Genome-wide identification and characterization of GATA family genes in wheat

Qian Yu
Qingdao Agricultural University
Wenxing Liu (liuwx@qau.edu.cn)
Qingdao Agricultural University
Xue Feng
Qingdao Agricultural University
Jianbin Zeng
Qingdao Agricultural University
Xiaoyan He
Qingdao Agricultural University

Research Article

Keywords: wheat (Triticum aestivum L.), GATA, Genome-wide, Abiotic stress, Expression patterns

Posted Date: March 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1420698/v1

Abstract

Transcription factors GATAs were a member of zinc finger protein, which could bind DNA regulatory regions to control expression of target genes, thus influencing plant growth and development either in normal condition or environmental stresses. Recently, *GATA* genes have been found and functionally characterized in a number of plant species. However, little information of *GATA* genes were annotated in wheat. In the current study, 79 *GATA* genes were identified in wheat, which were unevenly located on 21 chromosomes. According to the analysis of phylogenetic tree and functional domain structures, *TaGATAs* were classified into four subfamilies (I, II, III, and IV), consist of 35, 21, 12, and 11 genes, respectively. Meanwhile, the amino acids of 79 TaGATAs exhibited apparent difference in four subfamilies according to GATA domains comparison, gene structures and conserved motif analysis. We then analyze the gene duplication and synteny between the genomes of wheat and Arabidopsis, rice and barley, which provided insights into evolutionary characteristics. In addition, expression patterns of *TaGATAs* were analyzed, and they showed obvious difference in diverse tissues and abiotic stresses. In general, these results provide useful information for future *TaGATA* gene function analysis, and it helps to better understand molecular breeding and stress response in wheat.

Introduction

Plants face many environmental challenges during development, and they must optimize growth to adapt to all kinds of environmental condition including abiotic and biotic stress. During the long-term evolution, many kinds of plants have evolved a range of protective mechanisms in response to various environmental stress, among which transcriptional regulations play a dominant role (Xu et al., 2007). Transcription factors (TFs) are vital modulators to control gene expression level via specifically binding to promoter region of the downstream gene, thus influencing or regulating a lot of important biological processes, including cellular morphogenesis, signaling transduction, and environmental stress responses (Jin et al, 2015; Franco-Zorrilla et al., 2014). In plants, many well-known transcription factor families have been found, such as GATA (GATA-binding factor) (Reyes et al., 2004), WRKY (Wang et al., 2020), MYB (Hao et al., 2021), DREB (Dehydration-responsive element-binding protein) (Niu et al., 2020), bZIP (Basic region-leucine zipper) (Li et al., 2015), and MADS-box (Li et al., 2021).

GATAs are a class of DNA binding proteins widely existed in fungi, animals and plants, which belongs to a member of type IV zinc finger, and the DNA binding domain of which is consist of a basic region with C-X_2-C-X_{17-20}-C-X_2-C form (Reyes et al., 2004). The GATA proteins could modulate the transcription level of their target genes by recognizing and binding to the (T/A)GATA(A/G) sequences of genes promoter. The first GATA factor NTL1 was found in tobacco (*Nicotiana tabacum*), which was a homolog of NIT2 form *Neurospora crassa* (Daniel-Vedele and Caboche, 1993). The GATA family gene was successively identified in a number of plants, such as *Arabidopsis thaliana* (Reyes et al., 2004), *Oryza sativa* (He et al. 2018; Zhang et al., 2018), *Glycine max* (Zhang et al., 2015; Zhang et al., 2020), *Brachypodium distachyon* (Peng et al., 2021; Guo et al., 2021), *Capsicum annuum* (Yu et al., 2021), *Cucumis sativus* (Zhang et al., 2021b), *Gossypium genus* (Zhang et al., 2019) and so on. In plants, all GATA factors are featured with

one single zinc finger domain with 18 or 20 residues. In general, the GATA family were classified into four subfamilies in *Arabidopsis thaliana*, as subfamily I, II, III and IV, in terms of phylogenetic analysis, DNA binding domains and gene structures (Reyes et al., 2004).

With the rapid development of Next Generation Sequencing (NGS), GATAs family have been found both in monocots and dicots. There were 29 *GATA* genes in *Arabidopsis*, 28 in rice, 35 in apple, and 64 in soybean according to Genome-wide analyses (Chen et al., 2017; Reyes et al., 2004; Zhang et al., 2015). Plant GATA TFs have various roles like the chloroplast development (Hudson et al., 2013), photosynthesis and growth (An et al., 2020; Lu et al., 2017), seed dormancy (Ravindran et al., 2017)), host immune response (Liu et al., 2020), Grain shape (Zhang et al., 2018), and abiotic stress (Zhang et al., 2021a; Zhao et al., 2021; Nutan et al., 2021). In Arabidopsis, *GATA12* are involved in primary seed dormancy (Ravindran et al., 2017). SWI2/SNF2 ATPase BRM could associate with GNC (GATA, NITRATE-INDUCIBLE, Carbon metabolism Involved) to regulate SOC1 (Suppressor of Overexpression of Constans 1) expression and control bloom time in Arabidopsis (Yang et al., 2021). In rice, Over-expressed *OsGATA12* lead to reduction of leaf and tiller numbers, thus affecting yield-related characters (Lu et al., 2017). OsGATA7 regulated architecture and grain shape by mediating brassinosteroids content (Zhang et al., 2018). In poplar, *PdGATA19* was responsible for photosynthesis and growth (An et al., 2020). In soybean, low nitrogen treatment led to the obvious repression of *GATA44* and *GATA58* in seedlings (Zhang et al., 2015). Given the importance of GATA in plants, the above reports manifest that GATA TFs are needed to conduct a comprehensive assessment in development, growth and stress response.

Wheat is the second major cereal crop in the world. Hence, it is important to conduct genetic and physiological research. Many wheat *GATA* genes have been found and functionally characterized. For example, over-expressing TaZIM-A1 postponed flower time and led to the reduction of thousand seed weight (Liu et al., 2019). Liu et al (2020) reported that plants of over-expressed *TaGATA1* showed high resistance to *Rhizoctonia cerealis* in wheat. In spite of this, the function of GATA factors defined remains very little in wheat. In the current study, 79 candidate *TaGATA* genes were identified based on the bioinformatic analysis of wheat genome. Generally, we performed a genome-wide analysis in wheat *GATA* genes, such as phylogeny, conserved motifs, gene structures, chromosomal distribution, and expression profiles of *GATA* genes in different tissues and diverse abiotic stresses.

Materials And Methods

Identification of *TaGATA* Genes in Wheat

Gene and protein sequences were obtained from the Ensemble Plants database (http://plants.ensembl.org) (Kersey et al., 2018). To identify the candidate *TaGATA* genes, we used a Hidden Markov Model (HMM) to search the protein database in wheat genome by HMMER3.0, in which the profiles of the GATA protein domain, PF00320, were used as queries with default parameters. Within the same gene ID, we left the longest transcript sequence, and incomplete sequence without start or termination codon were discarded. Then, we used Pfam tool with e-value $<e^{-20}$ (Finn et al., 2016) and

Conserved Domain Database (CDD) to analyzed the left sequence. Ultimately, 79 *TaGATA* genes were identified. Furthermore, ExPASy tool (http://www.expasy.ch/tools/pi_tool.html) was used to calculate amino acids number, molecular weights (MW) and isoelectric point (pI) (Artimo et al., 2012).

