The rest two types were the same as the types defined in previous reports [2, 3]. The Z-type domain was defined when the spacing

between the second cysteine and the first histidine was <12 amino acids, and the D-type did not include the second histidine in the C₂H₂-ZF domain. This classification was well supported based on phylogenetic and protein conserved domain analysis. Most of the C₂H₂-ZF domains members in the same subset were clustered together (Fig 3, Fig 4). The results indicated that the sequences and biological functions of TaZFPs were relatively conservative.

Subfamily specific motifs (SFP1-, DL-, R-, PL-, L-, EK-, and NPL-box) of *T. aestivum* C₂H₂-ZF genes may contribute to the high adaptability of *T. aestivum*

Subsequently, sequence logo analysis to explore the sequence characteristics of the C₂H₂ domains from the ten major subsets were studied by the MEME motif search tool. The TaC₂H₂-ZF domains not only contain the C₂H₂ domain but also contain Nucleoplasmin-like (NPL) [44, 45] and Schizosaccharomyces pombe finger protein 1 (SFP1) domain [46, 47]. They also contain DL-box, R-box, PL-box, L-box, and EK-box (Fig 5). These motifs were previously reported in rice and *A. thaliana*, these findings were in line with previous findings in *T. aestivum* [25]. The DL-box is similar to the previously reported LDLSL-EAR motif and may contain transcriptional repressor functions [48]. For instance, overexpression of intact DREB/ERF proteins with an EAR motif led to the reduced expression levels of stress-related genes and decreased abiotic stress tolerance [26, 55, 56]. Overexpression of DREBs/ERFs with the EAR motif in rare cases led to increases in stress tolerance in soybean. Constitutive expression of *GmERF4* in transgenic tobacco plants increased tolerance to salt and drought stresses compared with wild-type plants [57]. Most of these C₂H₂-type zinc finger proteins contain a DLN-box/EAR-motif, which functions as a transcription repression domain, this core sequence has been regarded as a signature for transcriptional repressor activity in C₂H₂ zinc fingers.

The STZ/ZAT10 and ZPT2-3 with an EAR-motif have been proven to have the transcription repressors function [7, 58]. Thus, it could be speculated that some TaC₂H₂-ZF proteins might function as a transcription repressor due to the contain of DLN-box/EAR-motif in C₂H₂-ZF protein sequence.

The above analysis showed these TaC₂H₂-ZFPs had the high functional diversity in plant growth and development. This study cast a new light on multifunction of TaC₂H₂-ZF genes, and provided a certain reference value for people to understand the functions of the C₂H₂-ZF genes in *T. aestivum*. It is also worth discussing these significant facts revealed by the results of the above analysis.

Expression profile of *T. aestivum* C₂H₂-ZFPs in different tissues and different developmental stages and Multiple abiotic stresses

The RNA-seq analysis showed that 198 of 204 *T. aestivum* C₂H₂-ZF genes were expressed in 5 organs at 15 developmental stages. In all the tested organs and developmental stages, about 4.1% TaC₂H₂-ZF genes were almost highly and consecutively expressed, 73.7% TaC₂H₂-ZF genes were almost moderately expressed, and 22.2% TaC₂H₂-ZF genes were lowly expressed, illustrating the functional diversity of the C₂H₂-ZFP transcription factor in different tissues and different developmental stages (Fig 7a). Notably, 134 (67.7%) of the TaC₂H₂-ZF genes were moderately and consecutively expressed in root and leaf (Fig 7a). In plants, roots and leaves are the most important organs necessary for nutrient absorption and photosynthesis, respectively [59]. Also, *T. aestivum* leaves are important organs that first capture UV-B radiation [60]. The plant leaf through increasing stomatal closure, decreasing the transpiration rate, and reducing the leaf area provide an adaptive mechanism for plants undergoing abiotic stress [16, 61]. Roots are important plant organs that nourish entire plants by taking up water and essential minerals [62]. The previous study show that the plasticity of fine root growth may

increase plant productivity [63]. Roots can sense soil changes to reduce root damage and maintain plant growth under water shortage conditions [64]. These results indicated that bread wheat C_2H_2 -ZFPs played roles in root and leaf development more than other tissues or organs. There is considerable evidence that C_2H_2 -ZFP transcription factors are involved widely in the development of multiple organs and tissues and abiotic stresses, such as seed maturation, floral development, spike development [1], salt stress tolerance [15, 65], drought stress tolerance [34].

In this study, we examined the expression patterns of 43 selected genes. The 25 *TaC₂H₂-ZF* genes showed significant difference in different tissues expressions (chromosome 1-7 account for 12.1, 19.2%, 10.6%, 16.7%, 27.3%, 7.6%, and 6.6%, respectively). The 18 *TaC₂H₂-ZF* genes were very significant in response to drought and heat stresses (chromosome 1-7 account for 8.1%, 8.1%, 8.1%, 21.6%, 43.2%, 10.8%, and 0%, respectively) (Fig 7a). Previous studies have shown that some *C₂H₂-ZF* genes located in chromosome 5 play important roles in the bread wheat growth development stage and abiotic stress processes [66, 67, 68]. And most results were broadly consistent with the microarray profiles by real-time quantitative RT-PCR verification.

The *Traes_5DL_6CBF61AA4.1* and *5BL_7F56360BF.1* were homologous to *A. thaliana AtZAT10* (*AT1G27730.1*, also called *STZ*). The three genes belong to the same subsets and all of them contain not only the Q-type C_2H_2 -ZF domain but also a FDLNI (EAR-like motif) and a EK-box. *Traes_5DL_6CBF61AA4.1* and *5BL_7F56360BF.1* were highly expressed in the SR, RTLS, RMS, S1S, STNS, LTTS and L2DAAs stages. Moreover, transgenic *A. thaliana* overexpressing *AtZAT10* showed growth retardation as demonstrated in a previous study [7, 25], suggesting that the functions of three homologous genes were probably conserved in both species.

The *Traes_4BS_FE945D2FC.1*, homologous to *A. thaliana AT3G19580.2* (*AZF2*), contains 2

Q-type C₂H₂-ZF domains, a FDLNI (EAR-like-motif), an EK-box, and a L-box. Previous analysis showed that expression of *AZF2* was strongly induced by drought stresses [7, 25]. The *Traes_4BS_FE945D2FC.1* was significantly up-regulated at 6 h (up to 10-fold) under heat & drought stresses, suggesting it had a similar function in drought stresses with *AZF2*.

