Phyton-
International Journal of
Experimental Botany

Tech Science Press

DOI: 10.32604/phyton.2021.012726

ARTICLE

Genome-wide Analysis of a Plant AT-rich Sequence and Zinc-binding Protein (PLATZ) in *Triticum Aestivum*

Xiaohang He[1], Minjie Liu[2], Zhengwu Fang[1], Dongfang Ma[1,2,3,*], Yilin Zhou[1,3,*] and Junliang Yin[1]

[1]Engineering Research Center of Ecology and Agricultural Use of Wetland, Ministry of Education, Hubei Collaborative Innovation Center for Grain Industry, College of Agriculture, Yangtze University, Jingzhou, 434025, China

[2]Shanxi Key Laboratory of Integrated Pest Management in Agriculture, Institute of Plant Protection, Taiyuan, 030031, China

[3]State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, 100193, China

*Corresponding Authors: Dongfang Ma. Email: madf@yangtzeu.edu.cn; Yilin Zhou. Email: ylzhou@ippcaas.cn

Received: 10 July 2020 Accepted: 29 December 2020

ABSTRACT

Plant AT-rich sequence and zinc-binding protein (PLATZ) is a plant transcription factor that has been studied in corn. *PLATZ* can non-specifically bind to sequences rich in A/T bases to induce transcriptional repression. It is involved in the regulation of dehydration tolerance in seeds. In this study, we performed bioinformatics analysis to identify and characterize wheat *PLATZ*(*TaPLATZ*)genes. We identified 49 wheat *PLATZ* genes by searching the wheat genome by using known *PLATZ* gene sequences from rice, *Arabidopsis*, and maize. Phylogenetic analysis on *PLATZ* gene sequences from different species was performed. We found that *PLATZs* could be divided into three groups. The chromosome (chr) distribution analysis revealed that the 49 identified wheat *PLATZ* genes are distributed in 15 chrs. Gene structure and motif analyses indicated that most *PLATZ* genes possess conserved exon/intron arrangements and motif compositions. Our analysis of transcriptional data indicated that several wheat *PLATZ* genes may play an important role in abiotic stress resistance given that they are expressed under salt stress. The results of qRT-PCR further confirmed that *TaPLATZ* is involved in plant abiotic stress and is also related to the cell differentiation of plant tissues. Our results lay the foundation for further studies on the function of the wheat *PLATZ* gene family.

KEYWORDS

Transcription factor; *PLATZ*; abiotic stress; gene expression

1 Introduction

Several families of transcription factors (TFs) exist. TFs specifically bind to cis-acting elements in the promoter region to regulate the expression of multiple downstream genes and act as key regulators of transcriptional expression in biological processes [1,2]. A total of 84 putative TF families and other transcriptional regulators (TRs) have been identified from 19 plant species with genomes that have been completely sequenced and annotated [3]. TFs are proteins that bind to cis-elements in their target promoters in a sequence-specific manner, whereas TRs exert their regulatory function through protein–protein interactions or chromatin remodeling. TFs play an important role in plant-specific processes,

including secondary metabolism and response to phytohormones, as well as in the development of the specific characteristics of different cell types. In *Arabidopsis thaliana*, TF gene families are interspersed throughout the genome, and genes related to TFs account for approximately 50% of the total number of major TF families [4]. Additionally, 45% of TFs in corn are plant specific [5]. The *MYB*, *bHLH*, and *bZIP* gene families are large TF families present in different plant species [6–8]. TF-IIIA was first discovered by Miller et al. in *Xenopus laevis*. It has since been identified in plants, such as *Arabidopsis*, petunia, soybean, rice, cotton, and wheat [9,10]. Zinc finger proteins are well-researched nucleic acid-binding proteins that are widely distributed in diverse eukaryote species where they play important regulatory roles [11]. They can be classified as C_2H_2, C_4, C_6, C_4HC_3, C_3HC_4, C_2HC, C_3H, and combination types in accordance with the numbers and positions of their cysteine and histidine residues [12]. Zinc finger proteins are involved in cell differentiation, proliferation, apoptosis, and other important life processes. Moreover, some zinc finger proteins can regulate plant-stress tolerance [13]. Zinc finger proteins can play numerous roles in various plant tissues at different developmental stages under diverse abiotic stresses [10].

The zinc finger protein *TaZNF* can drastically improve the ability of transgenic plants to excrete Na^+ [13]. In *Arabidopsis*, the stress-associated zinc finger protein gene *AtSAP5* is mainly expressed in the roots and is induced by high salt, drought, and low temperatures. The overexpression of *AtSAP5* in *Arabidopsis* promotes the expression of other drought stress-related genes and enhances the drought tolerance of transgenic plants under normal growth conditions or drought stress [14]. The rice zinc finger protein genes *OsZFP* and *OsZF19* may participate in drought stress given that their expression is strongly induced under drought [15,16]. Jain *et al*. found that *Arabidopsis* transfected with the rice zinc finger protein gene *OsTOP6A1* exhibited increased salt tolerance and enhanced expression levels of numerous stress response genes after exposure to stress [17]. The wheat zinc finger protein gene *TaCHP* is mainly expressed in the roots of seedlings in the three-leaf stage. The overexpression of *TaCHP* can enhance the salt tolerance of wheat and *Arabidopsis* and increase the expression of various stress genes, such as *AtCBF3*, *AtDREB2 A*, *AtABI2*, and *AtABI1*, under salt stress [18]. *TaCHP* overexpression increases NaCl tolerance in the presence of exogenous 16-carbon DAG and 16-carbon PA. The K ion content and K^+/Na^+ ratio of *TaDSU*-overexpressing *Arabidopsis* considerably increased under stress conditions. The absolute salt ion content of plants affects salt tolerance. In contrast to salt-sensitive plants, salt-tolerant plants can reduce the toxic effects of ions by limiting the variation in ionic content under salt stress [19].

PLATZs represent a new class of plant-specific, zinc-dependent DNA-binding TFs that contain a conserved *PLATZ* domain [20]. They are novel zinc finger proteins that individually possesses a finger-like domain with two zinc finger domains. *PLATZ1* is the first reported *PLATZ* and was isolated from pea. *PLATZ* binds nonspecifically to A/T-rich sequences and affects transcription inhibition [9]. *Arabidopsis* has 12 *PLATZ* members. In *Arabidopsis*, *PLATZ5* is expressed in various tissues under salt stress and localizes in the nucleus and cytoplasm. The transcriptional activity of *PLATZ5* in *Arabidopsis* may be similar to that of *PLATZ1* in peas [21]. The maize *PLATZ* protein is involved in the regulation of the RNA PIII-mediated transcription of small noncoding RNAs in different tissues, including endosperm in maize [20,22]. The expression of Chinese narcissus *NtPLATZ1* gene could be induced by the treatment of paclobutrazol. The results of semi-quantitative interpretation RT-PCR showed that the expression of *NtPLATZ1* gene could be detected in leaves of Chinese narcissus from long leaf stage to blooming [23]. Transgenic tobacco growth is inhibited by *NtPLATZ1* overexpression and promoted by the inhibition of *NtPLATZ1*expression [24].

Wheat is an important food crop in the world, accounting for more than half of the total human consumption [25,26]. It has wide adaptability and highly resistant to storage degradation and damage. It is one of the most important food crops in China. Thus, the development of the wheat industry in China is directly related to national food security and social stability [27]. Wheat is affected by various

organisms (such as wheat stripe rust) and abiotic stresses such as drought, salinity and high temperature during its growth and development [28–30]. Therefore, improving the stress resistance of wheat is crucial for ensuring high and stable wheat yields. Zinc finger TFs play an important role in plant stress signal transduction and abiotic stress responses. Although the function of several TFs has been understood, that of *PLATZ* remains unclear. The completion of the genome-wide sequencing of wheat has enabled the size identification and functional analysis of the wheat *PLATZ* gene. In the present study, we systematically identified and structurally characterized the *PLATZ* gene from the wheat reference genome. We also described the expression pattern exhibited by the *PLATZ* gene under salt stress.