Phylogenetic Analysis of TaGATAs

Sequence alignment of 79 TaGATA protein was conducted using ClustalW (Larkin et al., 2007). We used MEGA 7.0 to construct Evolutionary tree by the Neighbor-Joining (NJ) method (Kumar et al., 2016). The phylogenetic tree was further beautified using the iTOL (https://itol.embl.de/) (Letunic and Bork, 2019).

Chromosomal Location and Gene Duplication

TaGATA genes localization on Chromosome was visualized by MapChart tools (v2.3.2) (Voorrips, 2002). Syntenic relationship of the orthologous *GATA* genes between *Triticum aestivum* and *other species*, including *Arabidopsis thaliana, Oryza sativa and Hordeum vulgare*, were analyzed by the MCScanX software (Wang et al., 2012). We then used KaKs_Calculator 2.0 to calculate non-synonymous (Ka) and synonymous (Ks) substitution of each duplicated *TaGATA* gene. Formula T= Ks/2R was used to assess Divergence time, where R is 1.5 × 10-8 synonymous substitutions per site per year (Wang et al., 2010).

Gene Structures and Protein Motifs Analysis

The exon/intron organization of *TaGATA* genes was identified using the Gene Structure Display Server (GSDS) tool (http://gsds.cbi.pku.edu.cn/) (Hu et al., 2015). The Multiple Expectation Maximization for Motif Elicitation (MEME) online program (http://meme.sdsc.edu/ meme/itro.html) was performed to identify conserved motifs of TaGATA proteins (Bailey et al., 2009). The exon-intron structure and conserved motif of TaGATA was examined by TBtools (Chen et al., 2020) and GFF3 database obtained from Ensemble Plants.

Cis-elements in the Promoter of *TaGATA* Genes

Promoter sequences (-1500 bp) of *TaGATA* gene was obtained from the wheat genome sequence, and cis-element of promoter region was analyzed using PlantCARE software (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al., 2002). The full graphics of Cis-elements were annotated by TBtools (Chen et al., 2020).

Gene Expression Analysis

The specific expression patterns of *GATA* gene from various tissues in the wheat Chinese spring were obtained from Wheatomics (http://202.194.139.32/) (Ma et al., 2021). The gene expression values are represented by transcript per Kilobase of exon per million reads mapped (TPM). The average expression level of three biological replicates was calculated and used to show their expression patterns in each tissue. The data were normalized to expression level in roots. Furthermore, transcriptome data under

abiotic stress were also obtained from Wheatomics. The genes with log2 ratio ≥0.5 and log2 ratio ≤ −0.5 were regarded as differentially expressed genes (DEGs). A heatmap of expression pattern profile on log2 (TPM+1) and log2 fold change scale was conducted by TBtools (Chen et al., 2020; Waterhouse et al., 2009).

Results

Identification of *TaGATAs* in Wheat

In total, 79 GATA family members were identified in wheat. The detailed information of genes and proteins were listed in Tables S1. For example, the amino acid length of 79 TaGATA proteins ranges from 146 to 499. Meanwhile, the molecular weight is ranged from 16.1 to 54.1 kDa. The GATA domain sequences were listed in Table S2.

Phylogenetic Analysis of TaGATA Proteins

To figure out the phylogenetic relationship of the GATA proteins, we constructed a evolutionary tree in terms of the alignment of 79 wheat TaGATAs and 29 *Arabidopsis* AtGATAs (Figure 1). The AtGATAs protein sequence were listed in Table S3. It was reported that 29 AtGATA proteins could be categorized into four clusters (Reyes et al., 2004). On the basis of classification standard used for *Arabidopsis*, the wheat GATA proteins were classified into four group. Group I, II, III, and IV consist of 35, 21, 12, and 11 TaGATA proteins, respectively (Figure1 and 2A).

Gene Structure and Protein Motif Analysis of TaGATA

We used the web server GSDS to analyze *TaGATA* genes structures. The results showed that *TaGATA* genes contained one to eight exons unevenly (Figure 2B). Protein motifs were determined by MEME. In general, 10 conserved motifs were found in TaGATA proteins and considered motifs 1-10 (Figure 2C). The detailed information of conserved motif were listed in Table S4. In total, 19 of 79 TaGATAsonly contain motif 1. Thirty five of 79 TaGATAs contain motif 1 and 2. The motif 1 were primarily presented in subfamily I and , and the motif 3-10 were detected in the members of group II and . In a word, similar gene structures and conserved motifs in the same subfamily forcefully back up phylogenetic analysis for subfamily classifications.

In addition, GATA domain analysis showed that TaGATAs in the subfamilies I, II and IV comprised 18 residues in the zinc finger loop between the second and the third Cys residues, while TaGATAs in the subfamily III comprised 20 residues, with the exception of TaGATA4 and TaGATA15 comprised 18 residues. In the GATA domains, many amino acid sites exhibited high conservation, such as LCNACG residues (Figure 3).

Chromosomal Location and Genome Synteny of *TaGATA* Genes

The chromosomal distribution of *TaGATA* gene were analyzed. In total, 79 *TaGATAs* were mapped to the wheat genome (Figure 4). The *TaGATA* genes were evenly located among A (29), B (25), D (25) subgenomes. This was consistent with the finding that a large proportion of *TaGATAs* have three homoeologous sequences distributed on three subgenomes. There were three *TaGATA* genes located on chromosome 3, 5. Six *TaGATAs* could be found on each of chromosomes 1 and 2. Four *TaGATA* genes were located on chromosome 6. Five *TaGATA* genes were distributed on chromosome 4A and three *TaGATA* genes were located on chromosome 4B and 4D. Chromosome 7 carried 2 *TaGATAs* which was the minimum number. With approach of BLAST and MCScanX, we detected 96 segmental duplication events in Ta*GATAs* (Figure 6; Table S5). All events were almost happened between the different chromosomes. Furthermore, 4 duplication events happened on the AA subgenome, 3 events on the BB subgenome, 4 events on the DD subgenome, and 85 events across AA/BB/CC subgenomes. The above results demonstrate that a number of *TaGATA* genes are likely to appear in the course of gene duplication, and the segmental duplication events could be of great importance in the expansion of *TaGATA* genes in wheat.

The colinearity of *TaGATA* gene pairs between *Hordeum vulgare* genome, *Arabidopsis thaliana* genome and *Oryza sativa* genome was compared. The result exhibited that three and ten *TaGATA* genes exhibited syntenic relationship with *AtGATA* and *OsGATA* genes, respectively (Figure 7; Table S6 and S7). For example, AT2G45050 showed syntenic relationship with *TaGATA3, TaGATA9* and *TaGATA14* (Table S6). However, 54 *TaGATAs* showed syntenic relationship with *GATAs* in barley (Table S8), implying that these genes may be responsible for the evolution of *TaGATAs* family.

To assess the evolutionary constraints acting, we calculated Ks values, Ka values, Ka/Ks ratios and divergence time of paralogous and orthologous on *GATA* family genes (Tables S9). Ka/Ks ratios were less than 1 in several segmental duplicated *TaGATA* gene pairs, while *TaGATA26/TaGATA31* were more than 1. The results demonstrated that *TaGATA*s family probably have suffered strong purifying selective stress in the course of evolution.