The *Traes_5BL_D53A846BE.1* highly homologous to *A. thaliana* *AT2G37430.1* (*ZAT11*) and *AT5G59820* (*ZAT12*), contains 2 Q-type C₂H₂-ZF domains. The *Traes_5BL_D53A846BE.1*, not only highly expressed in the RTLS, RMS and L2DAAs stages, but also responded to drought and heat stresses, making them good candidates for molecular breeding. While the *ZAT11* is highly expressed in roots and particularly in root tips [69]. The *ZAT12* can reduce heat-derived oxidative stress by activating multiple defense genes in tomato [70]. These results indicated that all the tested genes were more or less involved in development stages and stress response in *T. aestivum*, in a similar way as their orthologs in *A. thaliana*. Further tissue expression and stress responsive expression analysis of a larger number of *T. aestivum* C₂H₂-ZF genes, including those lowly expressed genes, would allow to identify the most active members in gramineae crops.

The above results indicated that these C₂H₂-ZF genes might be involved in the important regulatory networks or cross-talk triggered by different responses stresses. Overall, these findings provide insight into the potential functional roles of *T. aestivum* C₂H₂-ZF genes. The results provide a basis for further functional study of C₂H₂-ZF genes.

Conclusions

In the present study, we systematically analyzed the characterization of the TaZFP family. The *C₂H₂-ZF* genes in the bread wheat genome were identified (including 253 splice variants), according to defined *C₂H₂-ZF* domain types, and *C₂H₂-ZFPs* were classified into 10 distinct subsets. A total of 204 non-redundant candidate genes were obtained to encode 457 proteins. The 85 genes (41.67%) were encoded on the B sub-genome and 115 genes (56.37%) were distributed on long arms of the chromosomes. Synteny analysis showed that only 4 pairs of *C₂H₂* family genes in *T. aestivum*, 65 genes in *B. distachyon*, 66 genes in *A. tauschii*, 68 genes in rice, 9 genes in *A. thaliana*, were syntenic relationships respectively, indicating that TaZFPs were closely related to genes in the *Poaceae*. The specific conserved motifs of the *C₂H₂-type* zinc finger domain including SFP1-, DL- (EAR-like-motif), R-, PL-, L-, EK- and NPL-box were discovered, indicating the different subset members maybe contain a novel motif and execute different biological functions. According to public transcriptome data, totally 198 of 204 *TaC₂H₂-ZF* genes were found to have the expression data. The 25 *TaC₂H₂-ZF* genes were certificated to be significantly differentially expressed in 5 different organs and 15 different development stages by quantitative RT-PCR. The 18 *TaC₂H₂-ZF* genes were verified in response to heat, drought, and heat & drought stresses. Such as *Traes_5BL_D53A846BE.1*, it was not only highly expressed in L2DAAs, RTLS and RMS, but also endowed tolerance to drought and heat stresses, making them good candidates for molecular breeding. In summary, the present study provides the valuable reference information for further study of *C₂H₂-ZF* genes, and provides promising targets for further genetic engineering and genetic improvement in *T. Aestivum*.

Methods

Plant materials and abiotic stress treatments

Bread wheat (*Triticum aestivum* L. cv. Fielder) was used in this study. Seeds were germinated on wet paper towel and cultivated at 4°C for 5 days, then placed at 12°C for 5 day for germination. Germinated seedlings were grown at 22°C in a greenhouse. Two-week-old wheat seedlings were treated by different stresses (heat, drought, and heat & drought). The abiotic stress treatments were performed by submerging *T. aestivum* seedling roots in solutions of 37°C incubator, 25% PEG6000, respectively. *T. aestivum* seedlings treated with water were used as a mock control. The whole *T. aestivum* seedlings were sampled at different time points (0, 1, 6 h). Greenhouse-grown *T. aestivum* plants were used for measuring tissues specific expression patterns of 25 selected *TaC₂H₂-ZF* genes. Seedling roots (SR), roots at the three leaves (RTLS), roots at the meiosis (RMS), stems at the 1 cm spike (S1S), stems at the two nodes stems (STNS), stems at the anthesis stems (SATS), seedling leaves (SL), leaves at the three tillers (LTTS), leaves at the 2 days after anthesis (L2DAAs), spikes at the two nodes stems (SPTNS), spikes at the meiosis (SPMS), spikes at the anthesis stems (SPAS), grains at the 2 days after anthesis (G2DAAs), grains at the 14 days after anthesis (G14DAAs), grains at the 30 days after anthesis (G30DAAs) were collected. All collected tissue samples were immediately frozen in liquid nitrogen and stored at -80 °C for future analysis.

Data sets

The annotated genome sequences of *Aegilops tauschii*, *Brachypodium distachyon*, *Oryza sativa* Japonica (rice) and *Arabidopsis thaliana* were downloaded from plantTFDBv5.0, (http://planttfdb.cbi.pku.edu.cn/phylo_tree.php?sp=Tae&fam=C₂H₂) [29], iTAK software v1.6

(<http://itak.feilab.net/cgi-bin/itak/index.cgi>) [30], Ensembl plants
(http://plants.ensembl.org/Triticum_aestivum/Info/Index). TAIR (Release
10, ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release) and JCVI, (Release 6.1,
[ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/
version_6.1](ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_6.1)), respectively. The partially annotated genome sequences of *Triticum aestivum* and *Oryza
sativa* were downloaded from Prosite (<http://au.expasy.org/prosite/>) and NCBI
(<https://www.ncbi.nlm.nih.gov/>).

Identification and characterization of *Triticum aestivum* C₂H₂-ZFPs

For identification of *T. aestivum* C₂H₂-ZFPs family members, the information of *T. aestivum*
C₂H₂-ZFPs sequences was downloaded from the plantTFDBv5.0, iTAK database, and the Ensembl
plants. The C₂H₂-ZF domain was also queried for Hidden Markov Model (HMM) search through the
HMMER 3.0 program. The TaC₂H₂-ZF domain was further determined using the SMART database
(<http://smart.embl-heidelberg.de/>) [31].