2 Materials and Methods

2.1 Identification of PLATZ in the Wheat Genome

A computer-based method was used to identify members of the *PLATZ* family from the wheat (IWGSC v1.0, https://wheat-urgi.versailles.inra.fr/Seq-Repository/Annotations). We first collected information on previously reported *PLATZ* family members from model plants, including *Arabidopsis* (TAIR database1, http://www.arabidopsis.org/index.jsp), rice (http://rice.plantbiology.msu.edu/index.shtml) [31], and maize (https://www.maizegdb.org/) [1,32]. We queried the wheat genome [33] by using BLASTp ($E < 10^{-5}$) and known *PLATZ* sequences from model plants as seed sequences. We used a profiled Hidden Markov Model (HMM) implemented with default parameters in HMMER 3.0 to search for TaPLATZ proteins with the PLATZ domain in the Pfam (v32.0, http://pfam.xfam.org/) database. Finally, we determined by Pfam whether an obtained sequence contained a PLATZ-specific structural con-served domain and finally determined the number of PLATZ gene family members.

2.2 Phylogenetic Analysis of Wheat TaPLATZs

Phylogenetic relationships were inferred through the Neighbor-Joining (NJ) method with 1000 replicated-bootstraps in MEGA7.0 software [34]. A midpoint rooted base tree was produced using the Interactive Tree of Life (iTOL, version 3.2.317, http://itol.embl.de). The scale bar used corresponded to 0.1 amino acid substitutions.

2.3 Feature Analysis of TaPLATZ Proteins

PLATZ protein sequences were characterized by using the protein identification and analysis tools available on the ExPASy Server10 (https://prosite.expasy.org/). The characteristics of protein length, molecular weight (MW), isoelectric point (pI), stability, and grand average of hydropathicity (GRAVY) were predicted [35].

2.4 Protein Motif and Gene Structure Analysis of TaPLATZs

The protein sequences of 49 *PLATZs* were scanned for conserved motifs by using the MEME suite analysis tool version 4.9.1 (http://meme-suite.org/index.html) [36] and SMART motif search tool. The known PLATZs protein sequence was used as control sequences. Then the control sequences were applied to identify the conserved TaPLATZ motifs by the following criteria: each sequence may contain any number of non-overlapping occurrences of each motif, the number of different motifs as 20, and a range of motif widths from 6 to 50 aa. These motif patterns were drawn by TBtools software. According to GFF3 of TaPLATZs, using GSDS 2.0 (http://gsds.cbi.pku.edu.cn/index.php) to draw the genetic structures [37].

2.5 Chromosomal Localization of Gene Sequences

The wheat genome GFF3 gene annotation file was obtained from the wheat database IWGSC v1.0 and gene annotation of wheat TaPLATZs were extracted from the GFF3 file. MapInspect software was used to draw a physical map on the basis of the start and end location information of the PLATZs in corresponding chromosomes (chrs) [37].

2.6 Multiconditional Transcriptome Analysis of TaPLATZs

Original multiple RNA-seq data original from multiple conditional transcriptome analyses were download from National Center for Biotechnology Information (NCBI) Short Read Archive (SRA) database and mapped to wheat reference genome by hisat2. Then genes were assembled by Cufflinks to inspect the expression levels of TaPLATZs (normalized by TPM, transcripts per kilobase of exon model per million mapped reads). The R package "pheatmap" was used to generate the heatmap of TaPLATZs.

2.7 Growth and Stress Treatment of Wheat Seedlings

This study used hexaploid common wheat "Yangmai 20" as the experimental material. The selected standard seeds were surface sterilized with 1% hydrogen peroxide, rinsed thoroughly with distilled water, and germinate with water saturation at 25°C for 2 days in Petri dishes on three-layer of filter paper. The seedlings were then transferred and cultured in a quarter-strength Hoagland nutrient solution with continuous ventilation and increased to half-strength after 3 days [38]. When the wheat grows to one leaf and one heart, the plants are subsequently treated with 150 mM sodium chloride (NaCl) and 20% polyethylene glycol (PEG). PEG is used to simulate drought treatment. Use 1 M KOH or 0.2 M H_2SO_4 to adjust the pH of the nutrient solution to 6.0 every 2 days. Conditions during treatment are 25/20°C temperature and 16 h/8 h (day/night) photoperiod. Three biological replicates were set for each treatment. Collect the leaves and roots of wheat seedlings at 2, 6, 12, 24, 48 and 72 h after treatment. For IAA (indole-3-acetic acid) treatment, wheat seedlings were cultured in a 1/2 concentration Hoaglan nutrient solution containing 100 μM/L IAA. Collect the leaves and roots of wheat seedlings at 2, 6, 12 and 24 h after treatment. For 6-BA (6-benzylaminopurine) treatment, wheat seedlings were cultured in a 1/2 concentration Hoaglan nutrient solution containing 5 mg/L 6-BA. Collect the leaves and roots of wheat seedlings at 2, 6, 12, 24, 48 and 72 h after treatment. Each treatment included three biological replications. After sampling, the samples were quickly frozen in liquid nitrogen and stored at –80°C [39].

2.8 Quantitative Real-Time PCR and Data Analysis

In order to further clarify the functional specificity of *PLATZ*, quantitative real-time polymerase chain reaction (qRT-PCR) was performed to detect the expression level of *TaPLATZ* genes. TRizol reagent (Invi-trogen, USA) is used for the extraction of total RNA from the sample, and then the 5X all-in-one reverse transcription master mix kit (perfect real-time) is used to reverse transcribe the total RNA into cDNAs for quantitative polymerase chain reaction analysis. Primer 5.0 software was used to design gene-specific primers. The primers were used for PCR, the amplified products were detected by gel electrophoresis, and the desired bands were selected. For qRT-PCR experiments, the reaction system contains 10 μL of 2 × SYBR Green Mix, 0.4 μL of each primer (10 μM), 2 μL of template (about 400 ng/μL) and 7.2 μL of ddH2O to make the total volume reach 20 μL. The protocol was carried out as follows: Pre-denaturation at 95°C for 3 min (Step 1), denaturation at 95°C for 5 s (Step 2), primer annealing/extension and collection of fluorescence signal at 58°C for 30 s (Step 3). The next 40 cycles started at Step 2. Three technical replicates were used for each sample. ADP-ribosylation factor Ta2291 (forward: GCTCTCCAACAACATTGCCAAC, reverse: GCTTCTGCCTGTCACATACGC) was stably expressed under various conditions and was used as an internal reference gene for qRT-PCR analysis. Relative expression level was calculated by $2^{-\Delta\Delta Ct}$ method [39,40].

3 Results

3.1 Identification of PLATZ in the Wheat Genome

We acquired 48 *PLATZ* gene sequences from model plants to identify members of the *PLATZ* gene family in wheat. Specifically, we collected 12, 21, and 15 *PLATZ* sequences from *Arabidopsis*, maize, and rice, respectively [24]. We used these sequences and BLASTp (standard $E < 10^{-5}$) to search the wheat genome database. We screened 51 candidate wheat amino acid sequences of *TaPLATZs* to ensure

the reliability of the protein sequences. The obtained sequences were further screened by using Pfam. Finally, we obtained 49 wheat PLATZ (TaPLATZ) amino acid sequences.

3.2 Phylogenetic Analysis of PLATZ

To better understand the evolutionary history and evolutionary relationships of the *PLATZ* gene family in wheat, a phylogenetic tree was generated using the Neighbor-Joining (NJ) method with the full-length amino acid sequences (Fig. 1). The 97 identified *PLATZ* genes were divided into Subfamilies I, II and III. Subfamilies I, II and III contained 15, 13 and 21*TaPLATZs*, respectively. Moreover, we also found that the dicot *PLATZ*s (*Arabidopsis*) have more closely phylogenetic relationship related to monocot *PLATZ*s (wheat, rice and maize) in each clade with all plants. Members of the same family often have similar functions, so grouping in phylogenetic trees may reflect differences in gene function during evolution [41].