Cis-elements Analysis in *TaGATAs* Promoters

To explore the underlying function of *TaGATA* genes, we used Plant-CARE to detect the cis-elements in these genes promoter. 79 *TaGATAs* were estimated with cis-elements, such as ABRE, circadian, G-box, LTR, MSA, P-box, TCA, TGA TGACC-motif and MBS involving in ABA responses, circadian control, light response, low-temperature response, cell cycle regulation, gibberellin response, salicylic acid response, auxin response, MeJA response, drought-inducibility and flavonoid biosynthetic genes regulation (Figure 5, Table S10). In general, 69 *TaGATA* genes (87.3%) carried ABRE cis-elements, 75 *TaGATA* genes (94.9%) had G-box cis-elements, and 63 *TaGATA* genes (79.7%) carried TGACC cis-elements. In a word, the cis-elements analysis implied that a large portion of *TaGATA* genes are likely to be responded to various environmental stresses.

Expression Analysis of *TaGATAs* in Wheat Tissues

The expression patterns of 79 *TaGATAs* in 5 tissues of Chinese spring, including roots, leaves, stems, spikes, and grains, were compared (Figure 8; Table S11). On the basis of different expression pattern of these genes, they could be classified into two groups. Group 1 include 9 genes, and they were only expressed in some tissues. For example, *TaGATA4* were only expressed in spike, and no expressed in other tissues. Group 2 includes 70 genes, which displayed expression in all tissues analyzed in the current study. Group 2 can be divided into two subgroup. Twelve *TaGATAs* were assigned to the subgroup 1 with high expression levels (log_2^{TPM+1}>2) in all tissues. 10 *TaGATAs* were assigned to the subgroup 2 with low expression levels (log_2^{TPM+1}<0.5) in all tissues. The rest of 48 genes of 70 genes were belong to the subgroup 3. These results implied that *TaGATAs* showed different expression level and genes in the same subfamily also displayed different expression profile.

Expression Patterns of *TaGATAs* under Abiotic Stress

We analyzed the expression level of *TaGATA* genes under different abiotic stress using the wheat transcriptome data recently published, such as drought, heat, cold stresses and P starvation. Overall, the expression level of *TaGATA* genes significantly changed under diverse abiotic stresses (Figure 9; Table S12). Several *TaGATA* genes were in response to heat stress or P starvation. For example, the expression level of *TaGATA74, TaGATA76* and *TaGATA78* were extremely increased by P starvation. *TaGATA54, TaGATA57* and *TaGATA60* showed high expression level responding to heat stress. Meanwhile, some *TaGATA* genes were repressed by cold stress, such as *TaGATA53* and *TaGATA59*, or by P starvation, such as *TaGATA19*. In contrast, several *TaGATAs* were not induced by any abiotic stresses. For example, *TaGATA4* and *TaGATA20* displayed almost no expression alteration in response to all analyzed treatments. Instead, several genes displayed opposite expression patterns under different abiotic stress. For instance, *TaGATA78* was extremely induced by all treatments, which showed down-regulation in drought stress, but up-regulation in other treatments.

Discussion

Transcription factors take a vital regulatory role in plant growth and development. They are the key links in modulating many kinds of physiological activities. Thus far, the GATA family has been reported in a number of plant species, such as Arabidopsis, rice (Reyes et al., 2004), maize (Long et al., 2020), apple (Chen et al., 2017), and *Brassica napus* (Zhu et al., 2020). The gene structures, expression profile, characteristic features and functions have already been reported in some *GATA* genes. Nevertheless, a genome-wide analysis of the *GATA* family genes have not yet been reported in wheat (*Triticum aestivum* L.). In the present study, 79 members of *TaGATA* genes were found in the *Triticum aestivum* genome, which were identified as *TaGATA1* to *TaGATA79* based on their chromosome location (Fig. 1; Table S1). *TaGATAs* classified into four subfamilies showed obvious difference in genetic structures and expression patterns (Fig. 1 and Fig. 2; Table S11 and Table S12). The current study gives a valuable information for future functional characterization of *GATA* genes and it contributes to increase adaptive capacity when plants subjected to abiotic stress.

In plants, *GATA* genes showed low conservation in their exon/introns structures. In wheat, exons number in *TaGATA* genes ranges from 1 to 8 (Fig. 2), which is very similar to that of *Brassica napus* (1 to 9) (Zhu et al., 2020) and Arabidopsis (2 to 8), and rice (2 to 9) (Reyes et al., 2004). Except for the zinc finger, the low level of similarity in flanking sequences suggested that the different subfamilies have appeared by modular evolution through shuffling of exons encoding the zinc finger domains (Reyes et al., 2004). Large divergences in *TaGATA* gene and protein structures could cause functional differences. For instance, the GATA domain featured with 20 residues in the zinc finger in subfamily III, while other three subfamilies showed 18 residues (Fig. 3; Table S4). The CCT and TIFY domains were particularly existed in the subfamily III, which were found to be responsible for flowering, hypocotyl and root development in *Arabidopsis thaliana* (Richter, et al., 2013). For instance, AtGATA23 modulates the auxin response factors ARF7 and ARF19 and influences the lateral root initiation cell differentiation and root branching pattern (Rybel et al., 2010). In this study, *TaGATA75* and *TaGATA77* showed high expression level in most tissues of wheat (Fig. 8, Table S11). Meanwhile, the promoter of these genes had TGA-element involved in auxin-responsive. It suggested the importance of these genes in root development. The subfamily I genes were reported to be associated with plant growth and in response to environmental stresses. In *Arabidopsis*, BME3 (ortholog of TaGATA24) could enhance seed germination capacity (Liu et al., 2005). In comparison with wild-type plants, seeds in knockout of BME3 plants were more prone to dormancy and more vulnerable to cold stress. In this study, *TaGATA24* were highly expressed in all tissues and remarkably responded to heat and drought stresses (Fig. 8 and Fig. 9; Table S11 and Table S12). Meanwhile, Ravindran et al (2017) found that RGL2-DOF6 complex modulates GATA12 (GATA subfamily I) to promote seed dormancy in *Arabidopsis* (Ravindran et al., 2017). *GATA* genes in subfamily Ⅱ may be associated with flowering and also in response to abiotic stresses. Expression pattern analysis exhibit that *GATA* genes respond to diverse abiotic stresses, such as high temperature, salinity, cold, and drought treatments in rice, *Brassica juncea*, Brassica napus, *Cucumis sativus*, and pepper (Bhardwaj et al., 2015; Gupta et al., 2017; Yu et al., 2021; Zhang et al., 2021b; Zhu et al., 2020). In *Arabidopsis*, *GNC* and *GNL* (ortholog of T*aGATA38*) were associated with germination, bloom and cold stress (Richter et al., 2013). In *Brassica napus*, the expression level of *BnGATA2.5* (ortholog of *TaGATA38*) was highly depressed under ABA, drought and cold stresses. In this study, *TaGATA38* was expressed across many tissues, and was down-regulated in cold stress and P starve, and up-regulated in heat and drought stress, thus showing its strong response to environmental stresses (Fig. 8 and Fig. 9; Table S11 and Table S12). Moreover, over-expressing *BdGATA13* in Arabidopsis led to darker green leaves, more delayed flowering, and more increased drought tolerance (Guo et al., 2021). In rice, over-expressing *OsGATA16* and *OsGATA8* enhances cold and drought tolerance, respectively (Nutan et al., 2020; Zhang et al., 2021a). Over-expression of *SlGATA17* increases drought tolerance in tomato (Zhao et al., 2021). In this study, *TaGATA54*, *TaGATA57* and *TaGATA60* were increased in heat stress and drought stress, but decreased in cold stress and P starve. However, GATA subfamily IV have known very little so far. Here, Expression pattern analysis showed that *TaGATA19, TaGATA22* and *TaGATA25* were down-regulated in response to heat stress and P starve, suggesting subfamily IV of *TaGATA* may be also associated with abiotic stress.

