Journal of Cotton Research

RESEARCH Open Access

Zinc finger transcription factor ZAT family genes confer multi-tolerances in *Gossypium hirsutum* L.

FAN Yapeng, ZHANG Yuexin, RUI Cun, XU Nan, ZHANG Hong, WANG Jing, MALIK Waqar Afzal, HAN Mingge, ZHAO Lanjie, LU Xuke, CHEN Xiugui, CHEN Chao and YE Wuwei*

Abstract

ZAT (Zinc Finger of *Arabidopsis thaliana*) proteins are composed of a plant-specific transcription factor family, which play an important role in plant growth, development, and stress resistance. To study the potential function of ZAT family in cotton, the whole genome identification, expression, and structure analysis of ZAT gene family were carried out. In this study, our analysis revealed the presence of 115, 56, 59, and 115 ZAT genes in *Gossypium hirsutum*, *G. raimondii*, *G. arboreum* and *G. barbadense*, respectively. According to the number of domains and phylogenetic characteristics, we divided ZAT genes of four *Gossypium* species into 4 different clades, and further divided them into 11 subfamilies. The results of collinearity analysis showed that segmental duplication was the main method to amplify the cotton ZAT genes family. Analysis of *cis*-elements of promoters indicated that most *GhZAT* genes contained *cis*-elements related to plant hormones and abiotic stress. According to heatmap analysis, the expression patterns of *GhZAT* genes under different stresses indicated that *GhZAT* genes were significantly involved in the response to cold, heat, salt, and PEG stress, possibly through different mechanisms. Among the highly expressed genes, we cloned a *G. hirsutum* gene *GhZAT67*. Through virus-induced gene silencing (VIGS), we found that its expression level decreased significantly after being silenced. Under alkaline treatment, the wilting degree of silenced plants was even greater than the wild type, which proved that *GhZAT67* gene was involved in the response to alkaline stress.

Keywords: ZAT, Phylogenetic analysis, Collinearity analysis, *cis*-elements, VIGS

Background

Plants were constantly affected by a wide variety of biotic and abiotic factors during their growth and development. In response to these challenges, some responsive genes, including zinc finger genes, were induced to resist adversity through physiological and morphological changes. Zinc finger proteins (ZFPs) were one kind of the most abundant transcription factors in higher plants (Takatsuji 1999). According to the number and location of cysteine and histidine residues, ZFPs were classified into different groups, including C_2H_2, C_2HC_5, C_2C_2, CCCH, C_3HC_4, C_4, C_4HC_3, C_6, and C_8 (Sun et al. 2019). Previous studies on ZFP transcription factors found that they played an important role to promote plant growth and development, and most of the ZFPs transcription factors proteins may participate in the response to various stresses, including cold, hot, drought, salt, oxidation stress, osmotic stress, and wound stress (Kiełbowicz-Matuk 2012; Wang et al. 2009).

ZAT genes belong to the subclass C1-2i of C_2H_2 ZFPs family (Xie et al. 2019). As an important part of the C_2H_2 ZFPs family, ZAT genes played an important role in resisting adversity. Studies had found that they were widely involved in the response to abiotic stresses. In *Arabidopsis*, *ZAT12* was mainly related to oxidative

* Correspondence: yew158@163.com
Key Laboratory of Cotton Biology, Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang 455000, Henan, China

stress (Le et al. 2016; Rizhsky et al. 2004). *ZAT12* was also involved in response to strong light, ROS signal transduction, and low temperature in *Arabidopsis* (Davletova et al. 2005; He et al. 2020; Vogel et al. 2005; Wang et al. 2020). The exogenous gene *GmZAT4* expressed in *Arabidopsis* can enhance the tolerance of transgenic *Arabidopsis* to drought and salt stress (Sun et al. 2019). The *ZAT7* transgenic plants in *Arabidopsis* can inhibit growth and show higher tolerance to salt stress (Ciftci-Yilmaz et al. 2007). Studies found that transgenic plants expressing *Zat10* gene had stronger resistance to drought, heat, osmotic stress, and salt stress (Mittler et al. 2006; Sakamoto et al. 2004). Overexpression of *Zat10* can increase the transcription of ascorbate peroxidase 1, 2 (*APX1*, *2*) and iron superoxide dismutase 1 (*FSD1*), thereby eliminating ROS in plants (Mittler et al. 2006). It was shown that the ZAT gene family indeed played an important role in adversity stresses. However, there was no study on ZAT genes in cotton, so it was necessary to analyze the ZAT gene family.

Cotton was not only a worldwide leading textile fiber crop but also a crop with significance for petroleum and biofuel products (Zhang et al. 2008). However, the growth and production of cotton were severely restricted by external and internal environmental factors. The analysis of the whole cotton genome and the development of transcriptomics had laid the foundation for the study of the expression patterns and functions of different gene families. In this study, we conducted a preliminary analysis of the gene structure, evolutionary relationship, expression pattern, and *cis*-acting elements of *Gossypium hirsutum*. A ZAT family gene, *GhZAT67*, was isolated and characterized. VIGS plants with *GhZAT67* gene silenced were sensitive to alkaline stress. This study provided potential candidate genes for the study of cotton gene function and provided a certain molecular basis for cotton breeding.

Materials and methods

Identification of *GhZAT* gene family members in *Gossypium hirsutum* L.

To identify the members of the *GhZAT* gene family, the protein sequence and genome annotation of upland cotton were downloaded from Cotton Functional Genomic Database (CottonFGD) (http://www.cottonfgd.org/) (Zhu et al. 2017). The Hidden Markov Model (HMM) of PF13912 was downloaded from the Pfam (https://pfam.xfam.org/) website. HMMER 3.0 software was used to obtain the protein sequence of PF13912, Online sites Pfam (https://pfam.xfam.org/) and SMART (http://smart.emblheidelberg.de/) with default parameters were used to further screen the genes. We also retrieved some other features of ZAT genes of upland cotton like isoelectric points (pIs), molecular weights (MWs), exon/intron lengths, grand average of hydropathy, and charge by using CottonFGD (http://www.cottonfgd.org/). Subcellular localization prediction of *GhZAT* genes was carried out by using several websites, such as TargetP (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson et al. 2000), ProtComp (http://linux1.softberry.com/berry.phtml?topic=protcomppl&group=programs&subgroup=proloc), WOLF-PSORT (https://wolfpsort.hgc.jp/) and CELLO v.2.5 (http://cello.life.nctu.edu.tw/) (Yu et al. 2006).

Phylogenetic analysis

To study the evolutionary relationship among *ZAT* genes in different species, CottonFGD (http://www.cottonfgd.org/) was used to obtain ZAT protein sequences in other three cotton species (*G. barbadense*, *G. arboreum*, and *G. raimondii*). Protein sequences of other species like *Theobroma cacao* (version 10), *A. thaliana* (TAIR 10), *Oryza sativa* (version 7), *Populus trichocarpa* (version 3.0), *Vitis vinifera* Genoscope (version 12), *Glycine max* (version 10), *Zea mays* (version 10) were obtained from Phytozome v12.1 (https://phytozome.jgi.doe.gov/pz/portal.html) by using PF13912 as the keyword, and MEGA7.0 software was used for protein sequence alignment. Subsequently, the neighbor-joining (NJ) tree with 1 000 bootstrap replicates was constructed using the Poisson substitution model with default parameters in MEGA 7.0 (Kumar et al. 2016).

Chromosomal locations of *ZAT* genes from four *Gossypium* species

Physical positions of chromosomal locations from four cotton species including *G. hirsutum*, *G. arboreum*, *G. raimondii,* and *G. barbadense* were drawn by TB Tools software (Chen et al. 2018). Genomic sequences, coding sequences of all of the four species were downloaded from CottonFGD (http://www.cottonfgd.org/) (Zhu et al. 2017). Genome sequences of four *Gossypium* species (*G. hirsutum*, NAU; *G. barbadense*, HAU; *G. arboreum*, CRI; and *G. raimondii*, JGI) were used to identify the gene family.

Collinearity analysis of *ZAT* genes in four *Gossypium* species

Synteny relationship between duplicated gene pairs from four cotton species *G. hirsutum*, *G. arboreum*, *G. raimondii,* and *G. barbadense* was analyzed by using MCScanX software (Wang et al. 2012) and diagrammatical results were visualized by using Circos-0.69 software (http://circos.ca/) (Krzywinski et al. 2009).

Calculation of selection pressure

Duplicated gene pairs from four *Gossypium* species including *G. arboreum*, *G. raimondii*, *G. hirsutum,* and *G. barbadense* were identified by using MCScanX software. The nonsynonymous to the synonymous ratio (*Ka*/*Ks*)

was calculated for duplicated genes to investigate the selection pressure by using TB Tools software.

Analysis of the conserved protein motifs and gene structure

Using Multiple Em for Motif Elicitation (MEME, http://meme-suite.org/) to identify the conserved motifs (Bailey et al. 2009). TB Tools software was used to map the evolutionary relationship, gene structure, and conserved motifs of *GhZAT* genes with MAST file from MEME website, NWK file from phylogenetic tree analysis, and GFF3 genome file of *G. hirsutum* L.