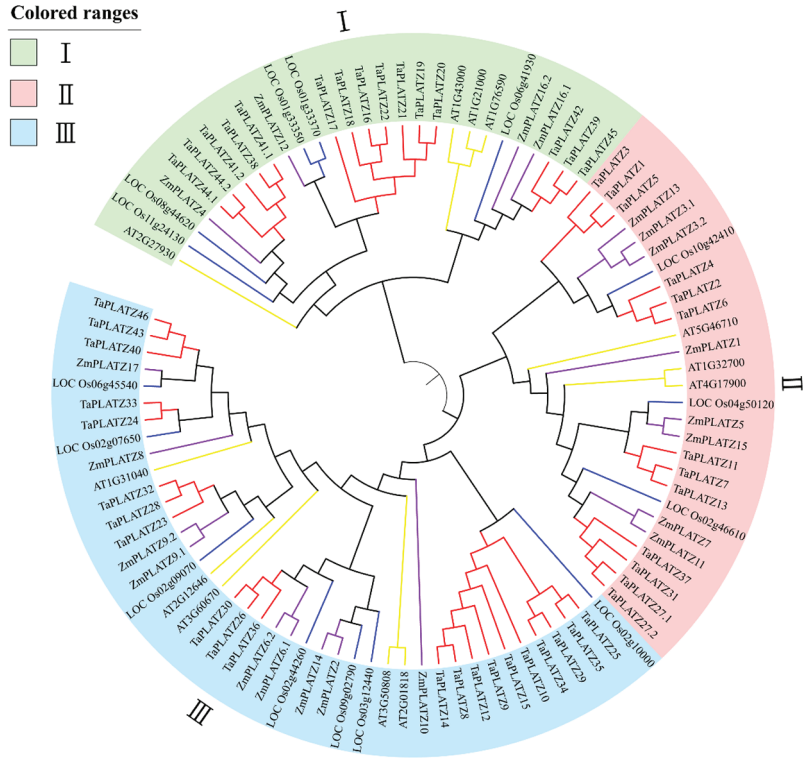

Figure 1: Phylogenetic analysis of the predicted amino acid sequences of *TaPLATZs* and *PLATZ* genes of other plant species. Predicted amino acid sequences were aligned by using the ClustalW2 sequence alignment program, and the phylogenetic tree was constructed with MEGA7.0 software through the bootstrap NJ tree method (1,000 replicates). Different colors indicate different groups, and PLATZs from wheat, rice, maize, and *Arabidopsis* are distinguished by lines with different colors: red-wheat, purple-corn, yellow-*Arabidopsis* and blue-rice

3.3 Protein Features and Domain Analysis of TaPLATZs

We used Pfam to confirm the existence of conserved sequence domains in PLATZ sequences. Analysis revealed that PLATZ proteins have an average isoelectric point of 8.19 (5.21–9.68). Protein feature analysis showed that the *TaPLATZs* have an average length of 229.63 aa (158–275 aa), an average molecular weight of 25.38 kD (17.98–29.76 kD) (Tab. 1), and average instability index of 55.58. Only *TaPLATZ18* was identified as a stable protein.

Table 1: Predicted sequence features of PLATZ proteins

New name	ID	Groups	[a]Length (aa)	[b]MW (kD)	[c]PI	[d]Instability index	[e]Gravy
TaPLATZ1	TraesCS1A01G068200.1	II	247	27.23	6.46	69.07	−0.68
TaPLATZ2	TraesCS1A01G192800.1	II	270	28.97	9.03	61.04	−0.489
TaPLATZ3	TraesCS1B01G086700.1	II	257	28.68	5.21	71.9	−0.7
TaPLATZ4	TraesCS1B01G207700.1	II	257	27.68	8.81	60.26	−0.435
TaPLATZ5	TraesCS1D01G069600.1	II	250	27.7	5.75	67.29	−0.668
TaPLATZ6	TraesCS1D01G196400.1	II	266	28.53	8.99	60.68	−0.46
TaPLATZ7	TraesCS2A01G448300.1	II	246	27.1	9.26	51.66	−0.425
TaPLATZ8	TraesCS2A01G496300.1	III	217	24.05	5.97	68.8	−0.23
TaPLATZ9	TraesCS2A01G496400.1	III	161	18.41	6.2	65.18	−0.444
TaPLATZ10	TraesCS2A01G496500.1	III	158	18.06	6.49	63.46	−0.384
TaPLATZ11	TraesCS2B01G469000.1	II	246	27.13	9.26	52.52	−0.435
TaPLATZ12	TraesCS2B01G524600.1	III	161	17.98	6.98	59.56	−0.379
TaPLATZ13	TraesCS2D01G447400.1	II	246	27.07	9.26	51.78	−0.415
TaPLATZ14	TraesCS2D01G496600.1	III	219	24.17	5.97	66.87	−0.217
TaPLATZ15	TraesCS2D01G496700.1	III	195	21.67	6.49	60.99	−0.316
TaPLATZ16	TraesCS3A01G497600.1	I	215	24.23	7.94	42.73	−0.467
TaPLATZ17	TraesCS3A01G497900.1	I	229	25.89	8.49	40.96	−0.448
TaPLATZ18	TraesCS3B01G560400.1	I	215	24.27	7.52	39.75	−0.414
TaPLATZ19	TraesCS3B01G560600.1	I	216	24.39	7	43.37	−0.462
TaPLATZ20	TraesCS3B01G561000.1	I	215	24.34	7	42.94	−0.503
TaPLATZ21	TraesCS3D01G505000.1	I	215	24.31	6.95	40.07	−0.496
TaPLATZ22	TraesCS3D01G505200.1	I	215	24.31	7.94	45.59	−0.463
TaPLATZ23	TraesCS6A01G145200.1	III	237	27.23	8.67	60.37	−0.542
TaPLATZ24	TraesCS6A01G156600.1	III	243	26.34	8.94	52.26	−0.345
TaPLATZ25	TraesCS6A01G168400.1	III	215	24.12	6.17	45.77	−0.313
TaPLATZ26	TraesCS6A01G248200.1	III	203	22.17	8.74	67.13	−0.473
TaPLATZ27.1	TraesCS6A01G264400.1	II	245	26.69	9.35	52.05	−0.518
TaPLATZ27.2	TraesCS6A01G264400.2	II	275	29.76	9.25	50.26	−0.441
TaPLATZ28	TraesCS6B01G173400.1	III	237	27.26	8.67	58.47	−0.541
TaPLATZ29	TraesCS6B01G196100.1	III	211	23.49	8.26	44.7	−0.399
TaPLATZ30	TraesCS6B01G276500.1	III	198	21.58	9.03	65.51	−0.432
TaPLATZ31	TraesCS6B01G291400.1	II	248	27	9.48	50.96	−0.516
TaPLATZ32	TraesCS6D01G134200.1	III	237	27.23	8.67	58.75	−0.551
TaPLATZ33	TraesCS6D01G146600.1	III	243	26.55	8.94	53.02	−0.366
TaPLATZ34	TraesCS6D01G157400.1	III	211	23.54	8.49	47.34	−0.479
TaPLATZ35	TraesCS6D01G157500.1	III	219	24.43	5.49	51.38	−0.444
TaPLATZ36	TraesCS6D01G230300.1	III	203	22.09	9.02	65.35	−0.483
TaPLATZ37	TraesCS6D01G250500.1	II	249	26.88	9.6	53.51	−0.47
TaPLATZ38	TraesCS7A01G281000.1	I	214	22.78	9.24	47.48	−0.356

(Continued)

Table 1 (continued).