In general, we conducted a comprehensive characterization of GATA family genes in wheat. All these results provide a basis information for manipulating *GATA* genes and facilitate marker-assisted breeding in wheat. Nevertheless, functional identification is necessary for further study to uncover the exact functional characteristic of *TaGATA* genes.

Declarations

Acknowledgements

We thank Dr. Lei Ge for his insightful advising and contribution in manuscript revision.

Author contributions

XF and QY conceived and designed the research. XF, QY, XH and JZ performed the experiments and data analyses. WL and XF wrote the article. All authors read and approved the final article.

Funding

This work was supported by the National Natural Science Foundation of China [grant No. 32101660; 32001449], and the Natural Science Foundation of Shandong Province [grant No. ZR2021QC052].

Availability of data and materials

All data analyzed during this study are included in this article and its Additional files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Conflict of interest

Authors have no conflict of interest.

References

1. An Y, Zhou Y, Han X, Shen C, Wang S, Liu C, Yin W, Xia X. 2020. *The GATA transcription factor GNC plays an important role in photosynthesis and growth in poplar*. Journal of Experimental Botany *71*, 1969–1984,
2. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K,

Redaschi N, Rossier G, Xenarios I, Stockinger H. 2012. **ExPASy: SIB bioinformatics resource portal**. Nucleic Acids Research **40**, 597–603.

3. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. **MEME SUITE: Tools for motif discovery and searching**. Nucleic Acids Research **37**, 202–208.
4. Bhardwaj AR, Joshi G, Kukreja B, Malik V, Arora P, Pandey R, **et al**. 2015. **Global insights into high temperature and drought stress regulated genes by RNA-Seq in economically important oilseed crop** *Brassica juncea*. BMC Plant Biology **15**, 9.
5. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. 2020. **TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data**. Molecular Plant **13**, 1194–1202.
6. Chen H, Shao H, Li K, Zhang D, Fan S, Li Y, **et al**. 2017. **Genome wide identification, evolution, and expression analysis of GATA transcription factors in apple (***Malus*× *domestica* **Borkh.)**. Gene **627**, 460–472.
7. Daniel-Vedele F, Caboche M. 1993. **A tobacco cDNA clone encoding a GATA-1 zinc fifinger protein homologous to regulators of nitrogen metabolism in fungi**. Molecular and General Genetics **240**, 365–373.
8. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. **The Pfam protein families database: towards a more sustainable future**. Nucleic Acids Research **44**, 279–285.
9. Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R. 2014. **DNA-binding specificities of plant transcription factors and their potential to define target genes**. Proceedings of the National Academy of Sciences of the United States of America **111**, 2367–2372.
10. Guo Ji, Bai X, Dai K, Yuan X, Guo P, Zhou M, Shi W, Hao C. 2021. **Identification of GATA Transcription Factors in** *Brachypodium distachyon* **and Functional Characterization of BdGATA13 in Drought Tolerance and Response to Gibberellins**. Frontiers in Plant Science **12**, 2386.
11. Gupta P, Nutan KK, Singla-Pareek S, Pareek A. 2017. **Abiotic stresses cause differential regulation of alternative splice forms of GATA transcription factor in rice**. <bi>Frontiers in Plant Science</bi> **8**,**1944**.
12. Hao L, Shi S, Guo H, Zhang J, Li P, Feng Y. 2021. *Transcriptome analysis reveals differentially expressed MYB transcription factors associated with silicon response in wheat*. Scientific Reports *11*, 4320.
13. He P<bi>,</bi> Wang X<bi>,</bi> Zhang X<bi>,</bi> Jiang Y<bi>,</bi> Tian W<bi>,</bi> Zhang X<bi>, </bi> Li Y<bi>,</bi> Sun Y<bi>,</bi> Xie J<bi>,</bi> Ni J<bi>,</bi> He G<bi>,</bi> Sang X<bi>.</bi> 2018<bi>.</bi> *Short and narrow flag leaf1*, *a GATA zinc finger domain-containing protein*, *regulates flag leaf size in rice (Oryza sativa)*<bi>.</bi> BMC Plant Biology **18**<bi>,</bi> 273<bi>.</bi>
14. Hu B, Jin J, Guo A, Zhang H, Luo J, Gao G. 2015. **GSDS 2.0: an upgraded gene feature visualization server**. Bioinformatics **31**, 1296–1297.
15. Hudson D, Guevara DR, Hand AJ, Xu ZH, Hao LX, Chen X, Zhu T, Bi YM, Rothstein SJ. 2013. **Rice cytokinin GATA transcription Factor1 regulates chloroplast development and plant architecture**. Plant

Physiology **162**, 132–144.

16. Jin J, He K, Tang X, Li Z, Lv L 2015. **An Arabidopsis transcriptional regulatory map reveals distinct functional and evolutionary features of novel transcription factors**. Molecular Biology Evolution **32**, 1767–1773.
17. Kersey PJ, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Grabmueller C, Kumar N, Liu Z, Maurel T, Moore B, McDowall MD, Maheswari U, Naamati G, Newman V, Ong CK, Paulini M, Pedro H, Perry E, Russell M, Sparrow H, Tapanari E, Taylor K, Vullo A, Williams G, Zadissia A, Olson A, Stein J, Wei S, Tello-Ruiz M, Ware D, Luciani A, Potter S, Finn RD, Urban M, Hammond-Kosack KE, Bolser DM, **De** Silva N, Howe KL, Langridge N, Maslen G, Staines DM, Yates A. 2018. **Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species**. Nucleic Acids Research **46**, 802–808.
18. Kumar S, Stecher G, Tamura K. 2016. **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets**. Molecular Biology and Evolution **33**, 1870–1874.
19. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. **Clustal W and clustal X version 2.0**. Bioinformatics **23**, 2947–2948.
20. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S. 2002. **PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences**. Nucleic Acids Research **30**, 325–327.
21. Letunic I, Bork P. 2019. **Interactive Tree Of Life (iTOL) v4: recent updates and new developments**. Nucleic Acids Research **47**, 256–259.
22. Li K, Debernardi JM, Li C, Lin Huiqiong, Zhang Chaozhong, *Jernstedt Judy, Maria von Korff, Zhong Jinshun, Dubcovsky Jorge.* 2021. *Interactions between SQUAMOSA and SHORT VEGETATIVE PHASE MADS-box proteins regulate meristem transitions during wheat spike development. The Plant Cell 33, 3621–3644.*
23. Li X, Gao S, Tang Y, Li L, Zhang F, Feng B, Fang Z, Ma L, Zhao C. 2015. *Genome-wide identification and evolutionary analyses of bZIP transcription factors in wheat and its relatives and expression profiles of anther development related TabZIP genes*. BMC Genomics *16*, 976.
24. Liu H, Li T, Wang YM, Zheng J, Li HF, Hao CY, Zhang XY. 2019. **TaZIM-A1 negatively regulates flowering time in common wheat (***Triticum aestivum* **L.)**. Journal of Integrated Plant Biology **61**, 359–376.
25. Liu P, Koizuka N, Martin RC, Nonogaki H. 2005. **Te BME3 (Blue Micropylar End 3) GATA zinc fnger transcription factor is a positive regulator of Arabidopsis seed germination**. Plant Journal **44**, 960–971.
26. Liu X, Zhu X, Wei X, Lu C, Shen F, Zhang X, Zhang Z. 2020. **The wheat LLM-domain-containing transcription factor TaGATA1 positively modulates host immune response to** *Rhizoctonia cerealis*. Journal of Experimental Botany **71**, 344–355.