Analysis of *GhZAT* promoter regions

DNA sequences of 2 000 bp in upstream regions of *GhZAT* genes were obtained from CottonFGD database (http://www.cottonfgd.org/) as promoters. PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used for the prediction of *cis*-acting elements in promoter regions of *GhZAT* genes.

Analysis of differentially expressed genes (DEGs)

RNA-Seq data (PRJNA248163) downloaded from National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/) was used to analyze differentially expressed genes under salt, PEG, cold, and heat stress (Hu et al. 2019). The heat map along with phylogenetic tree and *cis*-acting elements was generated through TB Tools software by using fragments per kilo base of exon per million fragments mapped (FPKM).

VIGS experiment of *GhZAT67*

The purified *GhZAT67* fragment was inserted into the empty pYL156 to form the pYL156: *GhZAT67* vector, with the sites of *Sac*I and *Bam*HI. The GV3101 strains carrying pYL156 (empty vector), pYL156: *GhZAT67* (VIGS), pYL156: PDS (positive control), and pYL192 (helper vector) were cultured to $OD_{600} = 1.21.5$. Each mixture was injected into the underside of cotyledons of cotton variety Zhong 9807 plants. After injection, the plants were placed in the dark overnight, and a 16 h light/8 h dark cycle was performed at 25 °C. The plants injected with pYL156 and pYL156: *GhZAT67* were treated with alkaline treatment after the cotton plants with pYL156: PDS developed an albino phenotype.

Results

Identification of ZAT proteins

To identify the *ZAT* genes in four cotton species, the protein sequences and genome sequences of *G. arboreum*, *G. raimondii*, *G. hirsutum,* and *G. barbadense* were compared using HMMER 3.0 software by using PF13912 as a query condition. Redundant sequences were detected and deleted by manual, 115, 115, 59, and 56 genes were identified in *G. hirsutum*, *G. barbadense*, *G. arboretum*, and *G. raimondii*, respectively. Statistical analysis of the number of genes in four *Gossypium* species showed that the number of *ZAT* genes in two tetraploid cotton species (*G. hirsutum* and *G. barbadense*) were almost double of their progenitors (*G. arboretum* and *G. raimondii*) (Fig. 1), and their numbers were also much higher than many of other plant species like *V. vinifera* (17), *T. cacao* (24) and *O. sativa* (46), indicating that the *ZAT* gene family in cotton had exhibited large scale expansion during the evolution.

To understand the *ZAT* gene family more conveniently, the *ZAT* genes were renamed according to the position of the gene on the cotton chromosome. For example, we assigned the name to ZAT genes as *GhZAT1-GhZAT115* for *G. hirsutum*, *GbZAT1-GbZAT115* for *G. barbadense*, *GaZAT1-GaZAT59* for *G. arboretum*, *GrZAT1-GrZAT56* for *G. raimondii* (Supplementary Table S1). Then we predicted the physical and chemical properties of *GhZAT* genes, including protein length, protein molecular weight (MWs), isoelectric point (pI), grand average of hydropathy analysis, and subcellular location (Supplementary Table S1). For *Gossypium hirsutum*, all of the 115 genes encoded proteins ranging from 134 (*GhZAT72*) to 577 (*GhZAT105*) amino acids, with an average of 283.6 amino acids. The isoelectric point (pI) varied from 4.936 (*GhZAT85*) to 10.171 (*GhZAT37*) with a mean of 7.98, and MWs varied from 15.105 (*GhZAT72*) kDa to 63.324 (*GhZAT105*) kDa. The values of the grand average of hydropathy were all negative, which proved that the proteins of ZAT family genes were hydrophilic. Subcellular location prediction showed that there were 110 genes located in the nuclear, 3 in the extracellular, one in the mitochondrial, and one in the plasma membrane. Most *GhZAT* genes located in the nuclear, which may be consistent with their interaction with DNA (Takatsuji et al. 1992).

Phylogenetic analysis of *ZAT* genes

To understand the evolutionary relationship of *ZAT* genes among four cotton species and other plant species, phylogenetic analysis of 662 ZAT protein sequences (Fig. 1) was performed to construct the tree based on multiple sequence alignment using the neighbor-joining (NJ) method in MEGA 7 with 1 000 bootstrap replicates, including 115 in *G. hirsutum*, 115 in *G. barbadense*, 59 in *G. arboretum*, 56 in *G. raimondii*, 42 in *A. thaliana*, 83 in *G. max*, 46 in *O. sativa*, 46 in *P. trichocarpa*, 24 in *T. cacao*, 17 in *V. vinifera* and 59 in *Z. mays* (Fig. 2). According to the number of domains in the genes, we divided the ZAT *ZAT* genes into 5 clades, and each clade can be divided into different classes (Fig. 2B). Among them, there were 2 classes in clade I, 6 in clade II, 2 in clade III, 1 in clade IV, and one in clade V (Fig. 2B). Among them, clade II had the most subgroups. As

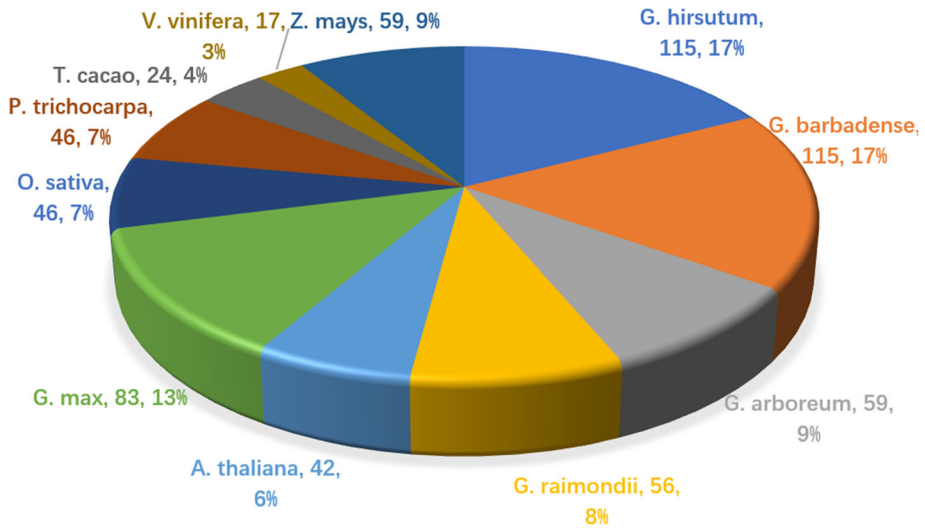

Fig. 1 Distribution of *ZAT* genes among eleven plant species

shown in the phylogenetic tree, we found that the gene distribution in subfamily Ia was the most extensive. There were 25 *GhZAT* genes in subfamily Ia, accounting for 21.7% of all *GhZAT* genes. The sequence similarity of the proteins between cotton and *T. cacao* was higher than that of any other plant species, which was consistent with a previous study that cotton and *T. cacao* were originated from the same ancestor. All species had gene pairs derived from the same node, demonstrating that the *ZAT* genes in all species had experienced gene duplication events that causing the expansion of the *ZAT* gene family. However, gene duplication among different groups varies from each other in different species. These results indicated that gene duplication was the main contributor toward *ZAT* gene family expansion in all plant species.

In addition, in order to study the relationship between the common ancestors of diploid cotton (*G. arboretum* and *G. raimondii*) and allotetraploid cotton (*G. hirsutum* and *G. barbadense*), we constructed NJ tree of four

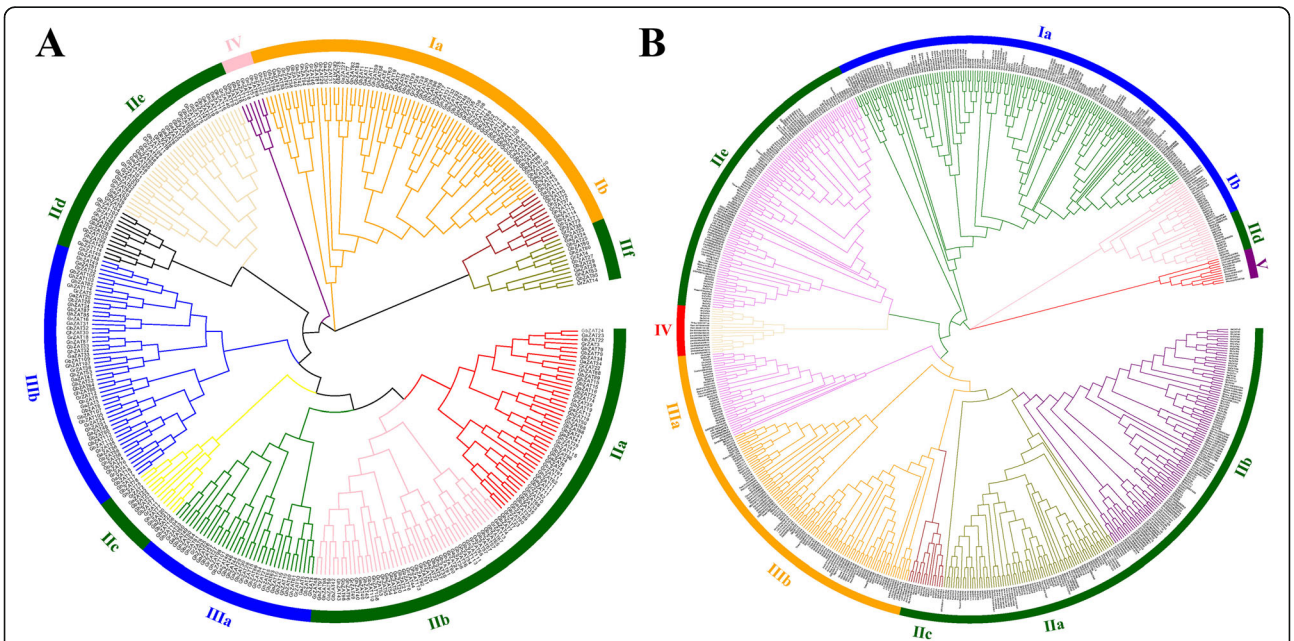

Fig. 2 Two rooted phylogeny trees constructed using MEGA-7 by the Neighbor-Joining (NJ) method. **A:** Phylogenetic relationship of the 384 identified *ZAT* genes from four cotton species; **B:** Phylogenetic relationship of the 662 identified *ZAT* genes from four cotton species and other seven species