New name	ID	Groups	[a]Length (aa)	[b]MW (kD)	[c]PI	[d]Instability index	[e]Gravy
TaPLATZ39	TraesCS7A01G394800.1	I	242	26.85	9.36	57.17	−0.45
TaPLATZ40	TraesCS7A01G479100.1	III	255	28.08	8.69	63.59	−0.256
TaPLATZ41.1	TraesCS7B01G180100.1	I	267	29.13	9.68	56.67	−0.266
TaPLATZ41.2	TraesCS7B01G180100.2	I	210	22.38	9.14	48.48	−0.37
TaPLATZ42	TraesCS7B01G296800.1	I	242	27.01	9.36	60.65	−0.467
TaPLATZ43	TraesCS7B01G381500.1	III	256	28.17	8.79	66.26	−0.286
TaPLATZ44.1	TraesCS7D01G279600.1	I	266	28.85	9.68	52.12	−0.289
TaPLATZ44.2	TraesCS7D01G279600.2	I	214	22.78	9.14	46.62	−0.353
TaPLATZ45	TraesCS7D01G390300.1	I	242	26.93	9.36	57.52	−0.462
TaPLATZ46	TraesCS7D01G466400.1	III	254	28.09	8.9	63.67	−0.28

Note: [a]Length (Amino acid length); [b]MW (Molecular weight, KD); [c]PI (Isoelectric point); [d]Ins. (Instability index); [e]GRAVY (Grand average of hydropathy).

3.4 Analyses of TaPLATZ Motifs and Gene Structures

We analyzed the structure of wheat *PLATZ* genes by using the annotated gff3 file. We found that wheat *PLATZ* genes contain 1–4 introns (Fig. 2). These introns all contain exons. Moreover, the genes have different structures. Seven gene contains 1 introns, 20 genes contain 2 introns, 20 genes contain 3 introns, and 2 genes contain 4 introns. These data indicate that there are intron acquisition and deletion in *PLATZ* gene. *TaPLATZ1*, *TaPLATZ2*, *TaPLATZ3*, *TaPLATZ5*, *TaPLATZ6*, *TaPLATZ41.1*, and *TaPLATZ44.1* contain only one intron. These genes may have diverged early in evolution and lost introns in the process. We applied MEME software InterPro Scan 5 to detect and annotate the structural diversity and predicted function of *TaPLATZs*. We detected 20 conserved motifs in *TaPLATZs*. Among them, motif 1 and motif 4 were identified as *PLATZ* transcription factors, and the others showed unknown. The sequences, positions and markers of the conserved motifs of each sub-family of *TaPLATZ* are shown in Supplementary Tab. S1. Proteins with relatively close evolutionary relationships are similar in motif composition mode, which indicates that the functions of PLATZ proteins in the same branch are similar [38,40]. The results show that all of the identified wheat PLATZ proteins contain a motif 1and motif 5 domain. While motif 19 only exist in Group I, motif 12 and motif 17 only exist in Group II, motif 9, motif 15 and motif 18 only exist in Group III, the existence of these special motifs may cause differences in *PLATZ* gene function to some extent. The phylogenetic analysis results and conserved motifs in wheat PLATZ proteins are depicted in Fig. 2.

3.5 Chromosomal Location and Gene Duplication Events Analysis of TaPLATZ

We applied Mapinspect software to construct chromosomal maps. We found that 49 *TaPLATZ* genes are unevenly distributed on 15 chromosomes, and chromosomes 6A and 6D contain the largest number of *TaPLATZ* genes. Chromosomes containing four *TaPLATZ* genes are 2A, 6B, 7B and 7D; and chromosomes containing two *TaPLATZ* genes are 1A, 2B, 1D, 2B, 3A and 3D. However, there is no gene distribution on chromosomes 4A, 5A, 4B, 5B, 4D and 5D (Fig. 3). We found gene clusters on 9 chromosomes (2A, 3A, 6A, 3B, 7B, 2D, 3D, 6D and 7D). There are 20 tandem repeat genes: *TaPLATZ8/TaPLATZ9/TaPLATZ*, *TaPLATZ16/TaPLATZ17*, *TaPLATZ27.1/TaPLATZ27.2*, *TaPLATZ18/TaPLATZ19/TaPLATZ20*, *TaPLATZ41.1/TaPLATZ41.2*, *TaPLATZ14/TaPLATZ15*, *TaPLATZ21/TaPLATZ22*, *TaPLATZ34/TaPLATZ35*, *TaPLATZ44.1/TaPLATZ44.2*. In addition, we also discovered the polyploidization of chromosomes. Therefore, we speculate that tandem duplication and chromosome polyploidization may be important ways of *TaPLATZ* genes amplification [42].

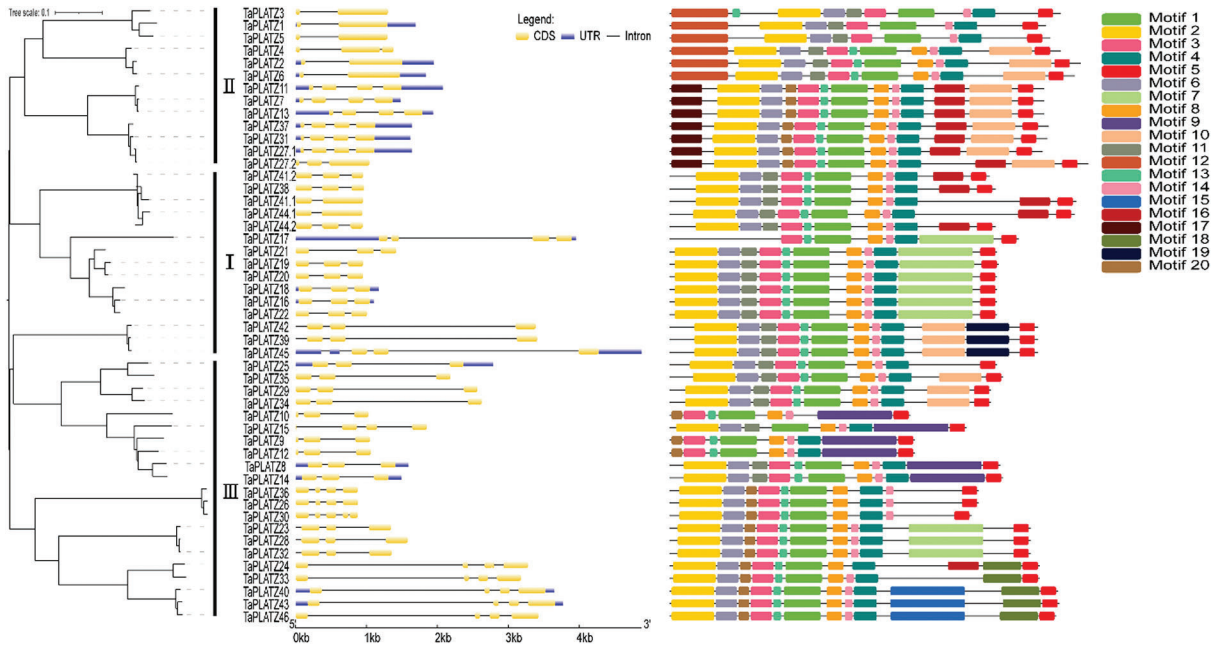

Figure 2: The tree was created with 1,000 bootstraps through the NJ method in MEGA7.0. The composition of motifs in TaPLATZ amino acid sequences was modeled and illustrated by using MAST. Exon-intron structure analyses were conducted using the GSDS. Lengths of exons and introns were displayed proportionally. The untranslated regions (UTRs) were indicated by blue boxes, the exons were indicated by yellow boxes, and the introns were indicated by black lines

3.6 Multiple Conditional Transcriptome Analysis of TaPLATZs

We downloaded original RNA sequence data from multiple conditional transcriptome analyses from NCBI. We then mapped these sequences to the wheat reference genome by applying hisat2. Then, we assembled these data by using cufflinks to inspect the expression levels of *TaPLATZs*. Expression levels were normalized on the basis of the fragments per kilobase of exon model per million mapped reads. We used the R package "pheatmap" to draw the Heat map of wheat *TaPLATZ* genes (Fig. 4). Different genes showed various responses under diverse conditions. For example, wheat *TaPLATZ* genes, *TaPLATZ2*, *TaPLATZ6*, *TaPLATZ27.1*, *TaPLATZ31* and *TaPLATZ37* were significantly up-regulated under Na treatment. In addition, *TaPLATZ7*, *TaPLATZ11*, *TaPLATZ13*, *TaPLATZ39*, *TaPLATZ42* and *TaPLATZ45* were up-regulated under salt treatment. However, more than half of the genes in this study are not expressed under abiotic stress, and these genes may play a role in biotic stress or other growth and development processes [43].

