Genome-Wide Analyses of the Genetic Screening of C2H2-Type Zinc Finger Transcription Factors and Abiotic and Biotic Stress Responses in Tomato (Solanum lycopersicum) Based on RNA-Seq Data

Tingting Zhao1†, Tairu Wu1†, Jia Zhang1, Ziyu Wang1, Tong Pei1, Huanhuan Yang1, Jingfu Li1 and Xiangyang Xu1,2*

1Laboratory of Genetic Breeding in Tomato, College of Horticulture and Landscape Architecture, Northeast Agricultural University, Harbin, China, 2Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (Northeast Region), Ministry of Agriculture and Rural Affairs, Northeast Agricultural University, Harbin, China

C2H2-type zinc finger proteins are classic and extensively studied members of the zinc finger family. C2H2-type zinc finger proteins participate in plant growth, development and stress responses. In this study, 99 C2H2-type zinc finger protein genes were identified and classified into four groups, and many functionally related cis-elements were identified. Differential C2H2-ZFP gene expression and specific responses were analyzed under drought, cold, salt, and pathogen stresses based on RNA-Seq data. Thirty-two C2H2 genes were identified in response to multiple stresses. Seven, 3, 5, and 8 genes were specifically expressed under drought, cold, salt, and pathogenic stresses, respectively. Five glycometabolism and sphingolipid-related pathways and the endocytosis pathway were enriched by KEGG analysis. The results of this study represent a foundation for further study of the function of C2H2-type zinc finger proteins and will provide us with genetic resources for stress tolerance breeding.

Keywords: C2H2-type zinc finger gene family, biotic stress, abiotic stress, transcription factor, tomato

INTRODUCTION

Tomato (Solanum lycopersicum) is one of the most important vegetable crops of Solanaceae (Mueller et al., 2005). However, the yield and quality of tomato are greatly affected by various biotic and abiotic stresses such as pathogen infection, low temperature, salt, and drought when the plants are exposed to complex environmental conditions. Upon stress perception, transcription factors (TFs) bind to their target genes to regulate their expression and orchestrate biochemical and physiological modifications critical for stress tolerance and the adaptation of plant growth (Hichri et al., 2014).

C2H2-type zinc finger proteins (C2H2-ZFPs), which are members of an important TF family, are also called TFIIIA-type ZFPs or classical ZFPs and are widely distributed in eukaryotic genomes (Huang et al., 2005). There are 176, 189, 211, and 109 C2H2-ZFPs in Arabidopsis, rice, maize, and poplar, respectively (Englbrecht et al., 2004; Agarwal et al., 2007; Liu et al., 2015;
Materials and Methods

Identification and Phylogenetic Analysis of the C2H2-ZFP Gene Family

Tomato genome sequence data were obtained from the Solanaceae Genomics Network (SGN)\(^1\) (Tomato Genome, 2012). The \(\text{C}_2\text{H}_2\)-ZFPs of eukaryotes generally have a specific conserved sequence consisting of 25–30 amino acids: X-X-C-X(1-5)-C-X(12)-H-X(3-6)-H (X: any amino acid; number: the number of amino acids). The two C (Cys) and two H (His) residues in the sequence form a coordination bond with a zinc ion and then form a tetrahedral structure composed of a two-stranded antiparallel \(\beta\)-sheet and an \(\alpha\)-helix (Carl et al., 2001). EPF1 (later renamed ZPT2-1) was identified in Petunia as the first plant-specific ZFP. Takatsui discovered that EPF1 interacts with the promoter region of the 5-enolpyruvylshikimate-3-phosphate synthase gene (EPSPS) and that the expression of EPF1 parallels the expression of EPSPS (Hiroshi et al., 1994). More zinc finger TFs have been subsequently identified in other plants and have been found to play crucial roles in the regulation of development and responses to biotic and abiotic stresses (Sang et al., 2004; Jiang and Pan, 2012).

Many \(\text{C}_2\text{H}_2\)-ZFPs genes involved in biotic and abiotic stresses have been studied in detail. GmZF1 and GmZF3, two soybean \(\text{C}_2\text{H}_2\)-ZFPs, positively regulate the cold response and negatively regulate the drought response, respectively. Both of these proteins might be involved in the ABA-dependent pathway during the stress response (Zhang et al., 2016a). OsMSR15 contains two \(\text{C}_2\text{H}_2\)-ZFP motifs, and its expression is strongly upregulated by cold, drought, and heat stresses in different rice tissues at different developmental stages (Zhang et al., 2016b). The expression of a novel ZFP gene, StZFP1, cloned from potato, is increased after salt stress as well as after infection by \textit{Phytophthora infestans} and exogenous ABA (Tian et al., 2010). Overexpression of the CAZFPI gene, a zinc-finger protein gene isolated from pepper leaves, enhances resistance against infection by the pathogens \textit{Xanthomonas campestris} and \textit{Colletotrichum cocodes} in transgenic Arabidopsis plants (Sang et al., 2004). In tobacco, ZFT1 functions as a transcription repressor and binds to the EPIS sequence. Overexpression of \textit{ZFT1} renders tobacco plants more tolerant to TMV (Uehara et al., 2005). In this study, \(\text{C}_2\text{H}_2\)-ZFP genes are identified from the tomato genome by a bioinformatic analysis method and are analyzed for many aspects, including their phylogenetic relationships, gene structures, conserved protein motifs, chromosomal locations and promoter \(\text{cis}\)-elements. The abiotic and biotic stress response genes of the \(\text{C}_2\text{H}_2\)-ZFP family are screened under drought, salt, cold, and pathogen infection stresses based on corresponding RNA-Seq data. The results of the present work will provide us with comprehensive genome-wide knowledge of the tomato \(\text{C}_2\text{H}_2\)-ZFP TF family and will also provide us with potential gene resources for biotic and abiotic stress tolerance breeding.