cotton species (Fig. 2A). In general, 345 *ZAT* genes in diploid cotton and allotetraploid cotton can be divided into four clades according to their number of domains. We can see that each branch contained genes from diploid and allotetraploid cotton species (Fig. 2A).

Analysis of chromosomal localization

To further study the genetic divergence and duplication events of the *ZAT* gene family, we constructed chromosomal maps of 345 *ZAT* genes from *G. hirsutum*, *G. barbadense*, *G. arboretum* and *G. raimondii* (Fig. 3). 334 out of 345 *ZAT* genes were distributed to their specific chromosomes, while the remaining only 11 *ZATs* were allocated to the unmapped scaffold. The distribution of *ZAT* genes on 13 chromosomes of different cotton species was uneven, and the number of each chromosome did not show a significant positive correlation with its length.

Among 115 identified *ZAT* genes of *G. hirsutum*, 56 genes were distributed on 12 chromosomes of At-subgenome (GhAt) while two genes were found at the scaffold, 53 genes were positioned on 12 chromosomes of Dt-subgenome (GhDt) while four genes at the scaffold (Fig. 3A, B). In *G. hirsutum*, Chr5, Chr7, Chr8, Chr10, Chr12 in GhAt had relatively more *ZAT* genes, including 7, 7, 6, 6, and 7 genes, respectively; while Chr5, Chr7, Chr8, Chr9, Chr10 in GhDt had relatively more *ZAT* genes, including 5, 6, 5, 5, 5 and 7 genes, respectively (Fig. 4). Furthermore, the number of genes located on the chromosomes of GhAt was not the same as that on its homologous chromosome of GhDt except GhA01-GhD01, GhA11-GhD11, GhA12-GhD12, GhA13-GhD13. Interestingly, we found that there was no *GhZAT* gene on Chr6, which may be related to the chromosome deletion of *G. hirsutum* or the translocation of large fragments during the evolution. Similarly, the number of genes located on the chromosomes of GhAt/GhDt subgenomes as compared with their homologous chromosome of A/D genome (*G. arboreum*/*G. raimondii*) was almost the same except GhAt01-A01, GhAt02-A02, GhAt03-A03, GhAt07-A07, GhAt08-A08, GhAt11-A11, GhDt03-D03, implied that some *ZAT* genes had lost or added during the evolution. At the same time, the uneven and random distribution of genes indicated that gene loss may occur during the evolutionary process, but it may also be the result of incomplete genome assembly.

Among 115 identified *ZAT* genes of *G. barbadense*, 57 genes were distributed on 13 chromosomes of the At-subgenome (GbAt) while only one gene was found at the scaffold, 55 *ZAT* genes were positioned on 13 chromosomes of Dt-subgenome (GbDt) while two genes were found at the scaffold (Fig. 3C, D). In *G. barbadense*, Chr5, Chr7, Chr10, and Chr12 in GbAt had relatively more *ZAT* genes, including 7, 7, 7, and 6 genes, respectively; while Chr7, Chr10, Chr12 in GbDt had relatively more *ZAT* genes, including 7, 7, 6 genes, respectively (Fig. 4). In addition, the number of genes located on the chromosomes of GbAt was not the same as that on its homologous chromosome of GbDt except GbA01-GbD01, GbA02-GbD02, GbA06-GbD06, GbA07-GbD07, GbA09-GbD09, GbA10-GbD10. We also found the number of genes on the GbAt/GbDt subgenome with the homologous chromosome A/D genome (*G. arboreum*/*G. raimondii*) was not the same, except for GbAt01-A01, GbAt02-A02, GbAt03-A03, Gbt06-A06, Gbt07-A07, GbAt09-A09, GhAt12-A12, and GhDt03-D03, which was similar to the situation of *G. hirsutum*.

Among 59 identified ZAT genes of *G. arboreum*, 57 genes were distributed on 13 chromosomes while two genes were found at the scaffold (Fig. 3E). Among 56 identified *ZAT* genes of *G. raimondii*, 56 genes were distributed on 13 chromosomes while no gene was found at the scaffold (Fig. 3F). In *G. arboreum*, Chr7, Chr8, Chr12 had relatively more *ZAT* genes, including 7, 6, 6 genes, respectively; while Chr1, Chr8, Chr9 in *G. raimondii* had relatively more *ZAT* genes, including 7, 7, 6 genes, respectively (Fig. 4).

Analysis of conserved protein motifs and gene structure

Diversity of gene structure and differentiation of conserved motifs were possible mechanisms for the evolution of polygenic families (Hu et al. 2010). We identified the conserved motifs of *ZAT* protein using the online website MEME and found 10 conserved motifs (named motif 1 to motif 10), the results were represented in schematic diagrams (Fig. 5B). ZAT genes in the same clade had similar motif composition, and the different composition of motifs indicating the diversity of gene function, which was particularly obvious in the class IIa and class IIb subfamily. Similar motif arrangements in the same subgroup revealed that protein structure was conserved in a specific subgroup, and the functions of most conserved motifs remain to be clarified. There were 1 ~ 11 motifs in each *GhZAT* gene, and all *GhZAT* genes had a conservative protein motif distribution. Except for the genes of subclasses Ia and Ib, all others had a conserved motif 3. Interestingly, in addition to the *GhZAT93* gene, the genes of subclass Ia contained a special motif 8, while the other subclasses did not have, which may be the reason that subclass Ia had obtained the special motif 8 through evolution or other subfamilies had lacked specific conserved motifs during evolution.

In general, genes in the same subclass had similar intron-exon arrangements in terms of intron number and exon length (Malik et al. 2020). As can be seen from Fig. 5C, most genes had no introns except for a few genes. The exon length of the sequence encoded by the