3.7 Quantitative-Real Time PCR Analysis (qRT-PCR)

To further reveal the potential functions of *TaPLATZ* under abiotic stress (NaCl, PEG, IAA and 6-BA), we used qRT-PCR to explore their expression patterns in different tissues of wheat. According to transcriptome analysis, 6 *TaPLATZs* with high expression levels (*TaPLATZ6*, *TaPLATZ7*, *TaPLATZ11*, *TaPLATZ13*, *TaPLATZ27.1*, *TaPLATZ31*) were selected. The results showed that the expression levels of the six genes had no significant difference after IAA application in roots and leaves at different stages (Fig. 5). Under NaCl treatment, compared with the control, 5 *TaPLATZs* (*TaPLATZ6*, *TaPLATZ7*, *TaPLATZ13*, *TaPLATZ27.1*, *TaPLATZ31*) were highly expressed at 48 h. In the root, *TaPLATZ6* at 48 h, *TaPLATZ7* at 12 h, and *TaPLATZ31* at 2 h and 72 h showed up-regulation. However, *TaPLATZ6*,

TaPLATZ7, *TaPLATZ11*, *TaPLATZ13*, and *TaPLATZ27.1* were down-regulated at 72 h. Under PEG treatment, in leaves, only *TaPLATZ7* was highly expressed at 6 h and *TaPLATZ27.1* at 2 h; in roots, all 6 *TaPLATZs* were up-regulated at 12 h. Under the 6-BA treatment, in the leaves, the expression levels of *TaPLATZ6* at 2 h, *TaPLATZ7* at 12 h and *TaPLATZ13* at 48h increased. In addition, the expression levels of *TaPLATZ27.1* and *TaPLATZ31* were low before 72 h. However, in the roots, the expression levels of the 6 *TaPLATZs* were up-regulated at 12 h and 48 h.

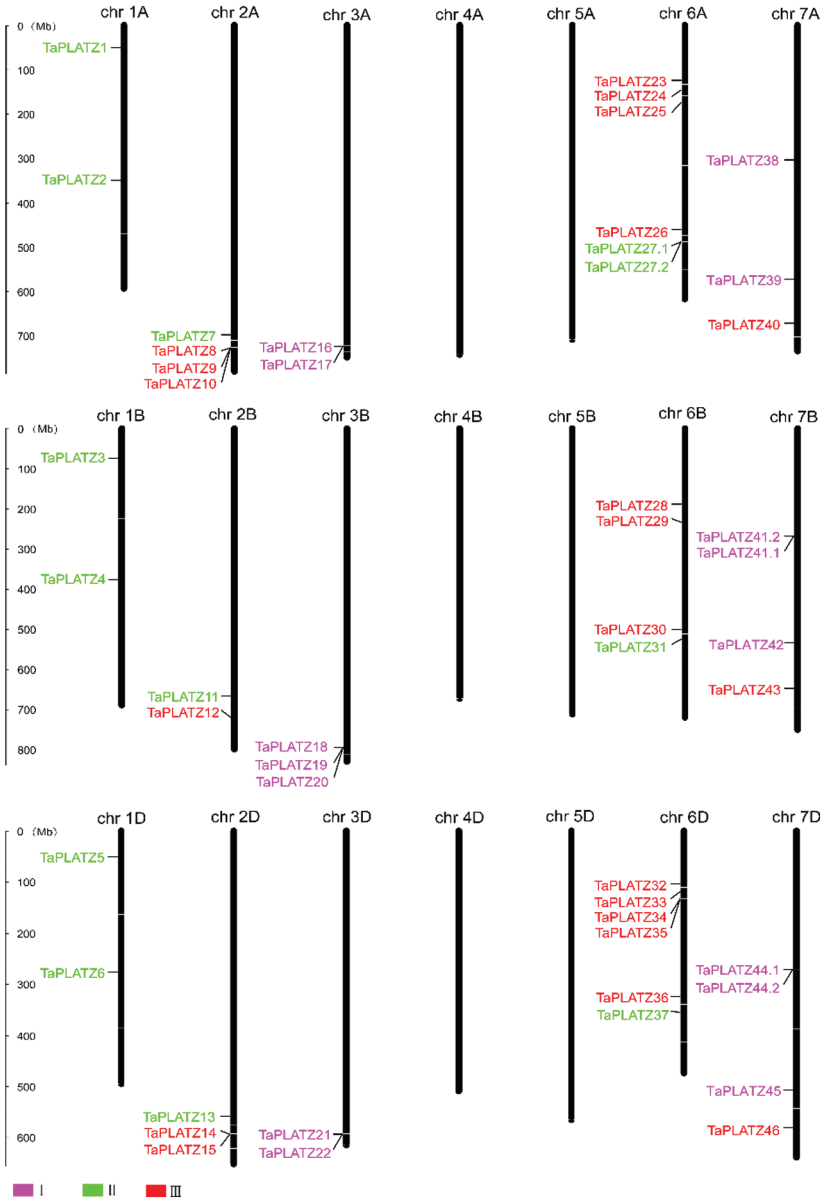

Figure 3: *TaPLATZ* distributions on wheat chromosomes. The chromosome name is indicated at the top of each bar. The rules on the left indicates the physical map distance among genes (Mbp). Rose, green, red colors represent for Groups I, II and III, respectively