\(^{1}\)\url{http://solgenomics.net}
Plant Materials and Stress Treatments
Tomato Micro-Tom and CGN18423 (which carries the Cf-19 gene that confers resistance to Cladosporium fulvum) were from the Tomato Research Institute of Northeast Agricultural University and were grown in a growth chamber (Conviron, Canada) with a light intensity of 120 µM photons m⁻² s⁻¹ (photoperiod 16 h, day/night temperature 22/18°C). Next, 100 healthy Micro-Tom seedlings were selected for subsequent analyses (10 seedlings used for organ analysis and 90 for drought, salt, cold stress treatments). The tomato seedlings used for drought, salt, and cold stress treatments were transferred to hydroponics and grown for 48 h when they were in the raising period of the four-leaf stage. Then, 30 seedlings were treated with 15% PEG 6000 to simulate drought stress. Young leaves were collected and frozen in liquid nitrogen at 0, 3 and 6 h after drought treatment. Thirty seedlings were transferred to a 5°C growth chamber for cold treatment. Young leaves were collected and frozen in liquid nitrogen at different time points (0, 4, and 12 h) after treatment. Additionally, 30 tomato seedlings were exposed to 200 mM sodium chloride (NaCl). Young leaves were collected and frozen in liquid nitrogen at different time points (0, 2, and 8 h) after treatment. Three biological replicates were performed for each time point. The materials and methods for pathogen stress are shown in Zhao et al. (2019).

RNA Isolation and cDNA Synthesis
Total RNA was extracted from leaf samples using a plant RNA mini kit (Watson, China) according to the manufacturer’s handbook. The first-strand cDNA was transcribed with the TransScript® One-Step gDNA Removal and cDNA Synthesis SuperMix Kit according to the manufacturer’s instructions.

Primer Design and qRT-PCR Verification for Different Tissues and Organs
Because of the large number of genes in the gene family and to rule out absolute tissue-specific preferences for subsequent analysis using transcriptome data (all sequencing materials were leaves), 25 genes that distribute in four groups were selected randomly for analysis to determine the tissue – specific general distribution of the family genes. 21 genes were successfully amplified by the primers we designed. The primers were shown in Supplementary Table S1. All the primers were checked for primer pair specificity using the NCBI (National Center for Biotechnology Information) tool.11 The Solanum lycopersicum actin (Tom41) gene (GenBank ID: U60480.1) was used as a reference gene. The primer efficiencies are listed in Supplementary Table S1. SYBR® Green Master Mix was used for quantitative PCR in an iQ5 real-time PCR detection system. The 2^(-ΔΔCT) method was used to calculate the gene expression (Livak and Schmittgen, 2001). Three replicates were performed for each abiotic stress treatment, and the standard errors of the three replicates were also calculated.

cDNA Library Construction and Sequencing
The analysis process of the disease-resistant response transcriptome is described in Zhao et al. (2019). The samples collected under the NaCl, PEG and low temperature stresses mentioned above were used for RNA-Seq analysis, which was completed by the Beijing Genomics Institute (BGI). First, mRNA with a polyA tail was enriched with oligo-dT magnetic beads. Subsequently, rRNA was hybridized with DNA probes, and then, the DNA/RNA hybrid chains were selectively digested by RNaseH. Finally, the DNA probes were digested with DNasel to purify RNA. After purification, the desired RNA was obtained. The enriched mRNA was broken into short fragments and then reverse transcribed into double stranded cDNA using random hexamer primers. Finally the end of the synthetic double stranded cDNA was smoothed, and the synthetic double stranded cDNA was amplified by PCR. The PCR product was thermally denatured into a single strand, and the single strand DNA was cycled to obtain a single stranded circular DNA library and then sequenced using the Illumina HiSeq platform.

Reads Mapping, Expression Levels, Gene Annotation, and Enrichment Analysis of C2H2-Type Zinc Finger TFs
Raw reads obtained from the Illumina HiSeq platform were processed to filter out adapters, shorter reads and low quality reads. The resulting reads (clean reads) were mapped onto the tomato reference genome using HISAT (Kim et al., 2015). To measure the gene expression levels, the total number of fragments per kilobases per million reads (FPKM) of each gene was calculated based on the length of the gene and the counts of reads mapped to the gene. We used getorf (EMBOSS:6.5.7.0) to detect the ORFs of genes. Then, hmmsearch was used to match each ORF to the transcription factor protein domain. Genes were identified according to the transcription factor family characteristics described by PlantTFDB. Then, we classified the C2H2-type Zinc Finger TFs functionally and used the phyper function in R software for GO and KEGG pathway enrichment analysis.