The theoretical isoelectric point (pI), molecular weight (MW), atomic composition, instability index,
aliphatic index, and grand average of hydropathicity (GRAVY) were downloaded from the ExPASy
server (<http://www.expasy.org/>) [32]. The subcellular localization of each TaC₂H₂-ZF protein was
predicted using the Cell-PLoc 2.0 (<http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/>) webserver [33].

Classification and chromosomal location of *Triticum aestivum* C₂H₂-ZFPs

The identified 457 TaC₂H₂-ZFPs were further manually analyzed to search the numbers of C₂H₂-ZF
domain, C₂H₂-ZF domain sequences, and the space length between C₂H₂-ZFPs using the SMART

database (<http://smart.embl-heidelberg.de/>). We defined the C₂H₂ type for each identified bread wheat
 C₂H₂ domain, according to the plant-specific amino acid residues and distances between two to nine
 C₂H₂-ZF domains, as have been previously adopted in soybean [3]. Tandem ZF was determined as ZF
 containing two to thirteen C₂H₂-ZF domains, and every two nearby domains are connected by <12
 amino acid residues, such as Tandem ZFs with two subsets assigned as Ta-t1-SF and Ta-t2-SF,
 respectively. ZF contains one ZF and/or two to four ZF domains each more than 11 amino acid residues
 between two nearby domains is considered to be scattered ZF. Also, different types of C₂H₂-ZFPs have
 been classified based on the variation of the plant-specific conserved domain "QALGGH" sequence
 and distances between metal ligands [1, 4, 34]. Generally, C₂H₂-ZFP were defined as four types: Q-type,
 M-type, Z-type, and D-type. C₂H₂-ZFP Q-type domains were defined as
 X2-C-X2-C-X7-QALGGH-X3-H. Those that have 1-5 amino acid degradations in the "QALGGH"
 domain and that there are certain modifications in the spacing between two cysteine and two histidines
 of Q type are defined as M-type. Those with more than 12 (Z1) and less than 12 (Z2) in their spacing
 between the second cysteine and the first histidine were defined as Z-type domains. However,
 compared with Q-type, M-type, and Z-type, the D-type does not include the second histidine in the
 C₂H₂-ZF domain. According to these defined C₂H₂-ZFP types, C₂H₂-ZFPs containing a single ZF
 domain were further classified into four clearly distinguishable subsets (Ta-1i-Q-SF, Ta-1i-M-SF,
 Ta-1i-Z-SF, and Ta-1i-D-SF). C₂H₂-ZFPs containing two Q-type or two M-type C₂H₂-ZF domains were
 defined as Ta-2i-Q-SF and Ta-2i-M-SF, respectively, while the C₂H₂-ZFPs containing two different
 types of C₂H₂-ZF domains were classified as Ta-2i-Mix-SF. All ZFPs containing three or four dispersed
 ZFs were classified into subsets of Ta-3i-SF or Ta-4i-SF, respectively [2, 3].

The *TaC₂H₂-ZF* genes approximate chromosomal location was determined by blastn search of

cDNA sequences against chromosome sequences of *T. aestivum* available at Ensembl Plants (http://plants.ensembl.org/Triticum_aestivum/Tools/Blast). The IWGSC bread wheat genome was assembled using the following parameters: E value $\leq 1e-10$ and identity 85% [35].

The chromosomal location image of the *TaC₂H₂-ZF* genes was generated by TBtools software, according to the chromosomal position information provided in the EnsemblPlants database.

Analyses of phylogeny, gene structure and conserved motif

For the phylogenetic investigation, multiple alignments were performed with the domain sequences of the TaZFP family proteins by the neighbor-joining (NJ) method available via MEGA7 software [37] to construct an unrooted phylogenetic tree. Sequences were aligned using the MUSCLE alignment algorithm [37]. The tree obtained was the consensus of 1000 single trees provided by bootstraps, gaps/missing data treatment: partial deletion, model/method: LG model, rates among sites: gamma distributed with invariant sites (G+I).

For the *TaC₂H₂-ZF* gene exon-intron structure, we used the Gene Structure Display Server (GSDS 2.0) (<http://gsds.cbi.pku.edu.cn>) [38]. We analyzed the additional motifs outside the C₂H₂ domain of 204 C₂H₂-ZFP amino acid sequences using the Multiple Em for Motif Elicitation (MEME) suite Version 5.1.1 online program (<http://meme-suite.org/tools/meme>). The optimized parameters of MEME were employed as the following: maximum number of motifs, 10; minimum motif width, 6; and maximum motif width, 50.

Chromosomal distribution and syntenic analysis of *TaC₂H₂-ZF* genes

The *TaC₂H₂-ZF* genes were mapped to the chromosome according to the chromosomal location given from the Ensembl plants database by using CIRCA (OMGenomics, <http://omgenomics.com/circa/>). Investigating the gene duplication events of having the Default parameters using the Multiple Collinearity Scan toolkit (MCScanX) [39].

To exhibit the syntenic relationship of the orthologous *C₂H₂-ZF* genes obtained from *T. aestivum* and *A. tauschii*, *B. distachyon*, *A. thaliana*, *O. Japonica* the syntenic analysis maps were constructed using the Dual Systemy Plotter software (<https://github.com/CJ-Chen/TBtools>) [39]. Synonymous (Ks) and non-synonymous (Ka) substitution rates were calculated for each duplicated *C₂H₂-ZF* genes using a maximum likelihood model averaging in Ka/Ks Calculator [39]. For each gene pair, the divergence time of collinear gene pairs was calculated using the mean Ks values from $T = Ks / (2\lambda \times 10^{-6})$ Mya ($\lambda = 6.5 \times 10^{-9}$) [40].

Expression analysis of *C₂H₂-ZF* genes in *Triticum aestivum*

The expression profiles of 198 *TaC₂H₂-ZF* genes in different tissues (roots, stem, leaves, spikes, and grains) and different development stages were determined based on the RNA seq data available from Wheat Expression Browser powered by expVIP (<http://www.wheat-expression.com/>) [41]. Briefly, the raw data for RNA-Seq were downloaded, and subsequent data processing was performed to remove weakly expressed genes using the R software package. Finally, the gene expression heatmap was generated by using the TBtools view software.