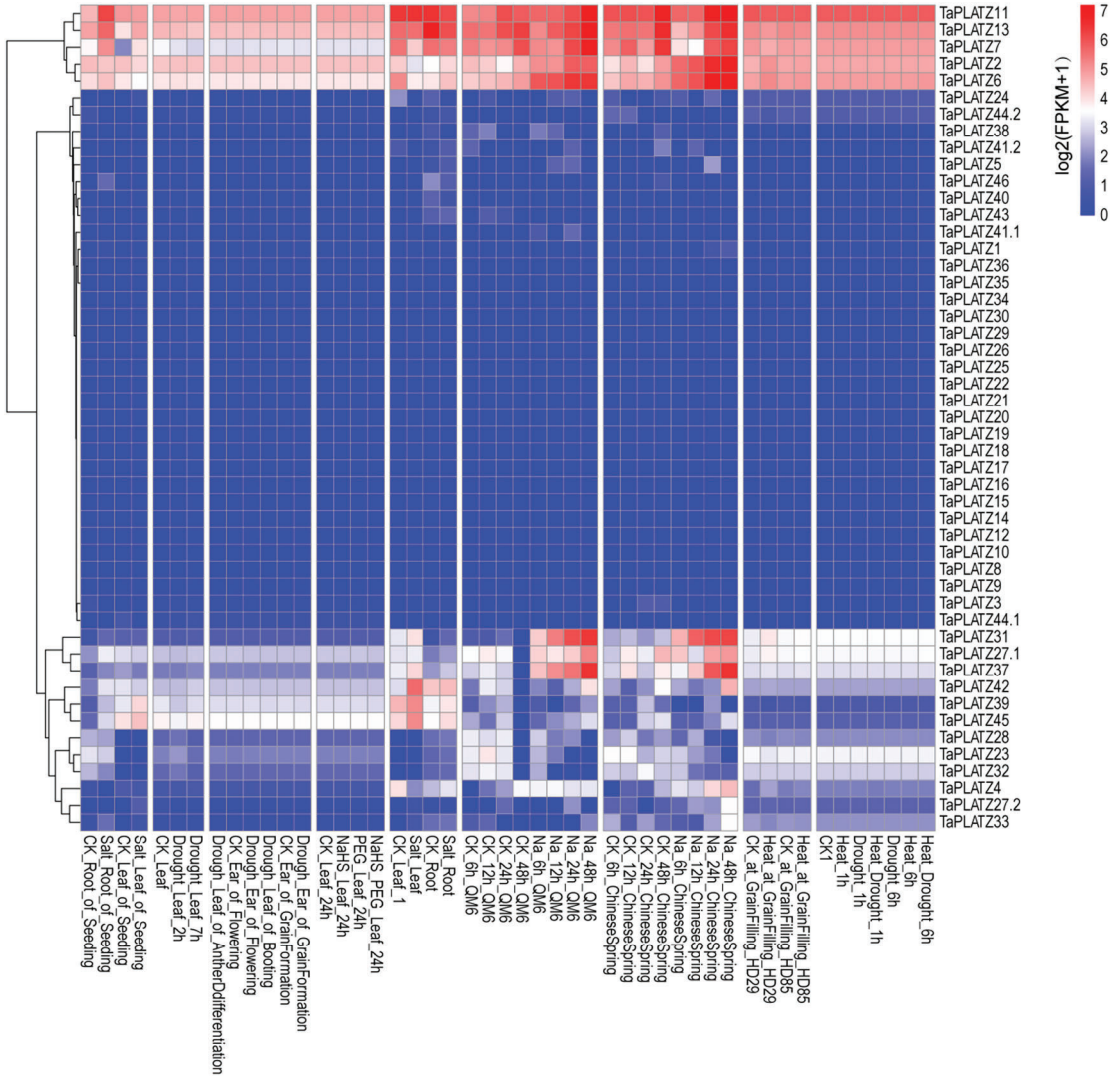

Figure 4: Transcriptome analyses of 49 *TaPLATZs* under different abiotic stress, including salt stress, Na stress, PEG stress, heat stress, drought stress, heat and drought stress (SRA numbers: PRJNA422010, PRJNA293629, PRJNA378325, PRJNA257938, PRJNA391522 and PRJNA257938). Red color indicates increased expression levels; blue color indicates decreased expression levels

4 Discussion

PLATZ protein is a novel class of plant specific zinc dependent DNA binding protein, a transcription factor (TFs), which plays an important role in gene expression and regulation [10,22]. We comprehensively analyzed the *PLATZ* gene family in wheat by identifying the complex functions and characteristics of *PLATZ* genes from model plants. Phylogenetic tree shows that all groups have formed clade. This clustering pattern indicates that these genes are evolutionarily conserved. The phylogenetic tree also shows coevolutionary relationships between species, and the relationships among wheat, rice, and maize are closer than those among wheat, rice, maize, and *Arabidopsis*. Therefore, we speculated that the common ancestors of *PLATZ* may have evolved independently during the evolution of plants from monocots, such as wheat, rice, and maize, to dicots (*Arabidopsis*) [40].

Figure 5: The qPCR analyses of 6 *TaPLATZ* genes under NaCl, PEG, IAA and 6-BA treatments. The qRT-PCR analysis of the *TaPLATZ* genes in the leaves and roots of plants treated with NaCl, PEG, IAA and 6-BA during the specified time period shown on the x-axis, the expression level is on the y-axis. The black square represents the leaf tissue, and the red square represents the root tissue. The standard deviation is shown with error bars. The expression levels of *TaPLATZ* genes were plotted using Origin software

Our research found that all *TaPLATZs* contain a *PLATZ* domain. Through analysis of gene structure and protein motifs, we found that *TaPLATZ* gene has intron deletion and acquisition. Studies have shown that the evolution of introns is affected by genetic mutation and selection, and the loss of introns determines evolution rather than gain of them. Genes contain more introns in the early stages of amplification, and will lose introns after differentiation [44]. Generally speaking, the number of introns in the same subfamily is specific, and the difference in the number of introns between different genes may be caused by insertion and deletion events, and this difference may have a driving effect on evolution. At the same time, this may also be an important reason for the diversity of gene structure and functional complexity. In addition, introns have positive or negative regulatory functions on gene expression. Therefore, the conservative and variant gene structure may affect the function of *TaPLATZs* and its functional diversity [45,46]. The existence of special motifs (For example: motif 9, motif 12, motif 15, motif 17, motif 18, motif 19) may also cause differences in the function of *TaPLATZ* genes.

Gene amplification is one of the most important driving forces in the evolution of the genome. Genes can be amplified in many ways, including genome-wide duplication, tandem duplication, fragment duplication, and so on. Among them, genome-wide duplication and tandem duplication have an important influence on the evolution of the genome and even the evolution of biological species [47]. Gene duplication events are considered to be frequent events in the evolution of organisms, and are also a manifestation of biological evolution. Homologous genes are more adaptable to environmental changes than single-copy genes [48,49]. Through chromosome mapping, we found some tandem duplications and chromosome polyploidization, which may be an important way for *TaPLATZ* genes family amplification. Studies have shown that tandem duplication tends to amplify genes related to biotic and abiotic stress. For example, *Thellungiella parvula* can withstand harsh environments, such as cold, drought, and salinity. This is because there are genes in the *T. parvula* genome that can resist extreme environments, and the cause of this gene is the existence of tandem duplication [50]. Therefore, we speculate that *PLATZ* gene is related to abiotic stress related genes.

Transcriptome analysis showed that *TaPLATZ* gene is related to abiotic stress. The results of qRT-PCR showed that 6 *TaPLATZs* (*TaPLATZ6, TaPLATZ7, TaPLATZ11, TaPLATZ13, TaPLATZ27.1, TaPLATZ31*) were up-regulated at different times under NaCl and PEG stress. Especially after PEG treatment of roots for 12 h, these 6 *TaPLATZs* showed high expression. We speculate that *TaPLATZ* genes participate in the abiotic stress response of plants and plays an important role in coping with environmental stress. Previous studies reported that *PLATZ1* is related to the cell differentiation of plant tissues [9]. It is worth noting that under the 6-BA stress treatment, the 6 *TaPLATZ*s in the leaves were up-regulated after 48 h of the stress treatment. In the roots, the stress treatment increased for 12 h and 24 h respectively, and then showed a downward trend. Therefore, *TaPLATZ* genes may also be involved in the cell differentiation of plant tissues.

PLATZ gene has been reported in *Arabidopsis thaliana* and maize [20,21]. Previously identified *PLATZ* sequences are important for identifying the *PLATZ* sequences of a particular species. Our systematic analysis identified 49 *TaPLATZ* genes in the wheat genome. Our initial result was further supported by phylogenetic, gene structural, and conserved protein motif analyses. By comparing the expression profiles of *TaPLATZ* genes in wheat, we found that some *TaPLATZ* genes may promote plant resistance to salt stress. At the same time, the qRT-PCR verification of some *TaPLATZ* genes showed that these genes respond significantly to abiotic stress and hormones. This also shows that the *TaPLATZ* genes plays an important role in the growth and development of plants. Our data provide new insights into the control of *TaPLATZ* genes expression on the transcriptional level and new clues for the further functional identification of *PLATZ* genes and the genetic improvement of wheat.

Acknowledgement: We are very grateful to Dr. Luo Yong for the language polishing of this manuscript. (Luo Yong, Department of Plant Pathology, China Agricultural University, Beijing, 100193, China).

Author Contributions: Junliang Yin and Dongfang Ma guided the design of the experiment. Junliang Yin and Zhengwu Fang directed the data analysis. He Xiaohang conducted experiments, data analysis, and manuscript writing. Minjie Liu and Yilin Zhou participated in the experiment and contributed to the manuscript writing. Dongfang Ma supervised the experiment and confirmed the manuscript. Xiaohang He is the guarantor of this work, so she can have full access to all the data in the research and is responsible for the integrity of the data and the accuracy of the data analysis. Thank all the above staff for the help of this study. The authors thank the reviewers for their valuable suggestions during the revision of the early manuscripts.

Funding Statement: This work was supported by the "National Key R&D Program of China (2018YFD0200500)", "Open Project Program of Engineering Research Center of Ecology and Agricultural Use of Wetland, Ministry of Education (KF201802)" and "Open Project Program of Shanxi Key Laboratory of Integrated Pest Management in Agriculture, Institute of Plant Protection (YHSW2018002)".