27. Long J, Yu X, Chen D, Hu F, Li J. Identification, **Phylogenetic Evolution and Expression Analysis of GATA Transcription Factor Family in Maize (***Zea mays*). International Journal of Agriculture and Biology **23**, 637–643.
28. Lu G, Casaretto JA, Ying S, Mahmood K, Liu F, Bi Yong, Rothstein SJ. 2017. **Overexpression of** *OsGATA12* **regulates chlorophyll content, delays plant senescence and improves rice yield under high density planting**. Plant Molecular Biology **94**, 215–227.
29. Ma S<bi>,</bi> Wang M<bi>,</bi> Wu J<bi>,</bi> Guo W<bi>,</bi> Chen Y<bi>,</bi> Li G<bi>,</bi> Wang Y<bi>,</bi> Shi W<bi>,</bi> Xia G<bi>,</bi> Gu D<bi>,</bi> Kang Z<bi>,</bi> Ni F<bi>.</bi> 2021<bi>.</bi> *WheatOmics: a platform combining multiple omics data to accelerate functional genomics studies in wheat*<bi>.</bi> Molecular Plant **14**<bi>,</bi> 1965–1968<bi>.</bi>
30. Niu X, Luo T, Zhao H, Su Y, Li H. 2020. *Identification of wheat dreb genes and functional characterization of tadreb3 in response to abiotic stresses. Gene 740, 144514.*
31. Nutan KK, Singla-Pareek SL, Pareek A. 2020. **The Saltol QTL-localized transcription factor** *OsGATA8* **plays an important role in stress tolerance and seed development in** *Arabidopsis* **and rice**. Journal of Experimental Botany **71**, 684–698.
32. Peng W, Li W, Song N, Tang Z, Liu J, Wang Y, Pan S, Dai L, Wang B. 2021. **Genome-Wide Characterization, Evolution, and Expression Profifile Analysis of GATA Transcription Factors in** <bi>Brachypodium distachyon. International Journal of Molecular Science 22,</bi> **2026**.
33. Ravindran P, Verma V, Stamm P, Kumar PP. 2017. A novel RGL2-DOF6 complex contributes to primary seed dormancy in *Arabidopsis thaliana* by regulating a GATA transcription factor. Molecular Plant 10, 1307–1320.
34. Reyes JC, Muro-Pastor MI, Florencio FJ. 2004. **The GATA family of transcription factors in Arabidopsis and rice**. Plant Physiology **134**, 1718–1732.
35. Richter R, Bastakis E, Schwechheimer C. 2013. **Cross-repressive interactions between SOC1 and the GATAs GNC and GNL/CGA1 in the control of greening, cold tolerance, and flflowering time in** *Arabidopsis*. Plant Physiology **162**, 1992–2004.
36. Rybel BD<bi>,</bi> Vassileva V<bi>,</bi> Parizot B<bi>, et al.</bi> 2010<bi>.</bi> *A Novel Aux/IAA28 Signaling Cascade Activates GATA23-Dependent Specification of Lateral Root Founder Cell Identity*<bi>.</bi> Current Biology **20**<bi>,</bi> 1697–1706<bi>.</bi>
37. Voorrips RE. 2002. **MapChart: Software for the graphical presentation of linkage maps and QTLs**. Journal of Heredity **93**, 77–78.
38. Wang D, Zhang Z, Zhang Z, Zhu J, Yu J. 2010. **KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies**. Genomics Proteomics Bioinformatics **8**, 77–80.
39. Wang H, Zou S, Li Y. *et al.* 2020. *An ankyrin-repeat and WRKY-domain-containing immune receptor confers stripe rust resistance in wheat*. Nature Communication **11**<bi>,</bi> 1353.
40. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. 2012. **MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity**. Nucleic Acids Research **40**, e49.

41. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. **Jalview Version 2-a multiple sequence alignment editor and analysis workbench**. Bioinformatics **25**, 1189–1191.
42. Xu L, Yang L, Huang H. 2007. **Transcriptional, post-transcriptional and post-translational regulations of gene expression during leaf polarity formation**. Cell Research **17**, 512–519
43. Yang J, Xu YC, Wang Jianhao, Gao Sujuan, Huang Yisui, Hung Fu-Yu, Li Tao, Li Qing, Yue Lin, Wu Keqiang, *Yang Songguang*. 2021. *The chromatin remodelling ATPase BRAHMA interacts with GATA-family transcription factor GNC to regulate flowering time in Arabidopsis*, Journal of Experimental Botany *73*, 835–847.
44. Yu C, Li N, Yin Y, Wang F, Gao S, Jiao C, Yao M. 2021. **Genome-wide identification and function characterization of GATA transcription factors during development and in response to abiotic stresses and hormone treatments in pepper**. Journal of Applied Genetics **62**, 265–280.
45. Zhang C, Hou Y, Hao Q, Chen H, Chen L, Yuan S, Shan Z, Zhang X, Yang Z, Qiu D, **et al**. 2015. **Genome-wide survey of the soybean GATA transcription factor gene family and expression analysis under low nitrogen stress**. PLoS One **10**, e0125174.
46. Zhang C, Huang Y, Xiao Z, Yang H, Hao Q, Yuan S, **et al**. 2020. **A GATA transcription factor from soybean (**<bi>Glycine max</bi>**) regulates chlorophyll biosynthesis and suppresses growth in the transgenic** <bi>Arabidopsis thaliana. Plants-Basel</bi> **9**, **1036**.
47. Zhang H, Wu T, Li Z, Huang K, Kim NE, Ma Z, Kwon SW, Jiang W, Du X. 2021a. **OsGATA16, a GATA Transcription Factor, confers cold tolerance by repressing OsWRKY45-1 at the seedling stage in rice**. Rice **14**, 42.
48. Zhang K, Jia L, Yang D, Hu Y, Njogu MK, Wang P, **et al**. 2021b. **Genomewide identifification, phylogenetic and expression pattern analysis of GATA family genes in cucumber (***Cucumis sativus* **L.)**. Plants-Basel **10**, 1626.
49. Zhang Y, Zhang Y, Zhang L, Huang H, Yang B, Luan S, Xue H, **lin W**. 2018. **OsGATA7 modulates brassinosteroids-mediated growth regulation and inflfluences architecture and grain shape**. <bi>Plant Biotechnology Journal</bi> **16**, **1261**–**1264**.
50. Zhang Z, Zou X; Huang Z, Fan S, Qun G, Liu A, Gong J, Li J, Gong W, Shi Y, **et al**. 2019. **Genome-wide identifification and analysis of the evolution and expression patterns of the GATA transcription factors in three species of** <bi>Gossypium</bi> **genus**. <bi>Gene 680</bi>, **72**–**83**.
51. Zhao T, Wu T, Pei T, Wang Z, Yang H, Jiang J, Zhang H, Chen X, Li J, Xu X. 2021. **Over-expression of** *SlGATA17* **promotes drought tolerance in transgenic tomato plants by enhancing activation of the phenylpropanoid biosynthetic pathway**. Frontiers in Plant Science **12**, 634888.
52. Zhu W, Guo Y, Chen Y, Wu D, Jiang L. 2020. **Genome-wide identification, phylogenetic and expression pattern analysis of GATA family genes in** *Brassica napus*. BMC Plant Biology **20, 543.**