RNA extraction and qRT-PCR analyses

Total RNA from different tissues (root, stem, leave, spike, grain) and various stress treated materials

were extracted using Trizol reagent (Invitrogen) following the manufacturer's instructions. Subsequently, RNA integrity and quality were checked by electrophoresis of RNA on 1.2% denaturing agarose gel. For qPCR analysis, first-strand cDNA synthesis was performed with the SuperScript II Reverse Transcriptase (Invitrogen) by the manufacturer's instructions. 25 tissue-specific *TaC₂H₂-ZF* genes expression patterns and 18 *TaC₂H₂-ZF* genes significantly in response to heat and drought stresses expression patterns were determined by qRT-PCR analysis. Gene-specific primer pairs in Additional files 1: Table S4. Accuracy and specificity of primers were checked by the Blast algorithm and their products were electrophoretically checked on 1.5% agarose gel for amplification accuracy. *TaRP15* (Additional files 1: Table S4) was used as the internal reference gene when examining gene expression of 25 selected *TaC₂H₂-ZF* genes. the qRT-PCR analysis was performed to detect relative mRNA expression levels. During the qRT-PCR analysis, each qRT-PCR analysis was repeated three times, and data analyzed by analysis software based on the comparative $2^{-\Delta\Delta CT}$ method of relative gene quantification [3]. Relative expression of 25 selected *TaC₂H₂-ZF* genes and probability value (P) were also calculated using the qRT-PCR analysis software.

Supplementary information

Additional files 1: Table S1 Classification of the 457 *Triticum aestivum* C₂H₂-ZFPs according to the organization of their contained C₂H₂ fingers, **Table S2** List of 204 *Triticum aestivum* C₂H₂-ZFP genes and their related information, **Table S3** One-to-one orthologous relationships between *T.aestivum* and *T.aestivum*, *T.aestivum* and *A.tauschii*, *T.aestivum* and *B.distachyon*, *T.aestivum* and Rice, *T.aestivum* and *Arabidopsis*, **Table S4** Primers used for qRT-PCR.

Abbreviations

C₂H₂-ZFPs: C₂H₂-type zinc finger proteins; *T. Aestivum*: *Triticum aestivum*; *A. tauschii*: *Aegilops tauschii*; *B. distachyon*: *Brachypodium distachyon*, *O. Japonica*: *Oryza sativa Japonica* (rice); *A. thaliana*: *Arabidopsis thaliana*; CDS: Coding sequence; pI: isoelectric point; MW: molecular weight; GRAVY: Grand average of hydropathicity; NJ: neighbor-joining; GSDS: Gene Structure Display Server; MEME: Multiple Em for Motif Elicitation; MCScanX: Multiple Collinearity Scan toolkit; SR: seedling roots; RTLS: roots at the three leaves; RMS: roots at the meiosis stems; S1S: at the 1 cm spike stems; STNS: at the two nodes stems; SATS: stems at the anthesis stems; SL: seedling leaves; LTTS: leaves at the three tillers, L2DAAs: leaves at the 2 days after anthesis; SPTNS: spikes at the two nodes stems; SPMS: spikes at the meiosis; SPAS: spikes at the anthesis stems; G2DAAs: grains at the 2 days after anthesis; G14DAAs: grains at the 14 days after anthesis; G30DAAs: grains at the 30 days after anthesis; qRT-PCR: Quantitative real-time PCR; TFs: Transcription factors

Acknowledgements

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

Author contributions

X.G. and M.Z. designed the experiments. Y.L. and X.G. wrote the main manuscript text. Y.L. and A.S. conducted the experiments. A.S., Q.W., X.Z., F.C., R.C. and H.X. collected and analysed phenotype data. Y.L., M.Z. and A.S. prepared Figures 1-7. All authors read and approved the manuscript.

Funding

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

Availability of data and materials

The datasets analysed during the current study are available in the bread wheat Genome Project
(http://plants.ensembl.org/Triticum_aestivum/Tools/Blast), plantTFDBv5.0
(http://planttfdb.cbi.pku.edu.cn/phylo_tree.phpsp=Tae&fam=C2H2), iTAK software v1.6
(<http://itak.feilab.net/cgi-bin/itak/index.cgi>), Ensembl plants
(http://plants.ensembl.org/Triticum_aestivum/Info/Index). TAIR (Release
10, ftp://ftp.arabidopsis.org/home/tair/Genes/TAIR10_genome_release) and JCVI (Release 6.1,
[ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/](ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_6.1)
version_6.1) repository.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