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

References

1. Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K., Yokoyama, S. (2013). DNA-binding domains of plant-specific transcription factors: Structure, function, and evolution. *Trends in Plant Science, 18(5),* 267–276. DOI 10.1016/j.tplants.2012.09.001.
2. Fan, Z. Q., Yan, J. F., Lu, W. J., Chen, J. Y. (2015). Progress in the regulation of fruit ripening and senescence by transcription factors. *Journal of Horticulture, 42,* 1649–1663, DOI 10.16420/j.issn.0513-353x.2015-0356.
3. Pérez-Rodríguez, P., Riaño-Pachón, D. M., Corrêa, L. G., Rensing, S. A., Kersten, B. et al. (2010). PlnTFDB: Updated content and new features of the plant transcription factor database. *Nucleic Acids Research, 38 (suppl_1),* D822–D827. DOI 10.1093/nar/gkp805.
4. Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C. et al. (2000). Arabidopsis transcription factors: Genome-Wide comparative analysis among eukaryotes. *Science, 290(5499),* 2105–2110. DOI 10.1126/science.290.5499.2105.
5. Jiang, Y., Zeng, B., Zhao, H. N., Zhang, M., Xie, S. J. et al. (2012). Genome-wide transcription factor gene prediction and their expressional tissue-specificities in maize. *Journal of Integrative Plant Biology, 54(09),* 616–630. DOI 10.1111/j.1744-7909.2012.01149.x.
6. Nijhawan, A., Jain, M., Tyagi, A. K., Khurana, J. P. (2008). Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. *Plant Physiology, 146(2),* 333–350. DOI 10.1104/pp.107.112821.
7. Carretero-Paulet, L., Galstyan, A., Roig-Villanova, I., Martínez-García, J. F., Bilbao-Castro, J. R. et al. (2010). Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, Poplar, Rice, Moss, and Algae. *Plant Physiology, 153(3),* 1398–1412. DOI 10.1104/pp.110.153593.
8. Katiyar, A., Smita, S., Lenka, S., Rajwanshi, R., Chinnusamy, V. et al. (2012). Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. *BMC Genomics, 13(1),* 544. DOI 10.1186/1471-2164-13-544.
9. Yukio, N., Hirofumi, F., Takehito, I., Yukiko, S. (2001). A novel class of plant-specific zinc-dependent DNA-binding protein that binds to A/T-rich DNA sequences. *Nucleic Acids Research, 29(20),* 4097–4105. DOI 10.1093/nar/29.20.4097.
10. Wang, W. Y., Li, H. M., Dai, Y. M., Lin, J. B. (2016). Progress in the function of plant zinc finger proteins. *Chinese Horticultural Digest, 32(7),* 3–5. DOI CNKI: SUN: YUWZ.0.2016-07-003.
11. Song, B., Hong, Y., Wang, P. W., Wang, H., Fu, Y. P. et al. (2010). Research progress of plant C_2H_2 zinc finger protein. *Genomics and Applied Biology, 29,* 1133–1141. DOI 10.3969/gab.029.001133.
12. Zhang, J., Liu, J. F., Zhao, T. T., Ren, J., Xu, X. Y. (2018). Research progress of plant C_2H_2 zinc finger protein. *Molecular Plant Breeding, 29(6),* 1133–1141. DOI 10.3969/gab.029.001133.

13. Ma, X., Liang, W., Gu, P., Huang, Z. (2016). Salt tolerance function of the novel C_2H_2-type zinc finger protein TaZNF in wheat. *Plant Physiology and Biochemistry, 106,* 129–140. DOI 10.1016/j.plaphy.2016.04.033.
14. Kang, M., Fokar, M., Abdelmageed, H., Allen, R. D. (2011). Arabidopsis SAP5 functions as a positive regulator of stress responses and exhibits E3 ubiquitin ligase activity. *Plant Molecular Biology, 75(4–5),* 451–466. DOI 10.1007/s11103-011-9748-2.
15. Huang, Y. M., Hu, H. H., Wu, C. Q. (2006). Isolation and identification of a stress inducible zinc finger gene in rice. *Journal of Huazhong Agricultural University, 25,* 581–585. DOI 10.3321/j.issn:1000-2421.2006.06.001.
16. Islam, M. S., Hur, J. H., Wang, M. H. (2009). The influence of abiotic stresses on expression of zinc finger protein gene in rice. *Russian Journal of Plant Physiology, 56(5),* 695–701. DOI 10.1134/S1021443709050161.
17. Jain, M., Tyagi, A. K., Khurana, J. P. (2008). Constitutive expression of a meiotic recombination protein gene homolog, OsTOP6A1, from rice confers abiotic stress tolerance in transgenic Arabidopsis plants. *Plant Cell Reports, 27(4),* 767–778. DOI 10.1007/s00299-007-0491-8.
18. Tian, L. M., Huang, C. L., Zhang, X. H., Zhang, L. S., Wu, Z. Y. (2005). Advances in research on zinc finger proteins in stress-related plants. *Biotechnology Bulletin, 6,* 12–16. DOI CNKI: SUN: SWJT.0.2005-06-002.
19. Zhang, H. J. (2014). *Functional study of wheat stress response genes TaCHP and TaDSU*. Shandong University, China. DOI 10.7666/d.Y2594373.
20. Wang, J. C., Ji, C., Li, Q., Zhou, Y., Wu, Y. R. (2018). Genome-wide analysis of the plant-specific PLATZ proteins in maize and identification of their general role in interaction with RNA polymerase III complex. *BMC Plant Biology, 18(1),* D822. DOI 10.1186/s12870-018-1443-x.
21. Yang, R. (2017). *Functional study of Arabidopsis thaliana zinc finger transcription factor PLATZ5 in salt stress response*. Shandong Agricultural University, China.
22. Ye, J., Zheng, X., Xiang, X., Li, C., Fu, M. et al. (2017). The maize imprinted gene Floury3 encodes a PLATZ protein required for tRNAs and 5S rRNA transcription through interaction with RNA polymerase III. *Plant Cell, 29(10),* 2661–2675. DOI 10.1105/tpc.17.00576.
23. Lin, J. B., Wang, W. Y., Li, H. M., Wu, J. S., Zou, H. et al. (2016). Cloning and expression analysis of zinc finger protein NtPLATZ1 from Chinese narcissus. *Journal of Northwest A&F University (Natural Science Edition), 44 (10),* 165–170. DOI 10.13207/j.cnki.jnwafu.2016.10.023.
24. Wang, W. Y., Li, H. M., Dai, Y. M., Lin, J. B. (2017). Construction of expression vectors of NtPLATZ1 positive and antisense plants in Chinese narcissus and transformation of tobacco. *Journal of Horticulture, 44(12),* 2399–2407. DOI CNKI: SUN: YYXB.0.2017-12-017.
25. Yin, J. L., Fang, Z. W., Sun, C., Zhang, P., Zhang, X. et al. (2018). Rapid identification of a stripe rust resistant gene in a space-induced wheat mutant using specific locus amplified fragment (SLAF) sequencing. *Scientific Reports, 8 (1),* 157. DOI 10.1038/s41598-018-21489-5.
26. Sun, C., Zhang, P., Fang, Z., Zhang, X., Yin, J. et al. (2018). Genetic analysis and molecular mapping of stripe rust resistance in an excellent wheat line sanshumai1. *Journal of Plant Pathology, 101,* 235–241. DOI 10.1007/s42161-018-0166-z.
27. Wang, Y. Q., Cheng, L. F. (2011). Restricting factors and countermeasures of wheat industry development. *Barley and Cereal Science, (4),* 80–81. DOI 10.3969/j.issn.1673-6486.2011.04.028.
28. Sun, C., Liu, Y. K., Chao, K. X., Fang, Z. W., Wang, S. P. et al. (2019). Characterization and molecular mapping of stripe rust resistance in wheat-psathyrostachys huashanica introgression line h9015-17-1-9-6. *Canadian Journal of Plant Pathology, 41(1),* 65–75. DOI 10.1080/07060661.2018.1523230.
29. Ma, D. F., Fang, Z. W., Yin, J. L., Chao, K. X., Jing, J. X. et al. (2016). Molecular mapping of stripe rust resistance gene YrHu derived from Psathyrostachys huashanica. *Molecular Breeding, 36(6),* 80. DOI 10.1007/s11032-016-0487-6.
30. Ma, D. M., Li, Q., Tang, M. S., Chao, K. X., Li, J. C. et al. (2015). Mapping of gene conferring adult-plant resistance to stripe rust in Chinese wheat landrace Baidatou. *Molecular Breeding, 35(8),* 385. DOI 10.1007/s11032-015-0244-2.
31. Maser, P., Thomine, S., Schroeder, J. I., Ward, J. M., Hirschi, K. et al. (2001). Phylogenetic relationships within cation transporter families of arabidopsis. *Plant Physiology, 126(4),* 1646–1667. DOI 10.1104/pp.126.4.1646.