Figures

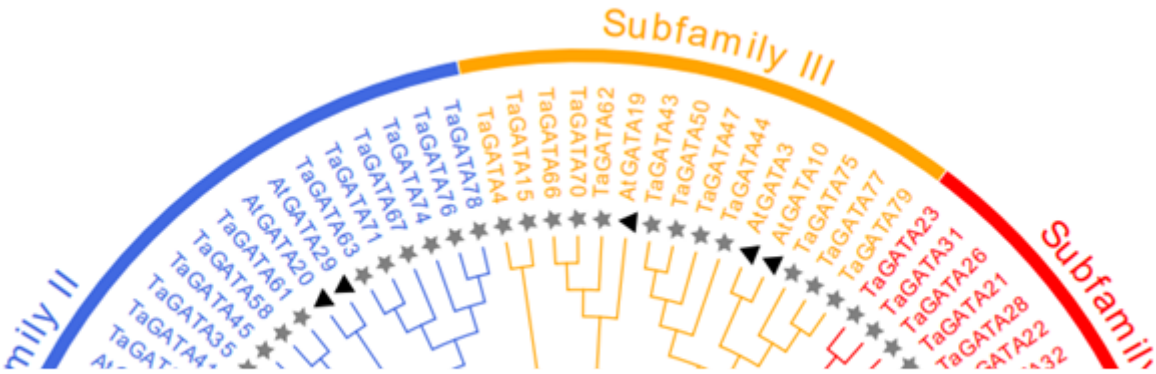

Figure 1

Phylogenetic tree of full-length TaGATA and AtGATA proteins. The different-colored arcs indicate subfamilies of the GATA proteins. The tree was constructed using identified 79 TaGATAs (asterisks) in wheat, 29 AtGATAs (triangle) from Arabidopsis. The unrooted Neighbour-Joining phylogenetic tree was constructed using MEGA7 with full-length amino acid sequences and the bootstrap test replicate was set as 1000 times.

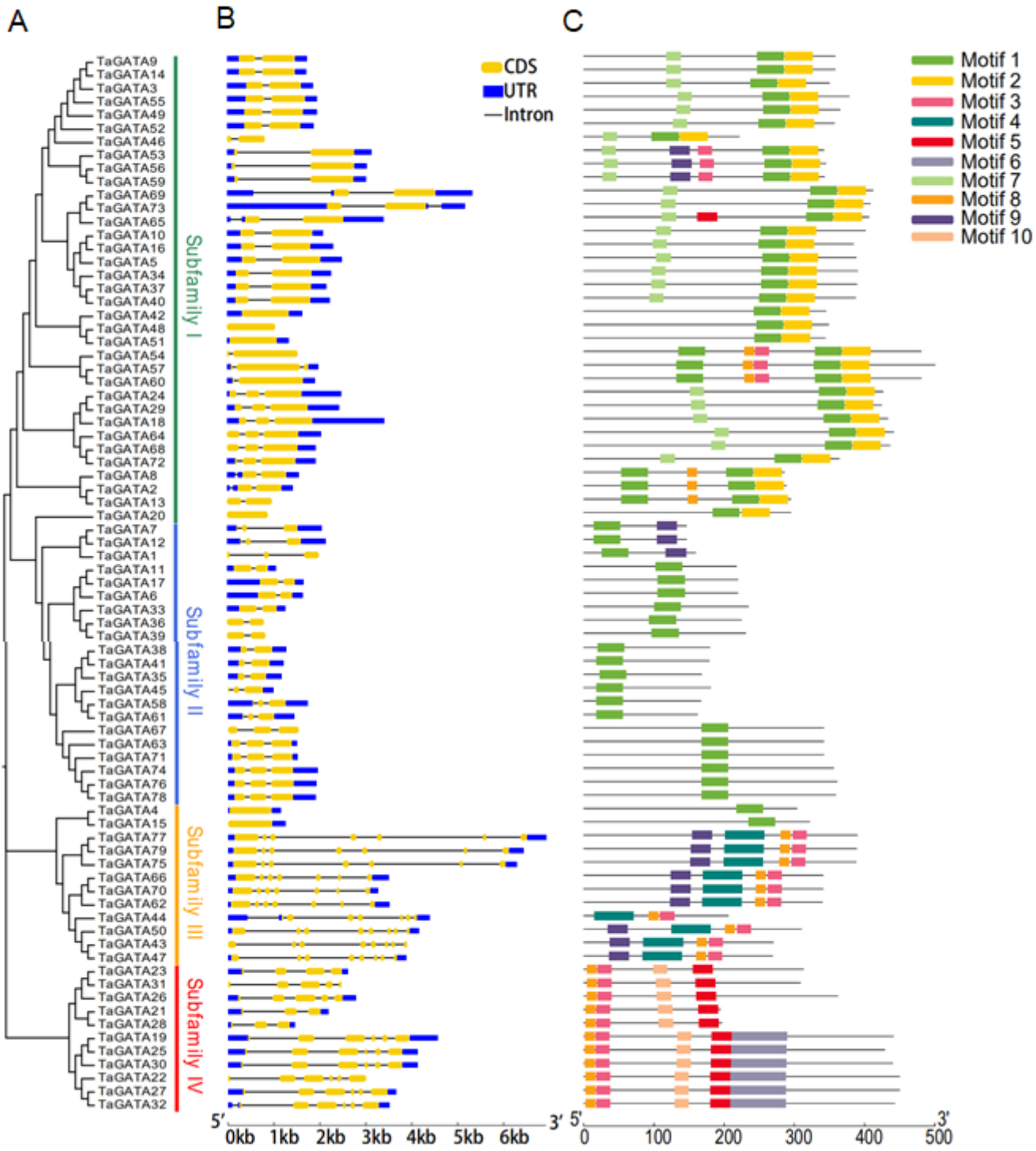

Figure 2

Phylogenetic relationships, architecture of conserved protein motifs and gene structure in *GATA* genes from wheat. (A) The phylogenetic tree was constructed based on the full-length sequences of wheat GATA proteins using MEGA 7 software. (B) Exon-intron structure of wheat *GATA* genes. Blue boxes indicate untranslated 5′- and 3′- regions; yellow boxes indicate exons; black lines indicate introns. (C) The motif composition of wheat GATA proteins. The motifs, numbers 1-10, are displayed in different colored boxes. The sequence information for each motif is provided in Supplementary Files. The length of protein can be estimated using the scale at the bottom.