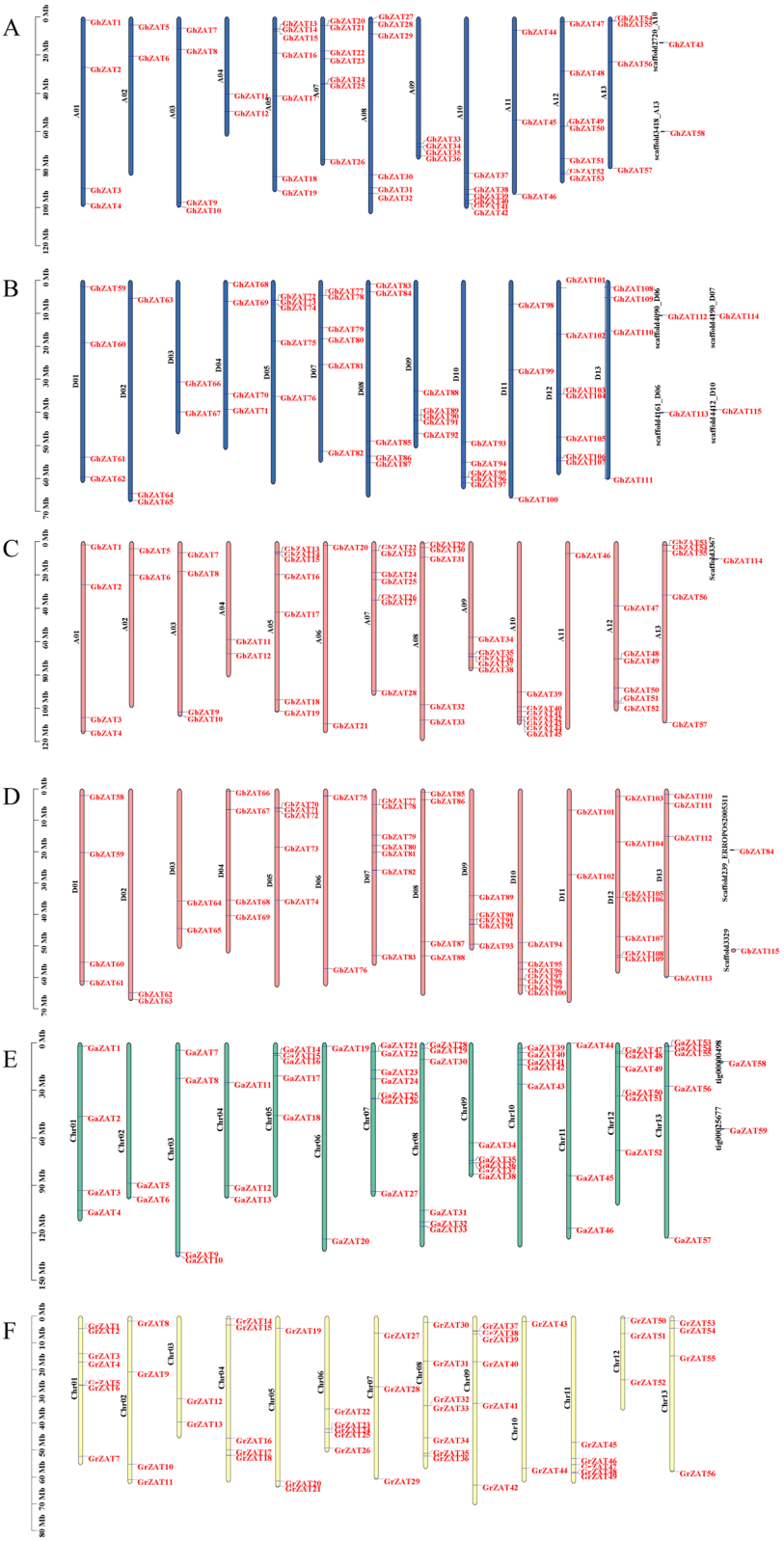

Fig. 3 (See legend on next page.)

(See figure on previous page.)
Fig. 3 Chromosomal location of *ZAT* genes from four *Gossypium* species. **A:** Location of *ZAT* genes on chromosomes in *G. hirsutum* At sub-genome (GhAt); **B:** Location of *ZAT* genes on chromosomes in *G. hirsutum* Dt sub-genome (GhDt); **C:** Location of *ZAT* genes on chromosomes in *G. barbadense* At sub-genome (GbAt); **D:** Location of *ZAT* genes on chromosomes in *G. barbadense* Dt sub-genome (GbDt); **E:** Location of *ZAT* genes on chromosomes in *G. arboreum* (Ga); **F:** Location of *ZAT* genes on chromosomes in *G. raimondii* (Gr). Note: The gene ID on the right side of each chromosome corresponds to the approximate location of each *ZAT* gene

gene had little difference, indicating that the genes were somewhat conserved.

In addition, 8 genes had only one intron, one gene had three introns, and many genes (106 of 115) had no introns. Fewer introns in the *ZAT* gene family indicated that *GhZAT* genes were less likely to undergo alternative splicing. Exon/intron analysis along with phylogenetic tree results displayed their random distribution among

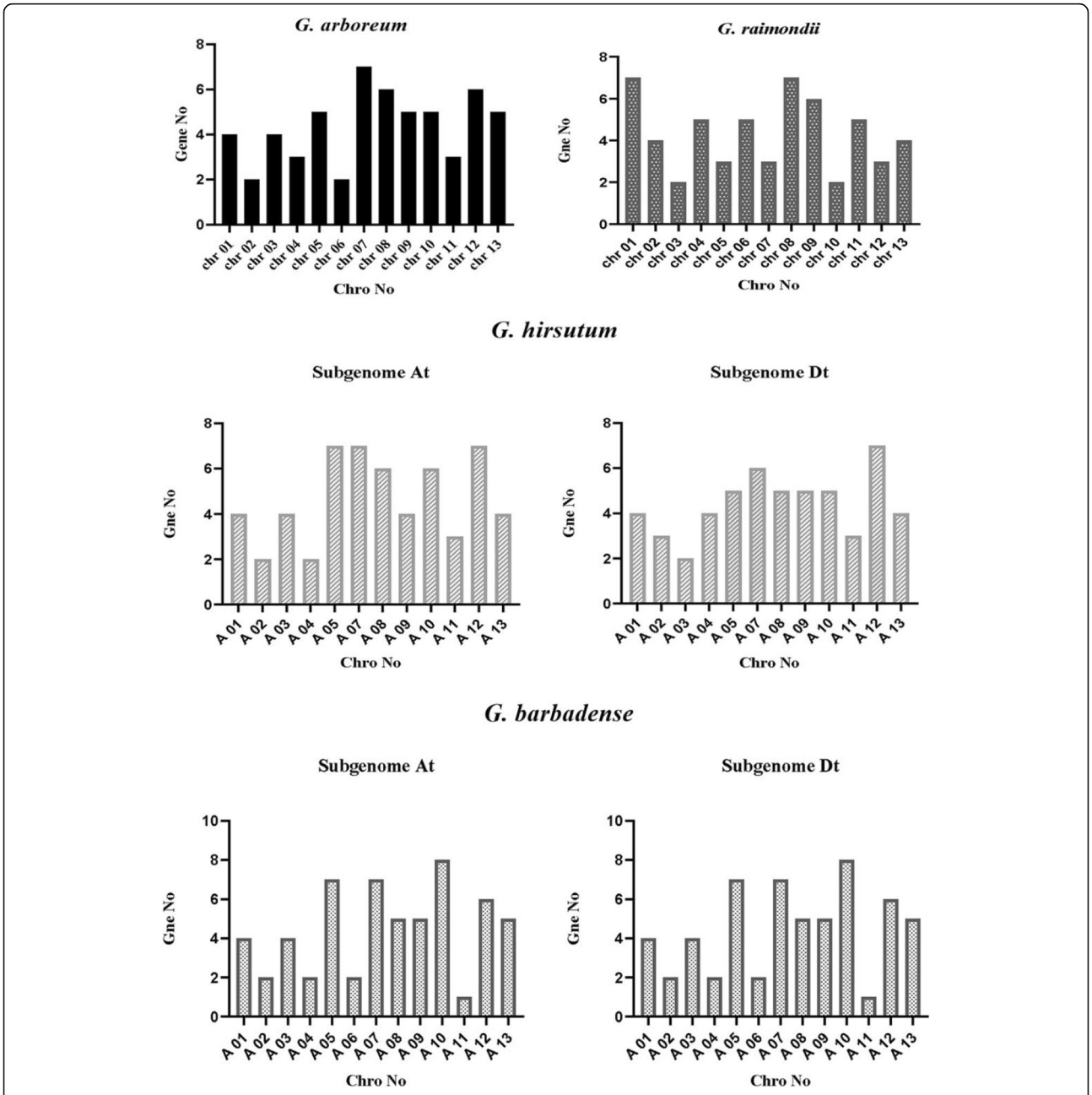

Fig. 4 Chromosomal distribution of upland cotton *ZAT* family. Note: The chromosome location is referred to the gff3 cotton genome dataset

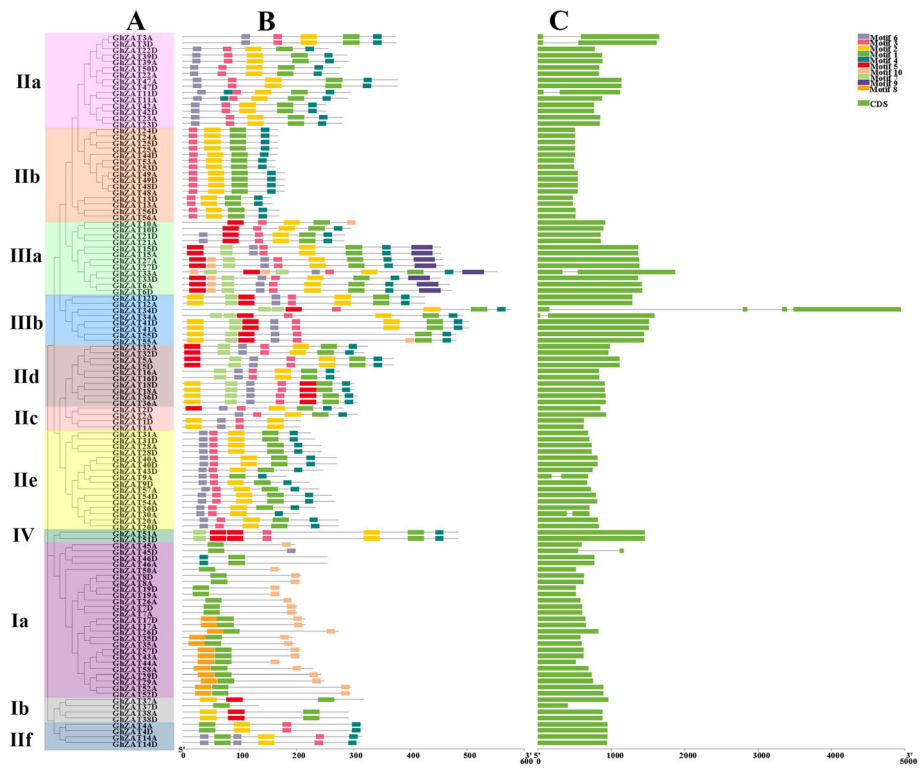

Fig. 5 Phylogenetic analysis, conserved motifs, and gene structures of *GhZAT* genes. **A:** Phylogenetic tree of *GhZAT* genes; **B:** Conserved motifs of *GhZAT* genes; **C:** Gene structures of *GhZAT* genes

GhZAT genes of different clades and showed that similar genes were clustered close to one another in the same group of phylogeny trees. It was worth noting that closely related genes were more similar in gene structure, but not necessarily the same in exon/intron length.