Real-Time PCR Validation
The mRNA that remained after cDNA library construction was transcribed using the method mentioned above with the cDNA Synthesis SuperMix Kit. To validate the accuracy of the RNA-Seq assay, 5 genes under abiotic and biotic stresses were selected randomly to confirm their relative expression by qRT-PCR. The primers were designed on NCBI. The primers are shown in Supplementary Table S2. qRT-PCR was applied using the method mentioned above.
RESULTS

Identification and Phylogenetic Analyses of C\textsubscript{2}H\textsubscript{2}-ZFP Genes in Tomato

After analysis of the entire genome of the C\textsubscript{2}H\textsubscript{2}-ZFP TF family, 99 C\textsubscript{2}H\textsubscript{2}-ZFPs were identified and named SlZF-1 to SlZF-99. Supplementary Table S3 shows that the molecular masses of the tomato C\textsubscript{2}H\textsubscript{2}-ZFPs range from 11.014 (SlZF-88) to 139.814 kDa (SlZF-4). The pI values of the predicted proteins varied from 3.5367 (SlZF-54) to 10.615 (SlZF-79). The lengths of proteins varied from 96 (SlZF-88) to 1,253 (SlZF-4) amino acids, almost consistent with the molecular masses.

The phylogenetic relationship of the C\textsubscript{2}H\textsubscript{2}-ZFPs was determined using the neighbor-joining (NJ) method with 1,000 bootstrap replicates to construct an unrooted phylogenetic tree from alignments of the full-length C\textsubscript{2}H\textsubscript{2}-ZFP sequences of the 99 proteins from tomato and 116 proteins from Arabidopsis (Figure 1A). Figure 1B shows that all the genes were clearly classified into four groups, labeled from I to IV. Group II, which consists of 29 C\textsubscript{2}H\textsubscript{2}-ZFP genes (SlZF-25 to SlZF-53), is the largest group, followed by groups I and IV, containing 26 C\textsubscript{2}H\textsubscript{2}-ZFP genes each (group I: SlZF-1 to SlZF-10, SlZF-12 to SlZF-24, and SlZF-54 to SlZF-56; group IV: SlZF-11 and SlZF-75 to SlZF-99), and the smallest group is group III, containing 18 C\textsubscript{2}H\textsubscript{2}-ZFP genes (SlZF-57 to SlZF-74). The C\textsubscript{2}H\textsubscript{2}-ZFP family genes of tomato and Arabidopsis are relatively evenly interspersed and distributed among the four groups, indicating that the C\textsubscript{2}H\textsubscript{2}-ZFP genes are closely related in Arabidopsis and tomato.

Gene Structure and Conserved Motifs of the C\textsubscript{2}H\textsubscript{2}-ZFP Genes in Tomato

To examine the structural evolution of the C\textsubscript{2}H\textsubscript{2}-ZFP genes in tomato, the organization of their introns and exons were analyzed using GSDS (Figure 2). Genes in the same group generally had similar numbers of introns. As the schematic structures suggest, the C\textsubscript{2}H\textsubscript{2}-ZFP group III and IV genes (other than SlZF-82 and SlZF-83, which had one intron each) had zero introns. Most of the C\textsubscript{2}H\textsubscript{2}-ZFP genes (69%) in group II had two or three introns. In contrast, the gene structure appeared to be more diverse among the members of group I, with striking differences in their numbers of introns (varying from 0 to 9).

In total, 15 distinct motifs were discovered (Supplementary Table S4). The logos of the 15 motifs are listed in Supplementary Table S5, and the motif distribution in the genes is shown in Figure 2. Most motifs in the same group were distributed similarly. In group II, the linked motifs 5, 1, 11, 4, 3, 2, 6, and
FIGURE 2 | Phylogenetic relations, gene structure and motif compositions of the tomato C\textsubscript{2}H\textsubscript{2}-ZFP genes. The 99 genes were divided into four groups. The four phylogenetic subgroups are marked with different colored backgrounds. (A) The exons and introns of each subgroup are represented by particular colored boxes and gray lines, respectively. (B) Each colored box represents a motif, and the black lines represent non-conserved sequences.

14 and the linked motifs 12, 1, 8, 3, and 7 were exhibited by the groups SlZF-25 to SlZF-39 and SlZF-49 to SlZF-53, respectively. Additionally, all of the C\textsubscript{2}H\textsubscript{2}-ZFPs in group III (16/18) except SlZF-69 and SlZF-73 had the linked motifs 1 and 15; those two exceptions both had the linked motifs 9, 1, and 10. Nearly half of the C\textsubscript{2}H\textsubscript{2}-ZFPs in group III (7/18) had the linked motifs 9, 1, 3, and 15. Most of the C\textsubscript{2}H\textsubscript{2}-ZFPs in group IV (21/26) had the linked motifs 1 and 13.

Chromosomal Locations of Tomato C\textsubscript{2}H\textsubscript{2}-ZFP Genes
To unravel the evolutionary patterns of this C\textsubscript{2}H\textsubscript{2}-ZFP gene family, we analyzed the locations of the 99 tomato C\textsubscript{2}H\textsubscript{2}-ZFP genes, finding that combinations of the 96 were distributed across all 12 chromosomes (Figure 3). The other three genes, SlZF-14, SlZF-79 and SlZF-88, were not be mapped due to the lack of complete annotation in SGN (Sol Genomics Network).