32. Andorf, C. M., Cannon, E. K., Portwood, J. L., Gardiner, J. M., Harper, L. C. et al. (2016). MaizeGDB update: New tools, data and interface for the maize model organism database. *Nucleic Acids Research, 44(D1),* D1195–D1201. DOI 10.1093/nar/gkv1007.
33. Clavijo, B. J., Venturini, L., Schudoma, C., Accinelli, G. G., Kaithakottil, G. et al. (2017). An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Research, 27(5),* 885–896. DOI 10.1101/gr.217117.116.
34. Kumar, S., Stecher, G., Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution, 33(7),* 1870–1874. DOI 10.1093/molbev/msw054.
35. Li, R., An, J. P., You, C. X., Shu, J., Wang, X. F. et al. (2018). Identification and expression of the CEP gene family in apple (Malus × domestica). *Journal of Integrative Agriculture, 17(2),* 348–358. DOI 10.1016/S2095-3119(17)61653-8.
36. Zhang, X. W., Yi, D. X., Shao, L. H., Li, C. (2017). *In silico* genome-wide identification, phylogeny and expression analysis of the R2R3-MYB gene family in Medicago truncatula. *Journal of Integrative Agriculture, 16(7),* 1576–1591. DOI 10.1016/S2095-3119(16)61521-6.
37. Jiang, W. Q., Yang, L., He, Y. Q., Zhang, H. T., Li, W. et al. (2019). Genome-wide identification and transcriptional expression analysis of superoxide dismutase (SOD) family in wheat (Triticum aestivum). *PeerJ, 7(2),* e8062. DOI 10.7717/peerj.8062.
38. Jiang, W. Q., Geng, Y. P., Liu, Y. K., Chen, S. H., Cao, S. L. et al. (2020). Genome-wide identification and characterization of SRO gene family in wheat: Molecular evolution and expression profiles during different stresses. *Plant Physiology and Biochemistry, 154(7),* 590–611. DOI 10.1016/j.plaphy.2020.07.006.
39. He, Y. Q., Huang, W. D., Yang, L., Li, Y. T., Lu, C. et al. (2020). Genome-wide analysis of ethylene-insensitive3 (EIN3/EIL) in Triticum aestivum. *Crop Science, 60(4),* 2019–2037. DOI 10.1002/csc2.20115.
40. Livak, K. J., Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods, 25(4),* 402–408. DOI 10.1006/meth.2001.1262.
41. Jiang, W. Q., Yin, J. L., Zhang, H. T., He, Y. Q., Shuai, S. M. et al. (2020). Genome-wide identification, characterization analysis and expression profiling of auxin-responsive GH3 family genes in wheat (triticum aestivum l.). *Molecular Biology Reports, 47(1),* 3885–3907, DOI 10.1007/s11033-020-05477-5..
42. Zhou, X. Y., Zhu, X. G., Shao, W. N., Song, J. H., Jiang, W. Q. et al. (2020). Genome-wide mining of wheat DUF966 gene family provides new insights into salt stress responses. *Frontiers in Plant Science, 11,* 569838. DOI 10.3389/fpls.2020.569838.
43. Fang, Z. W., He, Y. Q., Liu, Y. Y., Jiang, W. Q., Song, J. H. et al. (2020). Bioinformatic identification and analyses of the non-specific lipid transfer proteins in wheat. *Journal of Integrative Agriculture, 19(5),* 1170–1185. DOI 10.1016/S2095-3119(19)62776-0.
44. Knowles, D. G., Aoife, M. L. (2006). High rate of recent intron gain and loss in simultaneously duplicated Arabidopsis genes. *Molecular Biology and Evolution, 23(8),* 1548–1557. DOI 10.1093/molbev/msl017.
45. Zheng, B., Zhan, X. M. (2005). Group III intron and its effect on gene expression. *Biotechnology, 15(4),* 77–79. DOI 10.3969/j.issn.1004-311X.2005.04.035.
46. Cao, J., Shi, F. (2012). Dynamics of arginase gene evolution in metazoans. *Journal of Biomolecular Structure and Dynamics, 30(4),* 407–418. DOI 10.1080/07391102.2012.682207.
47. Fang, L., Cheng, F., Wu, J., Wang, X. W. (2012). Gene preservation of cabbage after whole genome and tandem duplication. *Biotechnology Bulletin, 11,* 9–14. DOI CNKI: SUN: SWJT.0.2012-11-001.
48. Mehan, M. R., Freimer, N. B., Ophoff, R. A. (2004). A genome-wide survey of segmental duplications that mediate common human genetic variation of chromosomal architecture. *Human Genomics, 1(5),* 335–344. DOI 10.1186/1479-7364-1-5-335.
49. Lockton, S., Gaut, B. S. (2005). Plant conserved non-coding sequences and paralogue evolution. *Trends in Genetics, 21(1),* 60–65. DOI 10.1016/j.tig.2004.11.013.
50. Dassanayake, M., Oh, D. H., Haas, J. S., Hernandez, A., Hong, H. et al. (2011). The genome of the extremophile crucifer Thellungiella parvula. *Nature Genetics, 43(9),* 913–918. DOI 10.1038/ng.889.

Supplementary File

Table S1: Identification of conserved motifs in these TaGH3 proteins

Motif	Width (aa)	Logo	Best possible match
1	25	PLATZ transcription factor	VLDISGVQTYVINSARVVFLNERPQ
2	29	Unknown	GGRVPGWLEPLLGTRFFVACAAHPDSHKN
3	15	Unknown	HHVIQIRRSSYHDVV
4	16	PLATZ transcription factor	FRFCSLGCKLKGMVSD
5	11	Unknown	RRKGIPHRAPF
6	15	Unknown	ECNMFCLDCAASACA
7	50	Unknown	MRTSIPTRDIIDYTRKDDDTDCSNTSGNSGNNEESCSDABYCKEKPSPPR
8	11	Unknown	GKGSANNCEVC
9	50	Unknown	PNLTFILDPECKWEYSDSDSTEEEDDGGHLPGPSNSQPIGGTSYGRQPRK
10	29	Unknown	STSSGSSDKSSVVQSFSPSTPPATASSYR
11	11	Unknown	LCYYCRSHHHD
12	38	Unknown	MAIDHDSPFKELLPKBRRIMGGGGPEPDEEEEEAZVAA
13	6	Unknown	RVAELE
14	6	Unknown	RSLLDP
15	50	Unknown	YMSGEPDVACFPRFENLRVGSGSADLLDDGCATGGQITPNSILEDPMHHY
16	21	Unknown	VDGGKKSKSEAEKGASSBSER
17	21	Unknown	MAIDHASPLALKSGGATGGAG
18	26	Unknown	VDVPVPVPRKKKSGGFFPQIVMSLNN
19	29	Unknown	GRFDRGVRWSDDEGSKSNTRPMTPTTPPI
20	8	Unknown	HCLPHHRD