Gene duplication and collinearity analysis

Previous studies had shown that gene duplication was one of the major drivers of the evolution of genomes and genetic systems (Moore and Purugganan 2003). Whole-genome duplication, segmental duplication, and tandem duplication were considered to be the main causes of plant gene family expansion (Xu et al. 2012). Segmental duplication produced multiple genes by polyploidy and chromosomal rearrangement (Yu et al. 2005), and it occurred in plants because most plants were diploid or polyploid and retained many duplicate chromosome blocks in their genomes (Cannon et al. 2004). Tandem duplication was characterized by multiple members of a family in the same or adjacent intergenic region (Ramamoorthy et al. 2008). A gene cluster usually contained two or more genes of chromosome regions in a 200 kb region (Holub 2001). Gene duplication can be large-scale whole-genome duplication (WGD), and can also be a small series of segmental duplication and tandem duplication (Ramsey and Schemske 1998). In this study, we performed to analyze the gene duplication and syntenic relationship from *G. hirsutum*, *G. barbadense*, *G. arboreum*, and *G. raimondii* using MCScanX (Fig. 6). We used the MCScanX software to analyze the gene pairs of tandem duplication, segmental duplication, and whole-genome duplication. The genes which lied on the same chromosomal block (e-value <1e−5) were categorized as tandem duplicated while the remaining from the same genomes were considered as segmental and else of all from different genomes and sub-genomes of the four *Gossypium* species were allocated in the whole-genome duplication (Malik et al. 2020). A total of 1 629 gene pairs in the four cotton species were identified as duplicated genes pairs, of which 19 pairs were tandem duplication, 386 pairs were segmental duplication, and 1 224 pairs were whole-genome duplication (Supplementary Table S2). A total of 19 tandem repeat gene pairs were identified in the four cotton species, of which 2 were Ga-Ga (*GaZAT21/22*, *GaZAT36/37*), six were Gh-Gh (*GhZAT20/21*, *GhZAT34/35*, *GhZAT41/42*, *GhZAT77/78*, *GhZAT90/91*, *GhZAT95/96*), 8 were Gb-Gb (*GbZAT22/23*, *GbZAT36/37*, *GbZAT42/43*, *GbZAT44/45*, *GbZAT77/78*, *GbZAT91/92*, *GbZAT97/98*, *GbZAT99/100*), three were Gr-Gr (*GrZAT1/2*, *GrZAT24/25*, *GrZAT48/49*). Therefore, tandem duplication played an indispensable role in the evolution of the *ZAT* gene family.

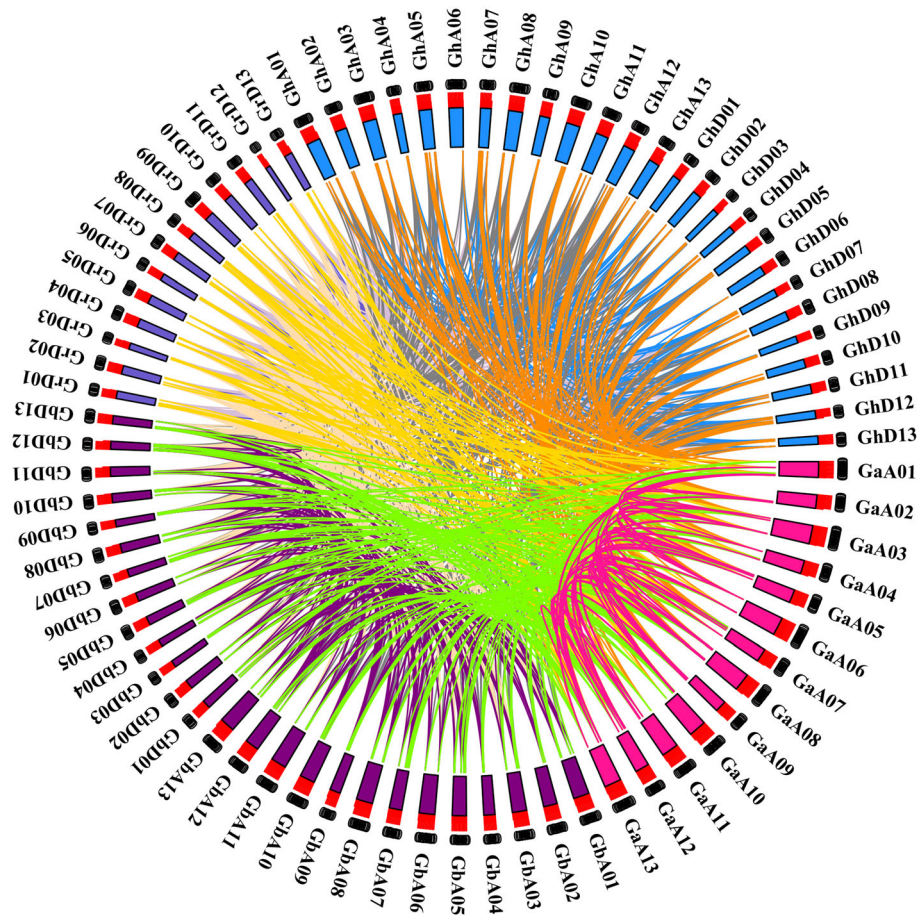

Fig. 6 Syntenic relationship of duplicated *ZAT* genes pairs from four *Gossypium* species (*G. hirsutum*, *G. barbadense*, *G. arboreum*, and *G. raimondii*)

In addition to tandem duplication, other replication mechanisms may also contribute to the extension of genes. It was suggested that parallel homologous genes were usually produced by segmental duplication and the whole-genome duplication before polyploidy, especially segmental duplication was the most significant factor in evolution (Malik et al. 2020). Segmental duplication of the *ZAT* genes in four cotton species was analyzed and 386 gene pairs were identified. There were 49, 150, 152, and 35 gene pairs among Ga-Ga, Gh-Gh, Gb-Gb, and Gr-Gr, respectively. In addition, there were 185, 241, 250, 254, 190, and 104 gene pairs between Gh-Ga, Gh-Gr, Gh-Gb, Gb-Gr, Gb-Ga, and Ga-Gr, respectively. From these results, we presumed that tandem duplication, segmental duplication, and whole-genome duplication had an impartment effect on the evolution of the *ZAT* gene family.

Calculation of non-synonymous (*Ka*) to synonymous (*Ks*) substitution rates

To investigate the differentiation mechanism of *ZAT* genes after polyploid duplication in cotton, *Ka*, *Ks*, and *Ka*/*Ks* of 1 332 homologous gene pairs from 10 combinations of four *Gossypium* species were determined, including Ga-Ga, Ga-Gb, Ga-Gr, Gb-Gb, Gb-Gr, Gh-Gh, Gh-Ga, Gh-Gb, Gh-Gr and Gr-Gr (Supplementary Table S3). *Ka*/*Ks* ratio was used to infer the selection pressure of duplicated gene pairs. $Ka/Ks < 1$ was considered as a purification selection, indicating that natural selection eliminated harmful mutations and kept the protein unchanged, $Ka/Ks > 1$ was considered as positive selection, indicating natural selection changed the protein, the mutation site was quickly fixed in the population, and the evolution of the gene was accelerated, $Ka/Ks = 1$ was considered as neutral selection, indicating that natural selection did not affect mutation (Hurst 2002).

Over all, we found 1 332 duplicated gene pairs in this analysis from four *Gossypium* species (Ga, Gr, Gh, and Gb). Among them, 1 288 (96.7%) duplicated gene pairs with *Ka*/*Ks* ratio < 1 including 1 045 gene pairs with *Ka*/*Ks* < 0.5 and 243 gene pairs between 0.5 and 0.99, exhibiting pure selection. Only 44 (3.3%) duplicated gene pairs with *Ka*/*Ks* > 1 and these gene pairs might have experienced relatively rapid evolution following duplication

and underwent positive selection pressure. Studies had shown that sub-functional duplicated genes generally differentiated at the transcriptional level first and tended to be subject to strong purification (Roulin et al. 2013). In general, gene duplication will produce two copies, on the one hand, under low selective pressure, the gene will evolve, on the other hand, it will produce some new functions, which can contribute to the evolution of the species.