Promoter cis-Element Analysis
Here, analysis of the Plant CARE database revealed multiple cis-acting elements in the promoter regions between −1 and 2000 bp upstream of the transcription start sites of all the putative tomato C\textsubscript{2}H\textsubscript{2}-ZFP genes. Figure 4 and Supplementary Table S6 show the presence of multiple cis-acting elements related to the phytohormone and environmental stress signal responsiveness of the C\textsubscript{2}H\textsubscript{2}-ZFP family promoters such as the C-repeat/DRE (cold- and dehydration-responsive element), HSE (heat stress-responsive element), LTR (low-temperature-responsive element), ABRE and CE3 (abscisic acid-responsive elements), CGTCA-motif and TGACG (MeJA-responsive elements), GARE-motif and TATC-box (gibberellin-responsive elements), SARE and TCA-element (salicylic acid-responsive elements), and ERE (ethylene-responsive element). The HSE and TCA-element were found in 63 and 58 promoters of the C\textsubscript{2}H\textsubscript{2}-ZFP genes, respectively. Several cis-acting elements involved in tissue-specific expression, e.g., an as-2-box (involved in shoot-specific expression), a GCN4\textsubscript{4} motif (involved in endosperm expression), an HD-Zip 2 (involved in the control of leaf morphology development), and an RY-element (involved in seed-specific regulation), were also found. Several cis-elements were unique to individual C\textsubscript{2}H\textsubscript{2}-ZFP genes; for instance, the C-repeat/DRE, CE3, and SARE motifs were present only in SlZF-13, SlZF-45, and SlZF-89, respectively.

Expression Profiles of the Tomato C\textsubscript{2}H\textsubscript{2}-ZFP Genes in Different Tissues and Organs
To further understand the tissue specificity of the C\textsubscript{2}H\textsubscript{2}-ZFP genes, the expression patterns of 21 genes across a variety of tomato tissues, including roots, stems, leaves, flowers, green...
fruits, mellow fruits and sepals, were analyzed by the qRT-PCR method. As shown in Figure 5, two genes (SlZF-24 and SlZF-28) were significantly highly expressed in root, and two genes (SlZF-6 and SlZF-13) were significantly highly expressed in flower. Furthermore, 10 genes (SlZF-6, SlZF-13, SlZF-15, SlZF-17, SlZF-18, SlZF-20, SlZF-22, SlZF-24, SlZF-28, and SlZF-30) and
Expression patterns of the \( \text{C}_2\text{H}_2\)-ZFP genes in different tomato tissues and organs. (A) Whole root; (B) young stems; (C) mature green leaves; (D) flowers at anthesis; (E) 20 days post-anthesis fruit; (F) breaker stage ripening fruit; (G) sepals. All tissues or organs were collected 4 months after sowing. Brown indicates higher expression levels, and bottle green indicates lower expression levels. Genes with similar profiles across arrays are grouped on the right by a hierarchical clustering method.

Eight genes (\( \text{SlZF-15, SlZF-17, SlZF-18, SlZF-22, SlZF-30, SlZF-31, SlZF-36, and SlZF-39} \)) were highly expressed in root and green fruit, respectively. \( \text{SlZF-22} \) was expressed in many tissues and organs. Interestingly, all of the 21 genes were expressed similarly in leaves. Genes with similar profiles across arrays were grouped on the right by a hierarchical clustering method.

\text{C}_2\text{H}_2\)-ZFP Gene Expression Pattern Analysis Based on RNA-Seq Data

Four groups of RNA-Seq data, including tomato drought, salt, cold, and pathogen infection response transcriptome data, were used for this analysis. The statistics of the pathogen infection response transcriptome data information are shown in Zhao et al. (2019), and around 6.4 G of data were generated per sample. The details of raw reads, clean reads and read quality of each sample are shown in Supplementary Table S7. Through RNA-Seq data analysis, 50 \( \text{C}_2\text{H}_2\)-ZFP genes were screened under drought stress (Figure 6), 49 \( \text{C}_2\text{H}_2\)-ZFP genes were screened under cold stress (Figure 7), 50 \( \text{C}_2\text{H}_2\)-ZFP genes were screened under salt stress (Figure 8) and 51 genes were screened under pathogen infection (Figure 9). The FPKM of each \( \text{C}_2\text{H}_2\)-ZFP genes in three replicates at different time points under the four stresses are shown in Supplementary Tables S8–S11. We screened and identified genes whose absolute value of \( \log_2(\text{FPKM of second time point/FPKM of CK0}) \) was greater than 1 as differentially expressed genes (DEGs) and obtained 28 drought-related genes, 31 salt-related genes, 13 cold-related genes and 24 pathogen-related genes. As shown in Figure 10A (a detailed gene list is shown in Supplementary Table S12), a small number of genes was involved in only one stress response, and most genes were involved in more than one stress response. Three genes (\( \text{SlZF-58,} \))
SlZF-61, SlZF-5) participated in all four stress responses. The numbers of upregulated genes and downregulated genes involved in the four stresses were similar. The numbers of upregulated and downregulated genes in the three abiotic stresses were similar, but the number of upregulated genes was significantly higher than the number of downregulated genes under biotic stress (Figure 10A). Further analysis shows (Figure 10B) that the highest proportion of the same response genes was 34% under drought and salt stresses, followed by pathogenic and salt stresses at 33%. Additionally, the lowest proportion of the same response genes was 15% under drought and cold stresses. The statistics of the genes that specifically respond to only one of the four kinds of stresses (including differentially expressed genes under only one stress and genes whose expression patterns were different from those under the other three stresses) were evaluated. The results are shown in Figure 10C. A total of 23 genes were screened out. Seven genes were specifically related to drought stress, five genes were specifically related to salt stress, three genes were specifically related to cold stress, and eight genes were specifically related to pathogen stress.

**GO Enrichment Analysis of the C$_2$H$_2$-ZFP Genes**

GO enrichment of all the C$_2$H$_2$-ZFP genes was performed. All 53 C$_2$H$_2$-ZFP genes identified through transcriptome data were involved in nucleic acid binding, organic cyclic compound binding, heterocyclic compound binding and binding. These four
terms all belong to the molecular function category. Twenty C2H2-ZFP genes were involved in the nucleus, intracellular membrane-bounded organelle, membrane-bounded organelle, organelle and intracellular organelle, which belong to the cellular component category. Additionally, a few C2H2-ZFP genes were enriched in other GO terms (Figure 11).