College of Biology, Hunan University, Changsha 410082, China

596 **References**

- 597 1. Ming N , Ma N , Jiao B , et al. Genome Wide Identification of C₂H₂-Type Zinc Finger
598 Proteins of Tomato and Expression Analysis Under Different Abiotic Stresses. Plant
599 Molecular Biology Reporter, 2020;38(1):75-94.
- 600 2. Intikhab, Alam, Khadija, et al. Comprehensive genomic survey, structural classification and
601 expression analysis of C₂H₂ zinc finger protein gene family in Brassica rapa L. PloS one,
602 2019;14(5):e0216071.
- 603 3. Yuan S, Li X, Li R, Wang L, Zhang C, Chen L, Hao Q, Zhang X, Chen H, Shan Z, Yang Z,
604 Chen S, Qiu D, Ke D, Zhou X. Genome-Wide Identification and Classification of Soybean
605 C₂H₂ Zinc Finger Proteins and Their Expression Analysis in Legume-Rhizobium Symbiosis.
606 Front Microbiol. 2018;9:126.
- 607 4. Liu D, He S, Song X, Zhai H, Liu N, Zhang D, et al. *IbSIMT1*, a novel salt-induced methyl
608 transferase gene from Ipomoea batatas, is involved in salt tolerance. Plant Cell Tissue Organ
609 Cult. 2015;120: 701–715.
- 610 5. Cheuk A, Houde M. Genome wide identification of C1-2i zinc finger proteins and their
611 response to abiotic stress in hexaploid wheat. Mol Genet Genomics. 2016;291(2):873-90.
- 612 6. Cheng X, Cao J, Gao C, Gao W, Yan S, Yao H, Xu K, Liu X, Xu D, Pan X, Lu J, Chang C,
613 Zhang H, Ma C. Identification of the wheat C3H gene family and expression analysis of
614 candidates associated with seed dormancy and germination. Plant Physiol Biochem.
615 2020;156:524-537.
- 616 7. Sakamoto H, Maruyama K, Sakuma Y, Meshi T, Iwabuchi M, Shinozaki K,
617 Yamaguchi-Shinozaki K. Arabidopsis Cys2/His2-type zinc-finger proteins function as
618 transcription repressors under drought, cold, and high-salinity stress conditions. Plant Physiol.
619 2004;136(1):2734-46.
- 620 8. Davletova S, Schlauch K, Coutu J, Mittler R. The zinc-finger protein *Zat12* plays a central
621 role in reactive oxygen and abiotic stress signaling in Arabidopsis. Plant Physiol. 2005;
622 139(2):847-856.
- 623 9. Mittler R, Kim Y, Song L, Coutu J, Coutu A, Ciftci-Yilmaz S, Lee H, Stevenson B, Zhu JK.
624 Gain- and loss-of-function mutations in *Zat10* enhance the tolerance of plants to abiotic stress.
625 FEBS Lett. 2006;580(28-29):6537-42.
- 626 10. Ciftci-Yilmaz S, Morsy MR, Song L, Coutu A, Krizek BA, Lewis MW, Warren D, Cushman
627 J, Connolly EL, Mittler R. The EAR-motif of the Cys2/His2-type zinc finger protein *Zat7*
628 plays a key role in the defense response of Arabidopsis to salinity stress. J Biol Chem.
629 2007;282(12):9260-8.
- 630 11. Devaiah BN, Nagarajan VK, Raghothama KG. Phosphate homeostasis and root development
631 in Arabidopsis are synchronized by the zinc finger transcription factor ZAT6. Plant Physiol.
632 2007;145(1):147-59.
- 633 12. Yamaguchi-Shinozaki K. Arabidopsis Cys2/His2 zinc-finger proteins AZF1 and AZF2
634 negatively regulate abscisic acid-repressive and auxin-inducible genes under abiotic stress
635 conditions. Plant Physiol. 2011;157(2):742-56.
- 636 13. Yin M, Wang Y, Zhang L, Li J, Quan W, Yang L, Wang Q, Chan Z. The Arabidopsis
637 Cys2/His2 zinc finger transcription factor ZAT18 is a positive regulator of plant tolerance to
638 drought stress. J Exp Bot. 2017;68(11):2991-3005.

- 639 14. Huang J, Wang JF, Wang QH, Zhang HS. Identification of a rice zinc finger protein whose
640 expression is transiently induced by drought, cold but not by salinity and abscisic acid. *DNA*
641 *Seq.* 2005;16(2):130-6.
- 642 15. Huang J, Yang X, Wang MM, Tang HJ, Ding LY, Shen Y, Zhang HS. A novel rice C₂H₂-type
643 zinc finger protein lacking DLN-box/EAR-motif plays a role in salt tolerance. *Biochim*
644 *Biophys Acta.* 2007;1769(4):220-7.
- 645 16. Huang XY, Chao DY, Gao JP, Zhu MZ, Shi M, Lin HX. A previously unknown zinc finger
646 protein, DST, regulates drought and salt tolerance in rice via stomatal aperture control. *Genes*
647 *Dev.* 2009;23(15):1805-17.
- 648 17. Xu DQ, Huang J, Guo SQ, Yang X, Bao YM, Tang HJ, Zhang HS. Overexpression of a
649 TFIIIA-type zinc finger protein gene *ZFP252* enhances drought and salt tolerance in rice
650 (*Oryza sativa* L.). *FEBS Lett.* 2008;582(7):1037-43.
- 651 18. Sun SJ, Guo SQ, Yang X, Bao YM, Tang HJ, Sun H, Huang J, Zhang HS. Functional analysis
652 of a novel Cys2/His2-type zinc finger protein involved in salt tolerance in rice. *J Exp Bot.*
653 2010;61(10):2807-18.
- 654 19. Zhang H, Liu Y, Wen F, Yao D, Wang L, Guo J, Ni L, Zhang A, Tan M, Jiang M. A novel rice
655 C₂H₂-type zinc finger protein, ZFP36, is a key player involved in abscisic acid-induced
656 antioxidant defence and oxidative stress tolerance in rice. *J Exp Bot.* 2014;65(20):5795-809.
- 657 20. Liu Z, Xin M, Qin J, Peng H, Ni Z, Yao Y, Sun Q. Temporal transcriptome profiling reveals
658 expression partitioning of homeologous genes contributing to heat and drought acclimation in
659 wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 2015;15(1):1-20.
- 660 21. Wu S, Tong X, Li C, Lu K, Tan D, Hu H, Liu H, Dai F. Genome-wide identification and
661 expression profiling of the C₂H₂-type zinc finger protein genes in the silkworm *Bombyx mori*.
662 *PeerJ.* 2019;7:e7222.
- 663 22. Muthamilarasan M, Bonthala VS, Mishra AK, Khandelwal R, Khan Y, Roy R, Prasad M.
664 C₂H₂-type of zinc finger transcription factors in foxtail millet define response to abiotic
665 stresses. *Funct Integr Genomics.* 2014;14(3):531-43.
- 666 23. Liu Y, Liu D, Hu R, Hua C, Ali I, Zhang A, Liu B, Wu M, Huang L, Gan Y. AtGIS, a C₂H₂
667 zinc-finger transcription factor from *Arabidopsis* regulates glandular trichome development
668 through GA signaling in tobacco. *Biochem Biophys Res Commun.* 2017;483(1):209-215.
- 669 24. Takatsuji H. Zinc-finger transcription factors in plants. *Cell Mol Life Sci.*
670 1998;54(6):582-96.
- 671 25. Kam J, Gresshoff PM, Shorter R, Xue GP. The Q-type C₂H₂ zinc finger subfamily of
672 transcription factors in *Triticum aestivum* is predominantly expressed in roots and enriched
673 with members containing an EAR repressor motif and responsive to drought stress. *Plant Mol*
674 *Biol.* 2008;67(3):305-22.
- 675 26. Amalraj A, Luang S, Kumar MY, Sornaraj P, Eini O, Kovalchuk N, Bazanova N, Li Y, Yang
676 N, Eliby S, Langridge P, Hrmova M, Lopato S. Change of function of the wheat
677 stress-responsive transcriptional repressor TaRAP2.1L by repressor motif modification. *Plant*
678 *Biotechnol J.* 2016;14(2):820-32.
- 679 27. Jing Y, Liu J, Liu P, Ming D, Sun J. Overexpression of *TaJAZ1* increases powdery mildew
680 resistance through promoting reactive oxygen species accumulation in bread wheat. *Sci Rep.*
681 2019; 9(1): 1-15.
- 682 28. Su Q, Zhang X, Zhang W, Zhang N, Song L, Liu L, Xue X, Liu G, Liu J, Meng D, Zhi L, Ji J,