Analysis of promoters and differentially expressed genes

To further understand the response of *GhZAT* genes to abiotic stress, we analyzed their expression under cold, heat, salt, and PEG stress using RNA-seq data and *cis*-acting elements of promoters under abiotic stress. Considering the importance of *cis*-acting elements in abiotic stress, we selected the region of 2 kb upstream of the start codon of the *ZAT* genes, then we used PlantCARE to study the *cis*-acting elements that may be involved in gene expression, and extracted the *GhZAT* gene promoter region which was associated with stress and plant hormone. In this study, most *GhZAT* genes in *G. hirsutum* contained *cis*-acting elements related to plant hormones (ABA, MeJA, GA, IAA) and various stresses (low temperature, light stress, and wound stress) (Fig. 7B), and genes in the same subfamily were functionally diverse despite the similar motifs. In terms of plant hormone response, we found the *cis*-acting elements including GARE-motif (gibberellin responsive elements), ERE (ethylene response), ABRE (involved in abscisic acid responsiveness), AuxRR-core (*cis*-acting regulatory elements involved in auxin responsiveness), and CGTCA-motif (*cis*-acting regulatory element involved in MeJA responsiveness). Almost all the promoters were found to contain several hormones responsive elements except *GhZAT25*, *GhZAT23*, *GhZAT100*, but we did not find

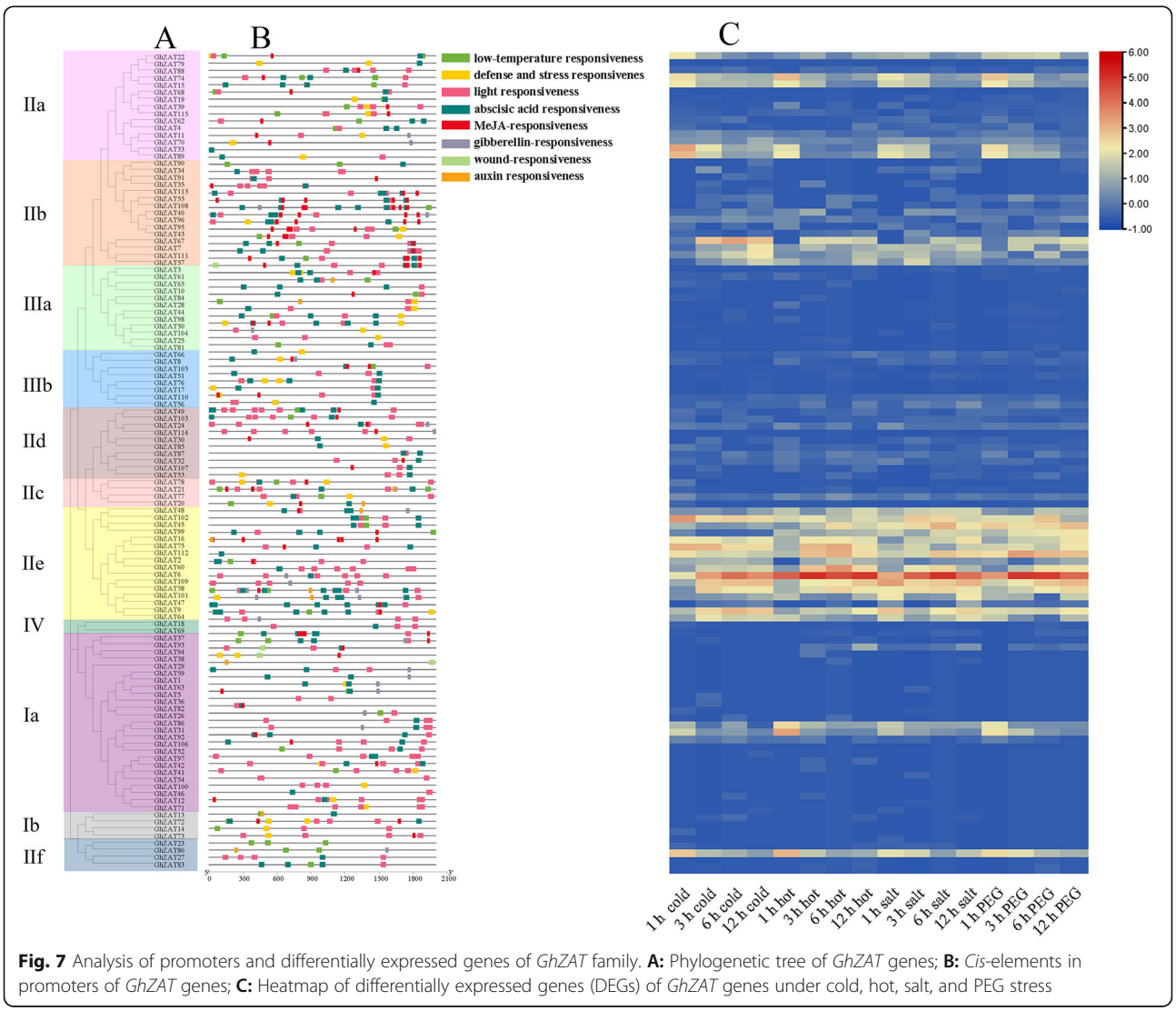

Fig. 7 Analysis of promoters and differentially expressed genes of *GhZAT* family. **A:** Phylogenetic tree of *GhZAT* genes; **B:** *Cis*-elements in promoters of *GhZAT* genes; **C:** Heatmap of differentially expressed genes (DEGs) of *GhZAT* genes under cold, hot, salt, and PEG stress

any close relationship of hormone-responsive elements with their subfamily. Therefore, *GhZAT* genes may be involved in regulating the growth and defense responses of cotton through specific hormonal signaling pathways.

Gene expression patterns can provide important references for functional analysis of genes, and gene expression was related to biological functions controlled by *cis*-acting elements. We detected the expression pattern of *GhZAT* genes in different times (1 h, 3 h, 6 h, 12 h) under different stresses (cold, hot, salt, and PEG) by analyzing RNA-seq data (Fig. 7C). The results showed that the genes of the IIb and IIe subfamily all responded to cold, heat, salt, and PEG stress, while the expression levels of other subfamily genes also changed, but the expression levels were not as high as these two types. The gene expression patterns of the same subfamily were almost the same, but the gene expression patterns of the same subfamily were slightly different. These results proved that *GhZAT* genes were involved in plant stress response.

Cotton plants with *GhZAT* gene silenced by VIGS were sensitive to alkaline stress

We also explored whether *ZAT* genes responded to alkaline stress. We selected a high-expressed gene *GhD03G1247.1* (*GhZAT67*) from transcriptome data treated with alkaline stress (125 mmol·L^{-1} $NaHCO_3$ treated for 12 h) for further study. VIGS vector pYL156: *GhZAT67* was constructed using In-Fusion technology to study their functions under $NaHCO_3$ stress. Afterward, the cotton carrying pYL156: PDS appeared albino, indicating that VIGS silencing was successful (Fig. 8A). qRT-PCR was used to detect the gene silencing effect, and the results showed that the *GhZAT67* expression level of pYL156: *GhZAT67* was significantly lower than that of the control pYL156 (Fig. 8B). After treatment with 125 mmol·L^{-1} $NaHCO_3$, we found that the pYL156: *GhZAT67* cotton seedlings wilted more severely than the pYL156. Therefore, *GhZAT67* was also involved in the adaptability of cotton to alkaline stress.

Discussion

Cotton was an important economic crop, widely distributed around the world, faced serious biotic and abiotic stress. However, the understanding of the functional role of *ZAT* genes in cotton was still very limited. In recent years, higher-quality *de-novo*-assembled genomes of two allotetraploid species (Hu et al. 2019) and *G. arboretum* (Du et al. 2018) had been published, which will open a new era of genome-wide analysis of cotton gene families.

In this study, we conducted a comprehensive identification and analysis of the *ZAT* genes of *G. hirsutum*, *G. barbadense*, *G. arboretum*, *G. raimondii* and seven other species. Among them, we identified 115, 115, 59, and 56 genes in *G. hirsutum*, *G. barbadense*, *G. arboretum*, *G. raimondii*, respectively, 42 genes in *T. cacao*, 24 genes in *Z. mays*, 59 genes in *V. vinifera*, 17 genes in *P. trichocarpa*, 46 genes in *G. max*, and 83 genes in *A. thaliana*. Previous research had shown that allotetraploid cotton was formed by inter-genomic hybridization of A-genome diploids and D-genome diploids, and followed by chromosome doubling (Paterson et al. 2012). And in our research, the number of *ZAT* genes in *G. hirsutum* and *G. barbadense* was the sum of the number of those in *G. arboretum* and *G. raimondii*. The subcellular localization results showed that about 96% (110 of 115) of *ZAT* genes were localized in the nucleus, while the other 5 genes were predicted to be located in the extracellular, mitochondrial and plasma membrane. This indicated that these genes had special characteristics compared with other members of the family. In addition, based on the analysis of WOLF PSORT software, we can roughly determine their location, but for a more accurate location, further experimental verification was needed.