**Analysis of the KEGG Metabolic Pathway Enrichment and Expression Patterns in Different Pathways of the C2H2-ZFP Genes**

The C2H2-ZFP genes were enriched into 6 KEGG pathways (Q-value ≤ 0.05), including glycosphingolipid biosynthesis, glycosaminoglycan degradation, sphingolipid metabolism, other glycan degradation, galactose metabolism, and endocytosis (Figure 12). Twenty-three C2H2-ZFP genes were enriched in the first five pathways. Seven C2H2-ZFP genes were enriched in the endocytosis pathway. Therefore, the genes enriched in these six pathways could be divided into two groups: the genes in the first five pathways were in one group (Group 1), and the genes in the endocytosis pathway were in another group (Group 2). The expression patterns of these genes were statistically analyzed under different stresses, and the results are shown in Figure 13A. The expression patterns of genes in the same group were different under different stresses, and the proportions of upregulated genes and downregulated genes were significantly different. In the first group, there were fewer upregulated genes than downregulated...
Validation of Gene Expression of Genes via qRT-PCR

To validate the expression data obtained from RNA-Seq, SlZF-58, SlZF-68, SlZF-30, SlZF-33, and SlZF-47 were chosen for confirmation by qRT-PCR under the four kinds of stresses. The expression levels of the five genes were basically consistent with the results of RNA-Seq (Figure 14). These results indicate that the RNA-Seq results in this study are valid.

DISCUSSION

The Evolution of the C2H2-ZFPs Family Among Different Species Is Conservative

C2H2-ZFPs are related to plant growth and development, abiotic stress, biotic stress, and other phenomena [14,18,19], showing that C2H2-ZFPs play an important role in plant life processes. In this study, we identified 99 potential C2H2-ZFPs in tomato, and the full-length C2H2-ZFP sequences of these 99 proteins were classified into four groups, labeled from I to IV, through an unrooted phylogenetic tree. We found that the phylogenetic relationship between the C2H2-ZFPs belonging to set A of the Arabidopsis C2H2-ZFPs and the group II C2H2-ZFPs of tomato was close. The phylogenetic relationship between the
C2H2-ZFPs belonging to set Cl of the Arabidopsis C2H2-ZFPs and groups III and IV C2H2-ZFPs of tomato was close, though all of the Arabidopsis C2H2-ZFPs are in group I, which may demonstrate that different plants have conserved different duplicates. Furthermore, the physical parameters of C2H2-ZFPs in different plants have been shown to be very similar. For example, the molecular masses of the tomato C2H2-ZFP genes range from 11.014 (SlZF-88) to 139.814 kDa (SlZF-4), the molecular masses of the maize C2H2-ZFP genes range from 10.535 to 179.433 kDa and the molecular masses of poplar C2H2-ZFP genes range from 17.626 to 191.195 kDa (Liu et al., 2015; Wei et al., 2015). This similarity suggests that the C2H2-ZFP family is relatively conserved in evolution. The protein-coding sequences of eukaryotes are interrupted by introns [20]. To further study the distribution of the introns in the genome of tomato, we investigated C2H2-ZFP TFs and the relationship between tomato C2H2-ZFP TFs and those of other species. The number of introns within the C2H2-ZFP groups III and IV genes was either 0 or 1, while gene structure of the numbers within the C2H2-ZFP groups I and II appeared to be more variable and irregular in terms of intron numbers. These values are very similar to the numbers
FIGURE 10 | Comparisons of differentially expressed genes and genes with different response patterns under the four stresses. (A) Venn diagrams showing the numbers the up- and down-regulated genes in the four comparisons. Nd(s/c/p)s: the number of differentially expressed genes under drought (salt/cold/pathogen) stress; Nud(s/c/p)s: the number of upregulated genes under drought (salt/cold/pathogen) stress; Ndd(s/c/p)s: the number of downregulated genes under drought (salt/cold/pathogen) stress. (B) Proportions of the identical DEGs under any two stresses. (C) A representation of a gene specifically expressed; “U” means upregulated; “D” means downregulated; “N” means no change.

of Maize and Poplar C2H2-ZFP introns. Most maize ZmZFP genes in their group C had 0 or 1 intron with the exception of a few members. In contrast, the gene intron–exon structures are more variable in group A and group B, which had more intron numbers. Poplar C2H2-ZFP genes in their group I and II possessed zero to three introns except for one gene. By contrast, the numbers of introns of genes in groups III and IV are large and irregular (Liu et al., 2015; Wei et al., 2015). Ninety-six of all 99 tomato C2H2-ZFP genes were unevenly distributed across all 12 chromosomes. Gene duplication plays an important role in genomic expansion (Vision et al., 2000), and the uneven distribution of the locations of the C2H2-ZFP genes on the chromosomes may be related to the duplication of the tomato genome (Ma et al., 2015). Nevertheless, all four groups of the C2H2-ZFP gene family were found on chromosomes 1, 3, 5, 6, 9, and 11, suggesting that members of different C2H2-ZFP gene family groups may interact via a network of proteins.