- Zhao X, Yang C, Tong Y, Liu Z, Li J. QTL Detection for Kernel Size and Weight in Bread Wheat (*Triticum aestivum* L.) Using a High-Density SNP and SSR-Based Linkage Map. *Front Plant Sci.* 2018;9:1484.
29. Tian F, Yang DC, Meng YQ, Jin J, Gao G. PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res.* 2020;48(D1):D1104-D1113.
30. Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ, Zhao PX, Rhee SY, Fei Z. iTAK: A Program for Genome-wide Prediction and Classification of Plant Transcription Factors, Transcriptional Regulators, and Protein Kinases. *Mol Plant.* 2016;9(12):1667-1670.
31. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* 2015 ;43(Database issue):D257-60.
32. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* 2003;31(13):3784-3788.
33. Chou KC, Shen HB. Cell-PLoc: a package of Web servers for predicting subcellular localization of proteins in various organisms. *Nat Protoc.* 2008;3(2):153-62.
34. Zhang D, Tong J, Xu Z, Wei P, Xu L, Wan Q, Huang Y, He X, Yang J, Shao H, Ma H. Soybean C₂H₂-Type Zinc Finger Protein GmZFP3 with Conserved QALGGH Motif Negatively Regulates Drought Responses in Transgenic Arabidopsis. *Front Plant Sci.* 2016;7:325.
35. Faraji S, Rasouli SH, Kazemitabar SK. Genome-wide exploration of C₂H₂ zinc finger family in durum wheat (*Triticum turgidum* ssp. Durum): insights into the roles in biological processes especially stress response. *Biometals.* 2018;31(6):1019-1042.
36. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2016;33(7):1870–1874.
37. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–1797.
38. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics.* 2015;31(8):1296-7.
39. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant.* 2020;13(8):1194-1202.
40. Zhu T, Liu Y, Ma L, Wang X, Zhang D, Han Y, Ding Q, Ma L. Genome-wide identification, phylogeny and expression analysis of the SPL gene family in wheat. *BMC Plant Biol.* 2020;20(1):420.
41. Okonechnikov K, Golosova O, Fursov M; UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics.* 2012;28(8):1166-7.
42. Danielli L, Li X, Tuler T, Daniel R. Quantifying the distribution of protein oligomerization degree reflects cellular information capacity. *Sci Rep.* 2020;10(1):17689.
43. Bailey TL, Elkan C. The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol.* 1995;3:21-9.
44. Kozowska M, Tarczewska A, Jakob M, Bystranowska D, Taube M, Kozak M, Czarnocki-Cieciura M, Dziembowski A, Orowski M, Tkocz K, Oyhar A. Nucleoplasmin-like domain of FKBP39 from *Drosophila melanogaster* forms a tetramer with partly disordered

- tentacle-like C-terminal segments. *Sci Rep.* 2017;7:40405.
45. Anselm E, Thomae AW, Jeyaprakash AA, Heun P. Oligomerization of Drosophila Nucleoplasmin-Like Protein is required for its centromere localization. *Nucleic Acids Res.* 2018;46(21):11274-11286.
 46. Albert B, Tomassetti S, Gloor Y, Dilg D, Mattarocci S, Kubik S, Hafner L, Shore D. Sfp1 regulates transcriptional networks driving cell growth and division through multiple promoter-binding modes. *Genes Dev.* 2019;33(5-6):288-293.
 47. Lee SY, Chen HF, Yeh YC, Xue YP, Lan CY. The Transcription Factor Sfp1 Regulates the Oxidative Stress Response in *Candida albicans*. *Microorganisms.* 2019;7(5):131.
 48. Kazan K. Negative regulation of defence and stress genes by EAR-motif-containing repressors. *Trends Plant Sci.* 2006;11(3):109-12.
 49. Zhang W, Yan H, Chen W, Liu J, Jiang C, Jiang H, Zhu S, Cheng B. Genome-wide identification and characterization of maize expansin genes expressed in endosperm. *Mol Genet Genomics.* 2014;289(6):1061-74.
 50. Borrill P, Harrington SA, Uauy C. Genome-Wide Sequence and Expression Analysis of the NAC Transcription Factor Family in Polyploid Wheat. *G3 (Bethesda).* 2017;7(9):3019-3029.
 51. Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E, Leroy P, Mangenot S, Guilhot N, Le Gouis J, Balfourier F, Alaux M, Jamilloux V, Poulain J, Durand C, Bellec A, Gaspin C, Safar J, Dolezel J, Rogers J, Vandepoele K, Aury JM, Mayer K, Berges H, Quesneville H, Wincker P, Feuillet C. Structural and functional partitioning of bread wheat chromosome 3B. *Science.* 2014;345(6194):1249721.
 52. Böhm S, Frishman D, Mewes HW. Variations of the C₂H₂ zinc finger motif in the yeast genome and classification of yeast zinc finger proteins. *Nucleic Acids Res.* 1997;25(12):2464-9.
 53. Wolfe SA, Nekludova L, Pabo CO. DNA recognition by Cys²His² zinc finger proteins. *Annu Rev Biophys Biomol Struct.* 2000;29:183-212.
 54. Englbrecht CC, Schoof H, Böhm S. Conservation, diversification and expansion of C₂H₂ zinc finger proteins in the *Arabidopsis thaliana* genome. *BMC Genomics.* 2004; 5(1):1-17.
 55. Feng BH, Han YC, Xiao YY, Kuang JF, Fan ZQ, Chen JY, Lu WJ. The banana fruit Dof transcription factor MaDof23 acts as a repressor and interacts with MaERF9 in regulating ripening-related genes. *J Exp Bot.* 2016;67(8):2263-75.
 56. Pan IC, Li CW, Su RC, Cheng CP, Lin CS, Chan MT. Ectopic expression of an EAR motif deletion mutant of *SIERF3* enhances tolerance to salt stress and *Ralstonia solanacearum* in tomato. *Planta.* 2010;232(5):1075-86.
 57. Zhang G, Chen M, Chen X, Xu Z, Li L, Guo J, Ma Y. Isolation and characterization of a novel EAR-motif-containing gene GmERF4 from soybean (*Glycine max* L.). *Mol Biol Rep.* 2010;37(2):809-18.
 58. Sugano S, Kaminaka H, Rybka Z, Catala R, Salinas J, Matsui K, Ohme-Takagi M, Takatsuji H. Stress-responsive zinc finger gene *ZPT2-3* plays a role in drought tolerance in petunia. *Plant J.* 2003;36(6):830-41.
 59. Cao H, Xu Y, Yuan L, Bian Y, Wang L, Zhen S, Hu Y, Yan Y. Molecular Characterization of the 14-3-3 Gene Family in *Brachypodium distachyon* L. Reveals High Evolutionary Conservation and Diverse Responses to Abiotic Stresses. *Front Plant Sci.* 2016;7:1099.
 60. Chen H, Han R. Characterization of Actin Filament Dynamics during Mitosis in Wheat