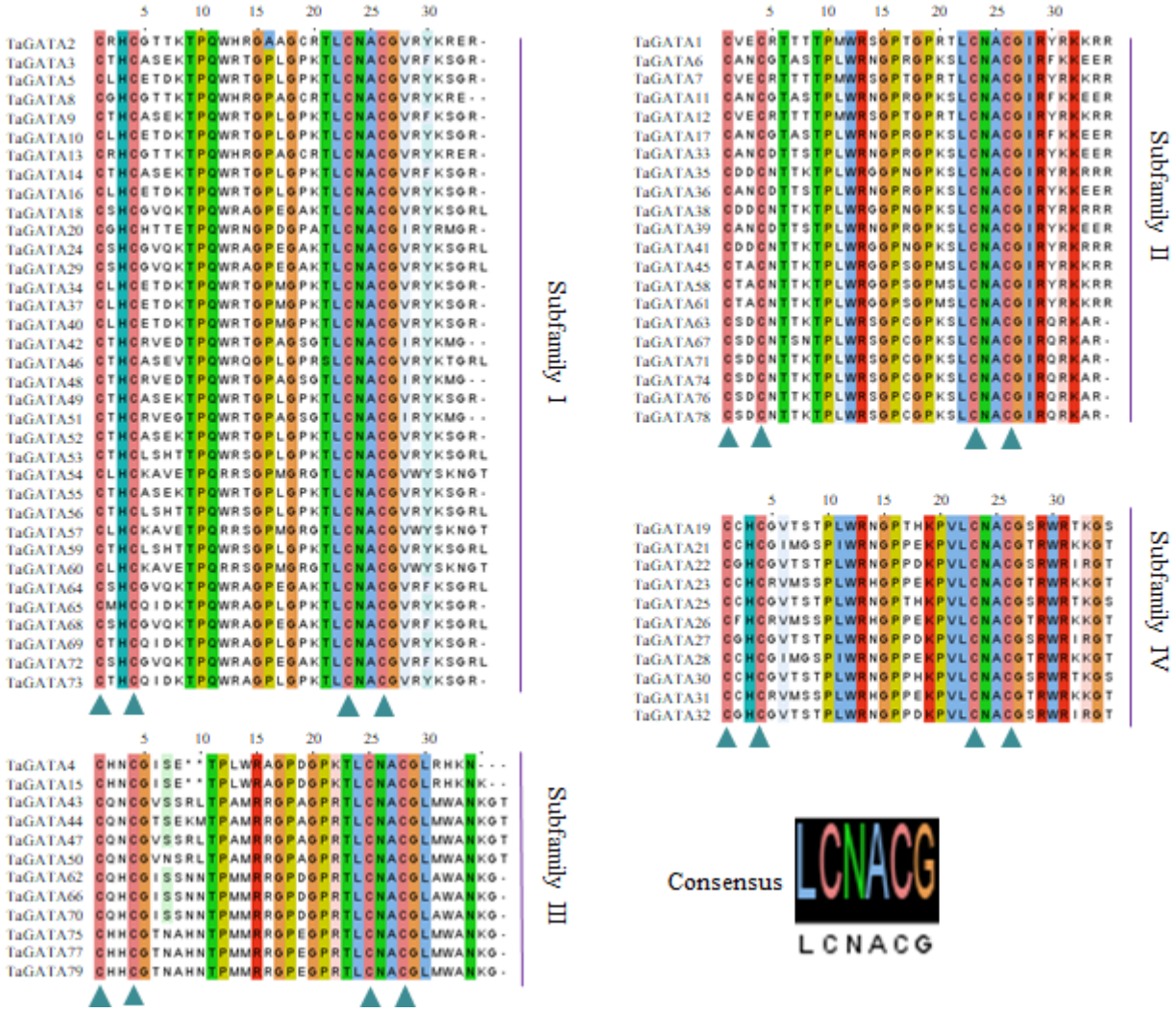

Figure 3

Alignments of GATA domain sequences of the GATA family members in wheat. Highly conserved amino acid positions are marked with letters and triangles at the bottom

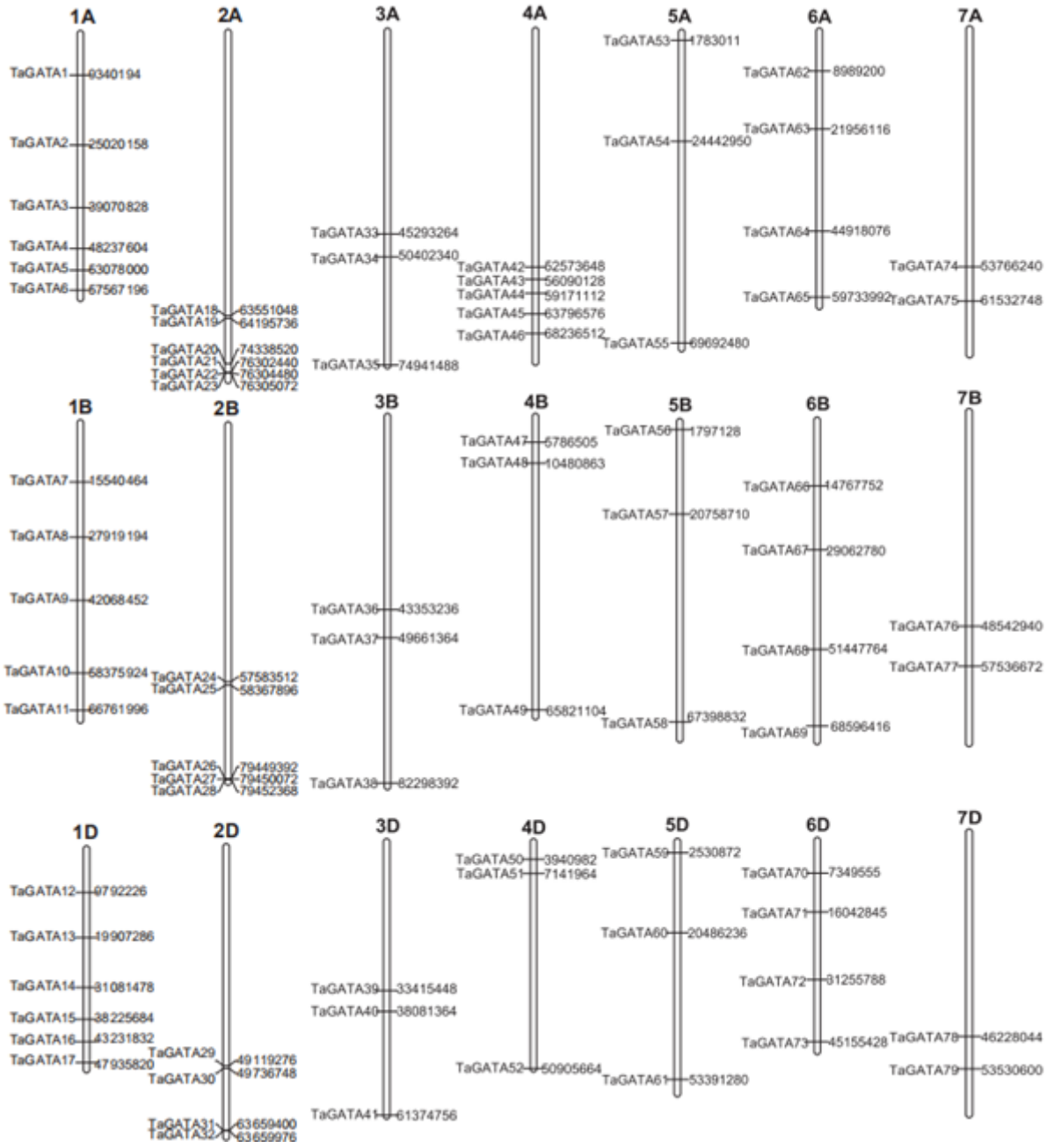

Figure 4

Distribution of *TaGATA* genes in wheat chromosomes. The chromosome numbers are indicated at the top of each chromosome image.