C2H2-ZFPs Gene Families May Have Multiple Functions
Fifteen motifs were found in this study. Most C2H2-ZFPs in tomato have motif 1, whose best possible match sequence contains the QALGGH amino acid sequence. Many plant ZFPs that are plant specific contain this motif (Agarwal et al., 2007; Liu et al., 2015; Wei et al., 2015). High conservation of the QALGGH sequence is critical for DNA-binding activity (Ken-ichi et al., 1998). In addition, ‘DLELRL’ (motif 13), a hexapeptide motif, has been found in 24 C2H2-ZFPs. This hexapeptide alone is sufficient to confer the ability to repress transcription to a DNA-binding domain in Arabidopsis (Hiratsu et al., 2004). Another repressor is the ERF-associated amphiphilic repression (EAR) motif, or DLN-box, which has been found in 22 C2H2-ZFPs. Stress-associated EAR-repressors may control the initiation of stress-activated gene expression under various stresses such as cold, salt, drought, and oxidative stress (Kazan, 2006; Ciftci-Yilmaz et al., 2007; Kielbowicz-Matuk, 2012). Motif 9 is a short, leucine-rich region...
with the core sequence EXEXXAXCLXXL (L-box) and is thought to play roles in protein–protein interactions or in maintaining folded structures (Kielbowicz-Matuk, 2012). We also discovered many motifs that may be related to DNA-binding or protein–protein interactions.

Most of the C$_2$H$_2$-ZFPs genes in tomato have abiotic stress-related and stress response cis-elements, including C-repeat/DRE, HSE, LTR, and TC-rich repeats. In addition, many C$_2$H$_2$-ZFPs genes have more than one phytohormone-related cis-elements (ABRE, CE3, CGTCA-motif, TGACG, GARE-motif, TATC-box, SARE, TCA-element, and ERE). Phytohormones such as salicylic acid (SA), jasmonic acid (JA), ethylene (ET), and abscisic acid (ABA) play important roles in plant responses to biotic and abiotic stresses and in the regulation...
of developmental processes (Vision et al., 2000). These findings suggest that C2H2-ZFPs genes may modulate the biotic and abiotic stress responses directly or indirectly in tomato.

In this study, transcriptome data analysis showed that some of the genes in this family had significant changes in expression under four stresses, indicating that these genes were involved in the stress resistance and disease resistance responses of tomato plants. A C2H2-ZFP encoded by OsMSR15 has been previously shown to have positive effects on cold, drought, and heat stress in different tissues of rice at different stages of development (Zhang et al., 2016a). SiZFP1, which was cloned from potato, can be induced under abiotic stresses, such as salt, drought, and exogenous ABA (Tian et al., 2010). GmZF1 in soybean regulates the expression of cold-regulated genes in transgenic Arabidopsis thaliana (A. thaliana) and enhances the anti-cold ability of A. thaliana (Yu et al., 2014). GO enrichment results showed that all the C2H2-ZFP genes screened through RNA-Seq were enriched in the molecular function category, indicating that all the C2H2-ZFP genes have multiple functions.

In addition, in the analysis of tissue-specific expression, we found that the expression of many genes was obviously tissue-specific. Previous studies have revealed that the C2H2-ZFP family of TFs have broad expression patterns under normal growth conditions in many organs and tissues. RABBIT EARS (RBE), a C2H2-ZFP gene, plays an important role in the development of petals in A. thaliana (Takeda et al., 2004). The product of RBE inhibits the expression of TCP4 in A. thaliana at the early stage of petal development. The expression of TCP4 in the flower of rbe-1 mutant plants is significantly higher than that in the flower of wild-type plants, which fully indicates that the expression of RBE is related to TCP4 (Li et al., 2016). In this study, the results showed that SIZF-24 and SIZF-28 were significantly highly expressed in root and that SIZF-6 and SIZF-13 were preferentially expressed in flower, providing some clues about the functions of these genes, which may be related to root and flower development. Previous studies have revealed that root and flower development are influenced by genes such as SUPERMAN, a C2H2-ZFP in Arabidopsis that is a regulator of floral homeotic genes, and TaZFP34, which is a transcriptional repressor in wheat and is involved in modulating the root-to-shoot ratio (Chang et al., 2016).

### Many C2H2-ZFP Genes of Tomato Respond to Multiple Stresses

Thirty-two genes were involved in differential expression under at least two stresses, while only 13 DEGs were involved in one type of stress, proving that most genes respond to multiple stresses. Previous studies have shown that some of the C2H2-ZFP genes are indeed associated with multiple stresses. IbZFP1, which is a Cys2/His2 zinc finger protein gene from sweet potato, is induced by NaCl and PEG. Overexpression of IbZFP1 significantly enhances salt and drought tolerance in transgenic Arabidopsis plants (Wang et al., 2015). ZFP245 is a cold- and drought-responsive gene that encodes a C2H2-ZFP in rice. Overexpression of ZFP245 results in enhanced tolerance under...
drought and cold stresses (Huang et al., 2009). A total of 11 DEGs were involved in the three stresses, and 7 genes had the same expression patterns under all three stresses, indicating that most of the DEGs involved in the three stresses had the same expression pattern. Six DEGs were identified under salt, pathogen and drought stresses, among which 4 genes were all upregulated and 1 gene was downregulated under these three stresses. However, only 1 DEG was the same under cold, drought and salt stresses (SLZF-73), and was downregulated under all three stresses. Under the four stresses, the cold-related DEGs were the fewest in number, and the cold-specific response genes were also the fewest in number. Indicating that the C2H2-ZFP genes have the weakest response to cold stress.