In this study, we divided the 345 *ZAT* genes from four *Gossypium* species into four clades according to the number of domains. Each clade contained different subfamilies, and clade II contained the largest number of genes. This proved that most genes of the *ZAT* family were conserved in the evolution, and most of them retained two main domains. However, genes containing only one domain also appeared in the class II subfamily, such as *GhZAT47* and *GhZAT2*. Through the analysis of

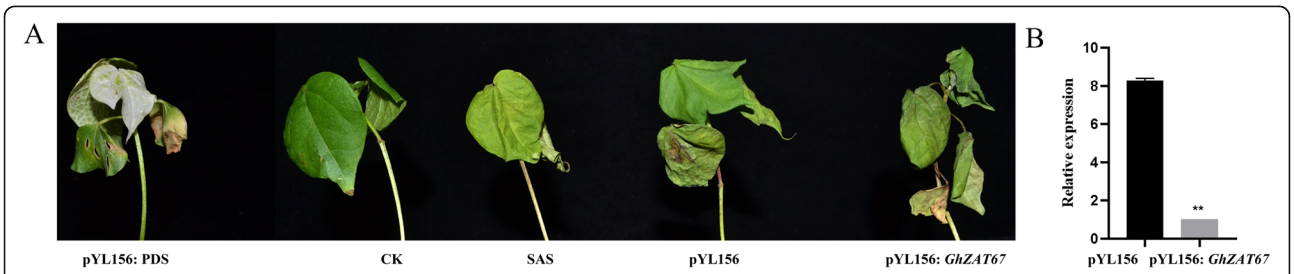

Fig. 8 The phenotype of cotton leaves after virus infection and expression analysis of *GhZAT67* under $NaHCO_3$ stress. **A:** Phenotype of cotton leaves after virus infection; **B:** qRT-PCR for *GhZAT67* under $NaHCO_3$ stress. Note: CK: No infection; SAS: WT under $NaHCO_3$ stress

the motifs of these genes, it was found that they were one motif less than the motifs of genes in the same subfamily. Genes that also contained 3 domains, such as *GhZAT78* and *GhZAT67*, were found to have one more motif compared with other members of the same subfamily. It was worth noting that these genes may have new functions or miss some functions in the process of evolution.

Phylogenetic analysis showed that the ZAT proteins in *Gossypium* were more closely related to the ZAT protein in *T. cacao*, which was consistent with a previous study that *Gossypium* and *T. cacao* came from the same ancestor (Li et al. 2015). Most *ZAT* genes in the same subfamily had similar intron and exon structures and conserved motifs, but there was a high degree of differentiation among different subfamilies. In addition, some exon-intron structure and motif composition mainly existed in a specific subgroup, which may be related to the functional diversity of this subgroup, such as both *GhZAT18* and *GhZAT69* had two motif 5, and this can be used as an important indicator to identify this subfamily.

Gene duplication produced functional differentiation, which was of great significance to environmental adaptability and speciation (Conant and Wolfe 2008). According to this study, the gene replication rate of plants was significantly higher than that of other eukaryotes (Hanada et al. 2008). Previous studies had proved that there were segmental duplication and tandem duplication in various plant transcription factor families such as Fbox, bZIP, and NAC factors (Jain et al. 2007; Puranik, et al. 2012; Wei et al. 2012). Therefore, we analyzed the gene duplication of ZAT transcription factor families. To reveal the expansion mechanism of the *ZAT* gene family, we used MCScanX to screen duplicated gene pairs, according to the results of the analysis, there were only a few tandem-duplicated events in both *G. hirsutum* and *G. barbadense*, most of which were segmental duplications, which indicated that segmental duplication played a vital role in the evolution of the *ZAT* gene family. This result was also consistent with previous studies (Wei et al. 2016). The number of homologous genes in the At and Dt subgenomes of *G. hirsutum* and *G. barbadense* were close, so we inferred that translocation and reverse transcript insertion rarely occurred in the *ZAT* gene family. In order to elucidate the differences after gene duplication, we calculated the ratio of non-synonymous (*Ka*) to synonymous (*Ks*) of four *Gossypium* species. We found the $Ka/Ks < 1$ in the four *Gossypium* species, indicating that the four *Gossypium* species had undergone strong purification selection. This finding was consistent with a recent study that MYB transcription factor genes were mostly evolved under negative selection (Salih et al. 2016).

ZAT proteins played an important role in the growth and development of plants. Some *ZAT* genes had been confirmed to act as an important role in the tolerance increasing to stress in *Arabidopsis*. *Cis*-acting elements played an important role when plants were subjected to abiotic stress (Nakashima et al. 2014). In this study, the difference in the *cis*-acting elements of the duplicated genes played a direct role in their expression and differentiation. The *cis*-acting element responded to the elicitor to regulate gene expression. Based on a large number of hormone response elements (salicylic acid, jasmonic acid, ethylene, and abscisic acid) in the promoter of *GhZAT* genes, we believed that plant hormones may be involved in the regulation of upstream of *GhZAT* genes. Studies had found that salicylic acid, jasmonic acid, ethylene, abscisic acid and other plant hormones interacted with corresponding *cis*-acting elements by inducing transcription factors, which were of great significance for plants to adapt to abiotic stress. This was consistent with our results (Fujita et al. 2006; Santner and Estelle 2009). ABA-responsive elements were widely distributed in promoters, we induced that ABA may be a key signal regulating the *GhZAT* gene family. Through the analysis of the *cis*-acting elements of *G. hirsutum*, we speculated that the ZAT gene of upland cotton was related to plant hormone response and stress resistance.

When subjected to abiotic stress, cotton will produce a series of physiological and biochemical reactions to resist the stress of adversity. Abiotic stresses such as drought, cold, heat, and salt stress will decrease the yield of cotton. The expression pattern of genes was often used as an indicator of their functions. This study used the published transcriptome data to study the expression profile of *GhZAT* genes under abiotic stress. The expression of *ZAT* genes in the IIe and IIb subfamily of *G. hirsutum* after cold, heat, salt, and PEG treatments was quite different. Among them, *GhZAT67* had the highest expression level at 6 h and 12 h after cold stress, *GhZAT7* and *GhZAT6* were both expressed in every treatments, *GhZAT111* and *GhZAT57* were highly expressed after cold stress for 12 h, *GhZAT109*, *GhZAT75*, *GhZAT102*, *GhZAT58*, *GhZAT9*, *GhZAT64* all played an important role in cold stress. Therefore, *GhZAT* genes may be involved in the adaptation of cotton to various stresses, especially cold stress. In summary, this study analyzed the gene structure, chromosome distribution, phylogenetic relationship, *cis*-acting elements, and collinearity relationship of *GhZAT* genes in *G. hirsutum*, which greatly enriched our understanding of the *ZAT* gene family. In addition, the expression profile and *cis*-acting element analysis of *GhZAT* genes indicated that they may be involved in plant hormone signal transduction and abiotic stress response, especially under cold stress conditions. These findings not only laid the foundation for further analysis of the function of cotton *ZAT* genes but also provided valuable information for potential candidate genes related to plant growth and development and adversity stress.

Conclusions

In conclusion, this study conducted a comprehensive analysis of the *ZAT* gene family of four kinds of cotton for the first time. We confirmed that *ZAT* subfamilies displayed a great evolutionary divergence by phylogenetic analysis, conserved motifs, and gene structure. *Cis*-acting elements analysis and expression profiles under environmental stresses suggested that *ZAT* family genes might be related to the responses of abiotic stresses and hormone stimuli. In addition, an alkaline-stress-induced gene *GhZAT67* was cloned and we found it positively responded to alkaline stress. Taken together, we established a foundation for further functional characterization of *ZAT* genes in response to abiotic stress in the future.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42397-021-00098-0.

Additional file 1.

Additional file 2.

Additional file 3.

Acknowledgments

We would like to thank the anonymous reviewers for their valuable comments and helpful suggestions which help to improve the manuscript.

Authors' contributions

Formal analysis, Fan YP; Investigation, Fan YP; Software, Zhang YX, Rui C, Xu N, Zhang H, Wang J, Malik WA, Han MG, Zhao LJ, Lu XK, Chen XG, and Chen C; Validation, Fan YP, Zhang YX, Rui C, Xu N, and Zhang H; Visualization, Fan YP, Zhang YX, and Rui C; Original draft, Fan YP; Reviewing and editing, Ye WW. The authors read and approved the final manuscript.

Funding

This work was supported by Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences. The funding bodies provided financial support to the research projects but didn't involve in study design, data collection, analysis, or preparation of the manuscript.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflict of interest.

Received: 15 February 2021 Accepted: 25 June 2021
Published online: 13 August 2021

References

Bailey TL, Boden M, Buske FA, et al. MEME suite: tools for motif discovery and searching. Nucleic Acid Res. 2009;37(suppl_2):W202–8. https://doi.org/10.1093/nar/gkp335.

Cannon SB, Mitra A, Baumgarten A, et al. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biol. 2004;4(1):1–21. https://doi.org/10.1186/1471-2229-4-10.

Chen C, Chen H, He Y, et al. TBtools, a toolkit for biologists integrating various biological data handling tools with a user-friendly interface. Molecular Plant. 2020;13(8):1194–202. https://doi.org/10.1016/j.molp.2020.06.009.

Ciftci-Yilmaz S, Morsy MR, Song L, et al. The EAR-motif of the Cys_2/His_2-type zinc finger protein Zat7 plays a key role in the defense response of *Arabidopsis* to salinity stress. J Biol Chem. 2007;282(12):9260–8. https://doi.org/10.1074/jbc.M611093200 .

Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet. 2008;9(12):938–50. https://doi.org/10.1038/nrg2482.

Davletova S, Schlauch K, Coutu J, Mittler R. The zinc-finger protein Zat12 plays a central role in reactive oxygen and abiotic stress signaling in *Arabidopsis*. Plant Physiol. 2005;139(2):847–56. https://doi.org/10.1104/pp.105.068254.