- Protoplasts under UV-B Radiation. *Sci Rep.* 2016;6:20115.
61. Caldeira CF, Bosio M, Parent B, Jeanguenin L, Chaumont F, Tardieu F. A hydraulic model is compatible with rapid changes in leaf elongation under fluctuating evaporative demand and soil water status. *Plant Physiol.* 2014;164(4):1718-30.
62. Zhang H, Wang R, Wang H, Liu B, Xu M, Guan Y, Yang Y, Qin L, Chen E, Li F, Huang R, Zhou Y. Heterogeneous root zone salinity mitigates salt injury to *Sorghum bicolor* (L.) Moench in a split-root system. *PLoS One.* 2019;14(12):e0227020.
63. Wu R, Grissom JE, McKeand SE, O'Malley DM. Phenotypic plasticity of fine root growth increases plant productivity in pine seedlings. *BMC Ecol.* 2004;4:14.
64. Lynch J. Root Architecture and Plant Productivity. *Plant Physiol.* 1995;109(1):7-13.
65. Sun B, Zhao Y, Shi S, Yang M, Xiao K. TaZFP1, a C₂H₂ type-ZFP gene of *T. aestivum*, mediates salt stress tolerance of plants by modulating diverse stress-defensive physiological processes. *Plant Physiol Biochem.* 2019;136:127-142.
66. Sakamoto A, Minami M, Huh GH, Iwabuchi M. The putative zinc-finger protein *WZF1* interacts with a cis-acting element of wheat histone genes. *Eur J Biochem.* 1993; 217(3):1049-56.
67. SUN Zhao-hua, DING Chang-huan, LI Xiao-juan, XIAO Kai .Molecular Characterization and Expression Analysis of TaZFP15, a C₂H₂-Type Zinc Finger Transcription Factor Gene in Wheat (*Triticum aestivum* L.).*Journal of Integrative Agriculture*,2012;11(01):31-42.
68. Chang H, Chen D, Kam J, Richardson T, Drenth J, Guo X, McIntyre CL, Chai S, Rae AL, Xue GP. Abiotic stress upregulated *TaZFP34* represses the expression of type-B response regulator and *SHY2* genes and enhances root to shoot ratio in wheat. *Plant Sci.* 2016;252:88-102.
69. Liu XM, An J, Han HJ, Kim SH, Lim CO, Yun DJ, Chung WS. ZAT11, a zinc finger transcription factor, is a negative regulator of nickel ion tolerance in *Arabidopsis*. *Plant Cell Rep.* 2014;33(12):2015-21.
70. Shah K, Singh M, Rai AC. Effect of heat-shock induced oxidative stress is suppressed in BcZAT12 expressing drought tolerant tomato. *Phytochemistry.* 2013;95:109-17.

800

801

802 **Figure 1** | Subset classification of *C₂H₂-ZF* genes in the *T.aestivum* genome. **(a)**
803 Numbers of *C₂H₂-ZFPs* in the 10 different subsets. **(b)** Numbers of *C₂H₂-ZFPs*
804 containing 1~ 7 and 9 *C₂H₂-ZF* domains.

805

806

807

808

809

810

811

812

Figure 2| Chromosome localization of TaC₂H₂ members. **(a)** The TaC₂H₂ members were localized in bread wheat chromosomes (Chr1A,1B,1D–7A,7B,7D). **(b)** Numbers of TaZFP genes on each *T.aestivum* chromosome.

Figure 3| Phylogenetic tree showing the spread of C₂H₂ domain proteins of *T.aestivum*. The phylogenetic tree was constructed using the neighbor-joining method as implemented in MEGA 7.0 from TaC₂H₂ protein sequences alignment. Bootstrap values from 1000 replicates are displayed at each node. The different C₂H₂ domain proteins on the tree can be divided by different colors.

Figure 4| Gene structure and motifs of *TaC₂H₂s* family. (a) The clustering of *TaC₂H₂* proteins based on Maximum Likelihood phylogenetic tree; (b) The exon-intron structure of *TaC₂H₂* genes, the yellow boxes are the UTR region of *TaC₂H₂* genes, the green boxes are the exon of CDS, and the black line is the intron. (c) Schematic depiction of 10 conserved motifs in *TaC₂H₂* proteins. MEME online tool was used to identify the motifs of the *TaC₂H₂* proteins. Each motif is denoted by different colored blocks with respective numbers in the center of the motifs. The number in boxes (1–15) signifies motif 1– motif 15. The length and position of each colored box indicate the actual size of the motif.