**Differences May Exist Among the Regulation of C2H2-ZFP Genes Under Different Stresses**

KEGG analysis showed that the genes of this family were mainly enriched in six pathways including glycosphingolipid biosynthesis, glycosaminoglycan degradation, sphingolipid metabolism, other glycan degradation, galactose metabolism, and endocytosis. The first five pathways are all related to the function of glycometabolism and sphingolipid, and the endocytosis pathway is related to membrane function, all of these functions are closely related to the resistance of plants to environmental stress (Jae et al., 2008; Ali et al., 2018; Couchoud et al., 2019). In this study, many C2H2-ZFP genes were differentially expressed in these pathways, suggesting that this family of genes may play a role in stress tolerance regulation through these pathways. In the pathways related to glycometabolism and sphingolipid, the proportion of upregulated C2H2-ZFP genes under biotic stress was higher than that of the proportion of genes upregulated under abiotic stresses, indicating that the sensitivity of the regulatory points of this family gene in the glycometabolism and sphingolipid pathways to biotic and abiotic stresses is different. In the endocytosis pathway, the proportion of upregulated genes under drought stress was the lowest, which was significantly different from the other three stresses. The proportions of upregulated genes under pathogen stress and cold stress were similar, and the data distribution was very different from that in the glycometabolism and sphingolipid -related pathways, indicating that the regulatory points of C2H2-ZFP genes in the two pathways are relatively independent. In the analysis of stress-specific response genes, we found that all the drought stress-specific response genes were enriched in the glycometabolism and sphingolipid pathways, and that all of them were upregulated, indicating that the C2H2-ZFP family has drought-specific response factors and that these response factors regulate the drought stress response only by enhancing processes related to glycometabolism and sphingolipid. There are specific.
response genes in both types of pathways under other stresses, the most notable of which is salt stress. Additionally, the specific response genes are all downregulated under salt stress. Therefore, it is speculated that under salt, cold and pathogen stresses, although some members of the C2H2-ZFP family regulate plant' resistance to environmental stress through the glycometabolism and sphingolipid-related pathways and the endocytosis pathway, the regulation process is more specific under salt stress compared to the processes under cold and pathogen stresses.

**CONCLUSION**

Ninety-nine C2H2-type zinc finger protein genes were identified and classified into four groups based on phylogenetic analysis. These genes were distributed across all 12 chromosomes and were expressed in different tissues or organs including roots, stems, leaves, flowers and fruits, some of which showed tissue-specific expression patterns. Thirty-two C2H2-ZFP genes were identified in response to multiple stresses. Seven, 3, 5, and 8 genes were specifically expressed under drought, cold, salt, and pathogen stresses, respectively. KEGG analysis revealed that the genes in the C2H2-ZFP family mainly exert anti-stress regulation through the glycosphingolipid biosynthesis, glycosaminoglycan degradation, sphingolipid metabolism, other glycan degradation, galactose metabolism, and endocytosis pathways. The results of this study are the foundation for further studies on the function of C2H2-type zinc finger proteins and will provide a basis for stress tolerance breeding.

**REFERENCES**

Agarwal, P., Arora, R., Ray, S., Singh, A. K., Singh, V. P., Takatsuji, H., et al. (2007). Genome-wide identification of C2H2 zinc-finger gene family in rice and their phylogeny and expression analysis. Plant. Mol. Biol. 65, 467–485. doi: 10.1007/s11103-007-9199-y

Ali, U., Li, H., Wang, X., and Guo, L. (2018). Emerging roles of sphingolipid signaling in plant response to biotic and abiotic stresses. Mol. Plant. 11, 1328–1343. doi: 10.1016/j.molp.2018.10.001

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37, W202–W208. doi: 10.1093/nar/gkp335

Carl, O. P., Ezra, P., and Robert, A. J. (2001). Design and selection of novel Cys2His2 Zinc Finger Proteins. Annu. Rev. Biochem. 70, 313–340. doi: 10.1146/annurev.biochem.70.1.313

Chang, H., Chen, D., Kam, J., Richardson, T., Drenth, J., Guo, X., et al. (2016). 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. 252, 88–102. doi: 10.1016/j.plantsci.2016.07.011

Ciftci-Yilmaz, S., Morsy, M. R., Song, L., Coutou, A., Krizek, B. A., Lewis, M. W., et al. (2007). The EAR-motif of the Cys2His2-type zinc finger protein Zat7 plays a key role in the defense response of Arabidopsis to salinity stress. J. Biol. Chem. 282, 9260–9268. doi: 10.1074/jbc.M610932000

Couchoud, M., Der, C., Girotet, S., Vernoud, V., Prudent, M., and Leborgne-Castel, N. (2019). Drought stress stimulates endocytosis and modifies membrane lipid order of rhizodermal cells of Medicago truncatula in a genotype-dependent manner. BMC Plant Biol. 19:221. doi: 10.1186/s12870-019-1814-y

Elisabeth, G., Christine, H., Alexandre, G. D., Marc, R. W., Ron, D. A., and Amos, B. (1999). “Protein identification and analysis tools on the ExPASy server,” in The Proteomics Protocols Handbook, ed. J. M. Walker (Totowa, NJ: Humana Press), 571–607. doi: 10.1385/1-59259-584-7:531

Engbrecht, C. C., Schoof, H., and Bohn, S. (2004). Conservation, diversification and expansion of C2H2 zinc finger proteins in the Arabidopsis thaliana genome. BMC Genomics 5:39. doi: 10.1186/1471-2164-5-39