Du X, Huang G, He S, et al. Resequencing of 243 diploid cotton accessions based on an updated a genome identifies the genetic basis of key agronomic traits. Nat Genet. 2018;50(6):796–802. https://doi.org/10.1038/s41588-018-0116-x.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol. 2000;300(4):1005–16. https://doi.org/10.1006/jmbi.2000.3903.

Fujita M, Fujita Y, Noutoshi Y, et al. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Curr Opin Plant Biol. 2006;9(4):436–42. https://doi.org/10.1016/j.pbi.2006.05.014.

Hanada K, Zou C, Lehti-Shiu MD, et al. Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiol. 2008;148(2):993–1003. https://doi.org/10.1104/pp.108.122457 .

He F, Niu MX, Feng CH, et al. *PeSTZ1* confers salt stress tolerance by scavenging the accumulation of ROS through regulating the expression of *PeZAT12* and *PeAPX2* in *Populus*. Tree Physiol. 2020;40(9):1292–311. https://doi.org/10.1093/treephys/tpaa050 .

Holub EB. The arms race is ancient history in *Arabidopsis*, the *wildflower*. Nat Rev Genet. 2001;2(7):516–27. https://doi.org/10.1038/35080508.

Hu R, Qi G, Kong Y, et al. Comprehensive analysis of NAC domain transcription factor gene family in *Populus trichocarpa*. BMC Plant Biol. 2010;10(1):1–23. https://doi.org/10.1186/1471-2229-10-145.

Hu Y, Chen J, Fang L, et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. Nat Genet. 2019;51(4):739–48. https://doi.org/10.1038/s41588-019-0371-5 .

Hurst LD. The *Ka/Ks* ratio: diagnosing the form of sequence evolution. Trends in Genetics. 2002;18(9):486–7. https://doi.org/10.1016/s0168-9525(02)02722-1 .

Jain M, Nijhawan A, Arora R, et al. F-box proteins in *rice*. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. Plant Physiol. 2007;143(4):1467–83. https://doi.org/10.1104/pp.106.091900.

Kiełbowicz-Matuk A. Involvement of plant C_2H_2-type zinc finger transcription factors in stress responses. Plant Sci. 2012;185:78–85. https://doi.org/10.1016/j.plantsci.2011.11.015.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45. https://doi.org/10.1101/gr.092759.109.

Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4. https://doi.org/10.1093/molbev/msw054.

Le CTT, Brumbarova T, Ivanov R, et al. Zinc finger of Arabidopsis thaliana12 (*ZAT12*) interacts with FER-like iron deficiency-induced transcription factor (*FIT*) linking iron deficiency and oxidative stress responses. Plant Physiol. 2016;170(1):540–57. https://doi.org/10.1104/pp.15.01589.

Li F, Fan G, Lu C, et al. Genome sequence of cultivated upland cotton (*Gossypium hirsutum* TM-1) provides insights into genome evolution. Nat Biotechnol. 2015;33(5):524–30. https://doi.org/10.1038/nbt.3208 .

Malik WA, Wang X, Wang X, et al. Genome-wide expression analysis suggests glutaredoxin genes response to various stresses in cotton. Int J Biol Macromol. 2020;153:470–91. https://doi.org/10.1016/j.ijbiomac.2020.03.021.

Mittler R, Kim YS, Song L, et al. Gain- and loss-of-function mutations in *Zat10* enhance the tolerance of plants to abiotic stress. FEBS Lett. 2006;580(28–29):6537–42. https://doi.org/10.1016/j.febslet.2006.11.002.

Moore RC, Purugganan MD. The early stages of duplicate gene evolution. Proc Natl Acad Sci. 2003;100(26):15682–7. https://doi.org/10.1073/pnas.2535513100.

Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. Front Plant Sci. 2014;5:170. https://doi.org/10.3389/fpls.2014.00170.

Paterson AH, Wendel JF, Gundlach H, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. Nature. 2012;492(7429):423–7. https://doi.org/10.1038/nature11798.

Puranik S, Sahu PP, Srivastava PS, Prasad M. NAC proteins: regulation and role in stress tolerance. Trends Plant Sci. 2012;17(6):369–81. https://doi.org/10.1016/j.tplants.2012.02.004.

Ramamoorthy R, Jiang SY, Kumar N, et al. A comprehensive transcriptional profiling of the *WRKY* gene family in rice under various abiotic and phytohormone treatments. Plant Cell Physiol. 2008;49(6):865–79. https://doi.org/10.1093/pcp/pcn061.

Ramsey J, Schemske DW. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu Rev Ecol Syst. 1998;29(1):467–501. https://doi.org/10.1146/annurev.ecolsys.29.1.467.

Rizhsky L, Davletova S, Liang H, Mittler R. The zinc finger protein Zat12 is required for cytosolic ascorbate peroxidase 1 expression during oxidative stress in *Arabidopsis*. J Biol Chem. 2004;279(12):11736–43. https://doi.org/10.1074/jbc.M313350200.

Roulin A, Auer PL, Libault M, et al. The fate of duplicated genes in a polyploid plant genome. Plant J. 2013;73(1):143–53. https://doi.org/10.1111/tpj.12026.

Sakamoto H, Maruyama K, Sakuma Y, et al. Arabidopsis Cys_2/His_2-type zinc-finger proteins function as transcription repressors under drought, cold, and high-salinity stress conditions. Plant Physiol. 2004;136(1):2734–46. https://doi.org/10.1104/pp.104.046599.

Salih H, Gong W, He S, et al. Genome-wide characterization and expression analysis of MYB transcription factors in *Gossypium hirsutum*. BMC Genet. 2016;17:129. https://doi.org/10.1186/s12863-016-0436-8.

Santner A, Estelle M. Recent advances and emerging trends in plant hormone signalling. Nature. 2009;459(7250):1071–8. https://doi.org/10.1038/nature08122.

Sun Z, Liu R, Guo B, et al. Ectopic expression of GmZAT4, a putative C_2H_2-type zinc finger protein, enhances PEG and NaCl stress tolerances in *Arabidopsis thaliana*. 3 Biotech. 2019;9(5):166. https://doi.org/10.1007/s13205-019-1673-0.

Takatsuji H. Zinc-finger proteins: the classical zinc finger emerges in contemporary plant science. Plant Mol Biol. 1999;39(6):1073–8. https://doi.org/10.1023/a:1006184519697.

Takatsuji H, Mori M, Benfey PN, et al. Characterization of a zinc finger DNA-binding protein expressed specifically in *Petunia* petals and seedlings. EMBO J. 1992;11(1):241–9. https://doi.org/10.1002/j.1460-2075.1992.tb05047.x

Vogel JT, Zarka DG, Van Buskirk HA, et al. Roles of the *CBF2* and *ZAT12* transcription factors in configuring the low temperature transcriptome of *Arabidopsis*. Plant J. 2005;41(2):195–211. https://doi.org/10.1111/j.1365-313X.2004.02288.x.

Wang M, Yuan J, Qin L, et al. *TaCYP81D5*, one member in a wheat cytochrome P450 gene cluster, confers salinity tolerance via reactive oxygen species scavenging. Plant Biotechnol J. 2020;18(3):791–804. https://doi.org/10.1111/pbi.13247.

Wang Y, Dou D, Wang X, et al. The PsCZF1 gene encoding a C H zinc finger protein is required for growth, development andpathogenesis in Phytophthora sojae. Microb Pathog. 2009;47(2):78–86 .https://doi.org/10.1016/j.micpath.2009.04.013 .

Wang Y, Tang H, DeBarry JD, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49. https://doi.org/10.1093/nar/gkr1293.

Wei K, Chen J, Wang Y, et al. Genome-wide analysis of *bZIP*-encoding genes in maize. DNA Res. 2012;19(6):463–76. https://doi.org/10.1093/dnares/dss026.

Wei K, Pan S, Li Y. Functional characterization of maize C_2H_2 zinc-finger gene family. Plant Mol Biol Report. 2016;34(4):761–76. https://doi.org/10.1007/s11105-015-0958-7.

Xie M, Sun J, Gong D, Kong Y. The roles of arabidopsis C1-2i subclass of C_2H_2-type zinc-finger transcription factors. Genes. 2019;10(9):653. https://doi.org/10.3390/genes10090653.

Xu G, Guo C, Shan H, Kong H. Divergence of duplicate genes in exon–intron structure. Proc Natl Acad Sci. 2012;109(4):1187–92. https://doi.org/10.1073/pnas.1109047109.

Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. Proteins. 2006;64(3):643–51. https://doi.org/10.1002/prot.21018.

Yu J, Wang J, Lin W, et al. The genomes of *Oryza sativa*: a history of duplications. PLoS Biol. 2005;3(2):e38. https://doi.org/10.1371/journal.pbio.0030038.

Zhang HB, Li Y, Wang B, Chee PW. Recent advances in cotton genomics. Int J Plant Genom. 2008;2008:742304. https://doi.org/10.1155/2008/742304.

Zhu T, Liang C, Meng Z, et al. CottonFGD: an integrated functional genomics database for cotton. BMC Plant Biol. 2017;17(1):1–9. https://doi.org/10.1186/s12870-017-1039-x.