Figure 5| The conserved functional motifs and sequence logos of ten subsets proteins were predicted by MEME software. The highlights show the wheat genome conservation of the C_2H_2 zinc finger motif. Numbers on the x-axis represent the sequence positions in zinc finger motifs. The y-axis represents the information content measured in bits.

Figure 6| The collinearity analysis of the C_2H_2 -ZF gene family in *T.aestivum*. **(a)** Colinearity analysis of *TaZFP* genes. Duplicated gene pairs in the wheat genome. **(b)** colinearity analysis of *TaZFP* genes with *Aegilops tauschii*. **(c)** *Brachypodium distachyon*. **(d)** *Arabidopsis thaliana*. **(e)** *Oryza sativa Japonica* (rice). Gray lines in the background indicate the collinear blocks within the tartary buckwheat and other plant genomes, while the dark lines highlight the syntenic *TaZFP* gene pairs.

Figure 7| Transcriptome analysis of *T.aestivum* C_2H_2 -ZF genes. **(a)** Expression profiles of 198 C_2H_2 -ZFP genes in different *T.aestivum* tissues. Hierarchical clustering of the relative transcript abundance profiles (log2 tpm) of genes. Different colors degrees represent different expression level. The scale bar indicates relative expression level as shown on the top right side. The different tissue types are shown on the bottom side: seedling roots (SR), roots at the three leaves (RTLS), roots at the meiosis (RMS), stems at the 1 cm spike (S1S), stems at the two nodes stems (STNS), stems at the anthesis stems (SATS), seedling leafs (SL), leafs at the three tillers (LTTS), leafs at the 2 days after anthesis (L2DAAs), spikes at the two nodes stems (SPTNS), spikes at the meiosis (SPMS), spikes at the anthesis stems (SPAS), grains at the 2 days after anthesis (G2DAAs), grains at the 14 days after anthesis (G14DAAs), grains at the 30 days after anthesis (G30DAAs). The individual gene names are indicated on the right side. **(b)** Expression analysis of 25 *TaC2H2-ZFP* genes in different tissues and developmental stages by qRT-PCR, vertical bars indicate standard deviation. **(c)** Expression analysis of 18 *TaC2H2-ZFP* genes in response to drought and heat stress by qRT-PCR, and vertical bars indicate standard deviation.

Figures

Figure 1

Subset classification of C2H2-ZF genes in the *T.aestivum* genome. (a) Numbers of C2H2-ZFPs in the 10 different subsets. (b) Numbers of C2H2-ZFPs containing 1–7 and 9 C2H2-ZF domains.

Figure 2

Chromosome localization of TaC2H2 members. (a) The TaC2H2 members were localized in bread wheat chromosomes (Chr1A,1B,1D–7A,7B,7D). (b) Numbers of TaZFP genes on each T.aestivum chromosome.

Figure 3

Phylogenetic tree showing the spread of C2H2 domain proteins of *T. aestivum*. The phylogenetic tree was constructed using the neighbor-joining method as implemented in MEGA 7.0 from TaC2H2 protein sequences alignment. Bootstrap values from 1000 replicates are displayed at each node. The different C2H2 domain proteins on the tree can be divided by different colors.

Figure 4

Gene structure and motifs of TaC2H2s family. (a) The clustering of TaC2H2 proteins based on Maximum Likelihood phylogenetic tree; (b) The exon-intron structure of TaC2H2 genes, the yellow boxes are the UTR region of TaC2H2 genes, the green boxes are the exon of CDS, and the black line is the intron. (c) Schematic depiction of 10 conserved motifs in TaC2H2 proteins. MEME online tool was used to identify the motifs of the TaC2H2 proteins. Each motif is denoted by different colored blocks with respective

numbers in the center of the motifs. The number in boxes (1–15) signifies motif 1– motif 15. The length and position of each colored box indicate the actual size of the motif.

Figure 5

The conserved functional motifs and sequence logos of ten subsets proteins were predicted by MEME software. The highlights show the wheat genome conservation of the C2H2 zinc finger motif. Numbers on the x-axis represent the sequence positions in zinc finger motifs. The y-axis represents the information content measured in bits.

Figure 6

The collinearity analysis of the C2H2-ZF gene family in *T.aestivum*. (a) Collinearity analysis of TaZFP genes. Duplicated gene pairs in the wheat genome. (b) collinearity analysis of TaZFP genes with *Aegilops tauschii*. (c) *Brachypodium distachyon*. (d) *Arabidopsis thaliana*. (e) *Oryza sativa Japonica* (rice). Gray lines in the background indicate the collinear blocks within the tartary buckwheat and other plant genomes, while the dark lines highlight the syntenic TaZFP gene pairs.

Figure 7

Transcriptome analysis of *T.aestivum* C2H2-ZF 862 genes. (a) Expression profiles of 198 C2H2-ZFP genes in different *T.aestivum* tissues. Hierarchical clustering of the relative transcript abundance profiles (log2 tpm) of genes. Different colors degrees represent different expression level. The scale bar indicates relative expression level as shown on the top right side. The different tissue types are shown on the bottom side—seedling roots (SR)—roots at the three leaves (RTLS)—roots at the meiosis (RMS), stems at the 1 cm spike (S1S), stems at the two nodes stems (STNS), stems at the anthesis stems (SATS), seedling leaves (SL), leaves at the three tillers (LTTS), leaves at the 2 days after anthesis (L2DAAs), spikes at the two nodes stems (SPTNS), spikes at the meiosis (SPMS), spikes at the anthesis stems (SPAS), grains at the 2 days after anthesis (G2DAAs), grains at the 14 days after anthesis (G14DAAs), grains at the 30 days after anthesis (G30DAAs). The individual gene names are indicated on the right side. (b) Expression analysis

of 25 TaC2H2-ZFP genes in different tissues and developmental stages by qRT-PCR, vertical bars indicate standard deviation. (c) Expression analysis of 18 TaC2H2-ZFP genes in response to drought and heat stress by qRT-PCR, and vertical bars indicate standard deviation.

Supplementary Files

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

- [SupplementalTableS1.xls](#)
- [SupplementalTableS2.xls](#)
- [SupplementaryTableS3.xlsx](#)
- [SupplementaryTableS4.xlsx](#)