Hiratsu, K., Mitsuda, N., Matsu, K., and Ohme-Takagi, M. (2004). Identification of the minimal repression domain of SUPERMAN shows that the DLELR hexapeptide is both necessary and sufficient for repression of transcription in Arabidopsis. Biochem. Biophys. Res. Commun. 321, 172–178. doi: 10.1016/j.bbrc.2004.06.115

Hiroshi, T., Norihisa, N., and Yukihisa, K. (1994). A new family of zinc finger proteins in petunia: structure, DNA sequence recognition, and floral organ-specific expression. Plant Cell 6, 947–958. doi: 10.1105/tpc.6.7.947

Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GDSD 2.0: an upgraded gene feature visualization server. Bioinformatics 31, 1296–1297. doi: 10.1093/bioinformatics/btu817

Huang, J., Sun, S. J., Xu, D. Q., Yang, X., Bao, Y. M., Wang, Z. F., et al. (2009). Increased tolerance of rice to cold, drought and oxidative stresses mediated by the overexpression of a gene that encodes the zinc finger protein ZFP245. Biochem. Biophys. Res. Commun. 389, 556–561. doi: 10.1016/j.bbrc.2009.09.032

Huang, J., Wang, J., and Zhang, H. (2005). Rice ZFP15 gene encoding for a novel C2H2-type zinc finger protein lacking DLN box, is regulated by spike development but not by abiotic stresses. Mol. Biol. Rep. 32, 177–183. doi: 10.1007/s11033-005-2338-2330

**DATA AVAILABILITY STATEMENT**

The datasets generated for this study can be found in the NCBI GEO accession GSE148530.

**AUTHOR CONTRIBUTIONS**

XX and TZ designed the study and supervised the project and revised the manuscript. TZ, TW, JZ, TP, and ZW implemented the study and acquired the data. TW and TZ drafted the manuscript. HY and JL helped to drafted the manuscript. All authors read and approved the final manuscript.

**FUNDING**

This study was supported by the National Key R&D Program of China (2017YFD0101900), the China Agriculture Research System (CARS-23-A-16), the China postdoctoral science foundation grant (2018 M630333), and the University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (UNPYSCT-2018169).

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00540/full#supplementary-material
Jae, S. K., Julia, F., Chang, H. K., Hiroyuki, K., Meenu, V., Akhiro, U., et al. (2008). Salt tolerance of Arabidopsis thaliana requires maturation of N-glycosylated proteins in the Golgi apparatus. Proc. Natl. Acad. Sci. U S A 105, 5933–5938. doi: 10.1073/pnas.0800237105

Jiang, L., and Pan, L. J. (2012). Identification and expression of C2H2 transcription factor genes in Carica papaya under abiotic and biotic stresses. Mol. Biol. Rep. 39, 7105–7115. doi: 10.1007/s11033-012-1542-y

Julie, D. T., Toby, J. G., Frédéric, P., François, J., and Desmond, G. H. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882. doi: 10.1093/nar/25.24.4876

Kazan, K. (2006). Negative regulation of defence and stress genes by EAR-motif-containing repressors. Trends Plant Sci. 11, 109–112. doi: 10.1016/j.tplants.2006.01.004

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360. doi: 10.1038/nmeth.3317

Li, J., Yanzhi, W., Yongxia, Z., Weiyao, W., Vivian, F. I., and Tengbo, H. (2016). Kazan, K. (2006). Negative regulation of defence and stress genes by EAR-motif-containing repressors. Trends Plant Sci. 11, 109–112. doi: 10.1016/j.tplants.2006.01.004

Ken-ichi, K., Ayako, S., Akira, K., Zbigniew, R., Yoshiaki, K., Hitoshi, N., et al. (2008). Identification and expression of C2H2-type zinc finger transcription factors and their responses to abiotic stresses in poplar (Populus trichocarpa). BMC Genomics 9:e109399. doi: 10.1371/journal.pone.0109399

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408. doi: 10.1006/meth.2001.1262

Ma, J., Li, M. Y., Wang, F., Tang, J., and Xiong, A. S. (2015). Genome-wide analysis of DoF family transcription factors and their responses to abiotic stresses in Chinese cabbage. BMC Genomics 16:33. doi: 10.1186/s12864-015-1242-1249

Mueller, L. A., Tanksley, S. D., Giovannoni, J. J., van Eck, J., Stack, S., Choi, D., et al. (2005). The tomato sequencing project, the first cornerstone of the international Solanaceae project (SOL). Comp. Funct. Genomics 6, 153–158. doi: 10.1002/cfg.468

Riano-Pachon, D. M., Ruzicic, S., Dreyer, I., and Mueller-Roeber, B. (2007). PhTFDB: an integrative plant transcription factor database. BMC Bioinformatics 8:42. doi: 10.1186/1471-2105-8-42

Sang, H. K., Jeum, K. H., Sung, C. L., Kee, H. S., Ho, W. J., and Byung, K. H. (2004). CAZFP1, Cys2/His2-type zinc-finger transcription factor gene functions as a pathogen-induced early-defense gene in Capsicum annuum. Plant Mol. Biol. Rep. 55, 883–904. doi: 10.1007/s11103-005-2151-2150

Takeda, S., Matsumoto, N., and Okada, K. (2004). RABBIT EARS, encoding a SUPERMAN-like zinc finger protein, regulates petal development in Arabidopsis thaliana. Development 131, 425–434. doi: 10.1242/dev.00938

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739. doi: 10.1093/molbev/msr121

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zhao, Wu, Zhang, Pei, Yang, Li and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.