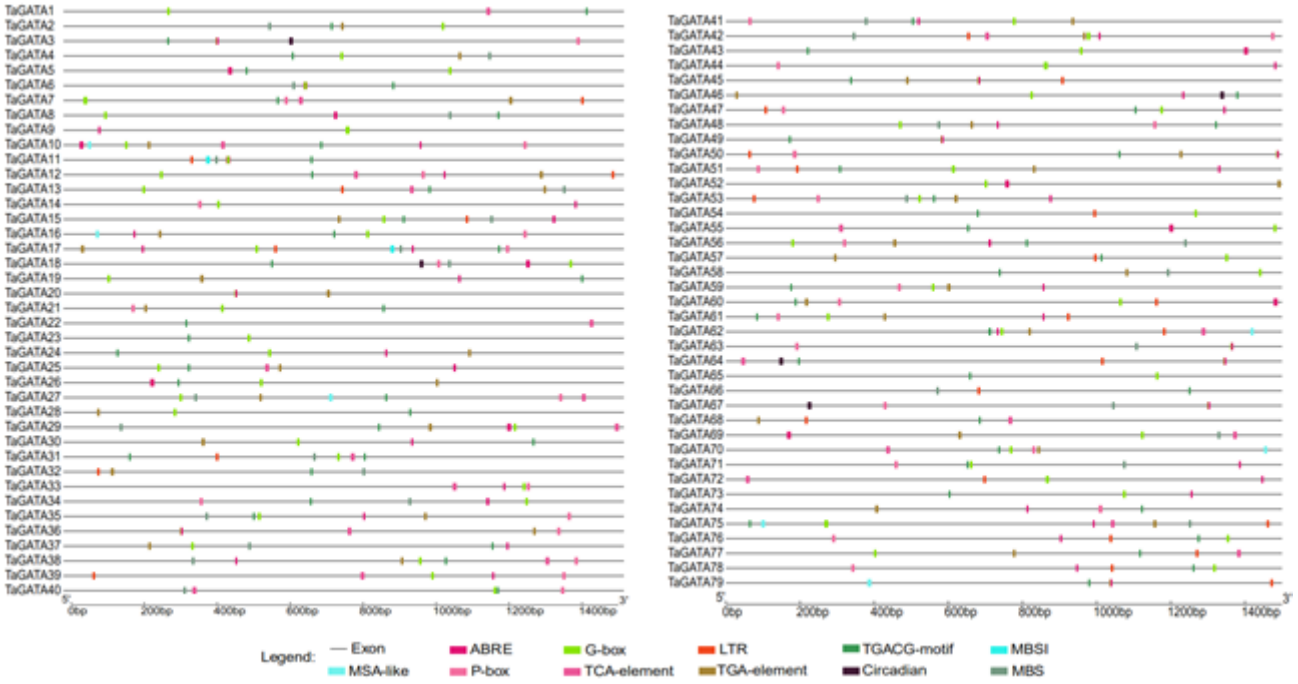

Figure 5

Predicted cis-elements in *TaGATA* promoters. Promoter sequences (−1500 bp) of 79 *TaGATA* genes were analyzed by PlantCARE. The upstream length to the translation starting site can be inferred according to the scale at the bottom.

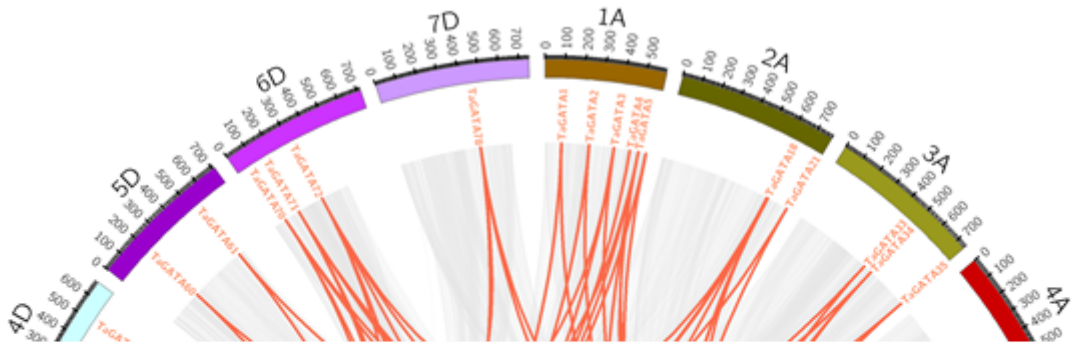

Figure 6

the synteny analysis of *TaGATA* family in wheat. Gray lines indicate all synteny blocks in the wheat genome, and the red lines indicate duplicated *TaGATA* gene pairs. The chromosome number is indicated at the bottom of each chromosome.

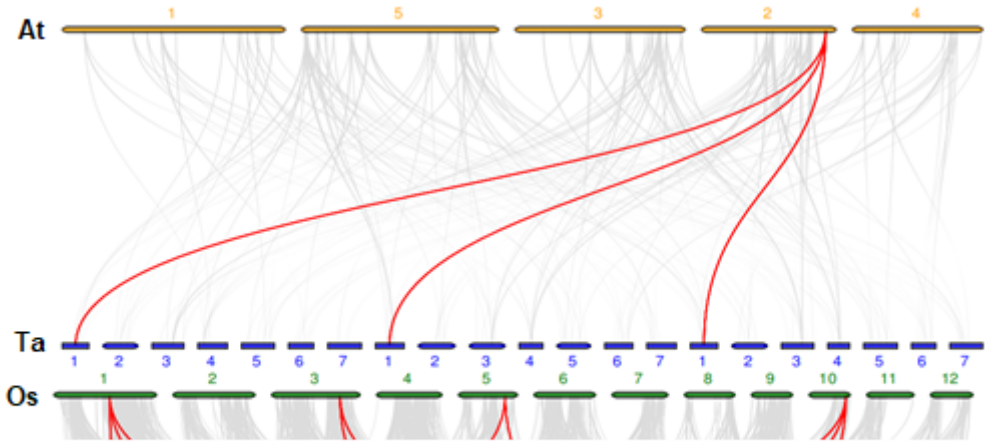

Figure 7

Synteny analysis of *GATA* genes between wheat, Arabidopsis, rice and barley. Gray lines in the background indicate the collinear blocks within wheat and other plant genomes, while the red lines highlight the syntenic gata gene pairs. The specie names with the prefixes, Hv, Ta, At and Os indicate barley, wheat, Arabidopsis and rice, respectively.

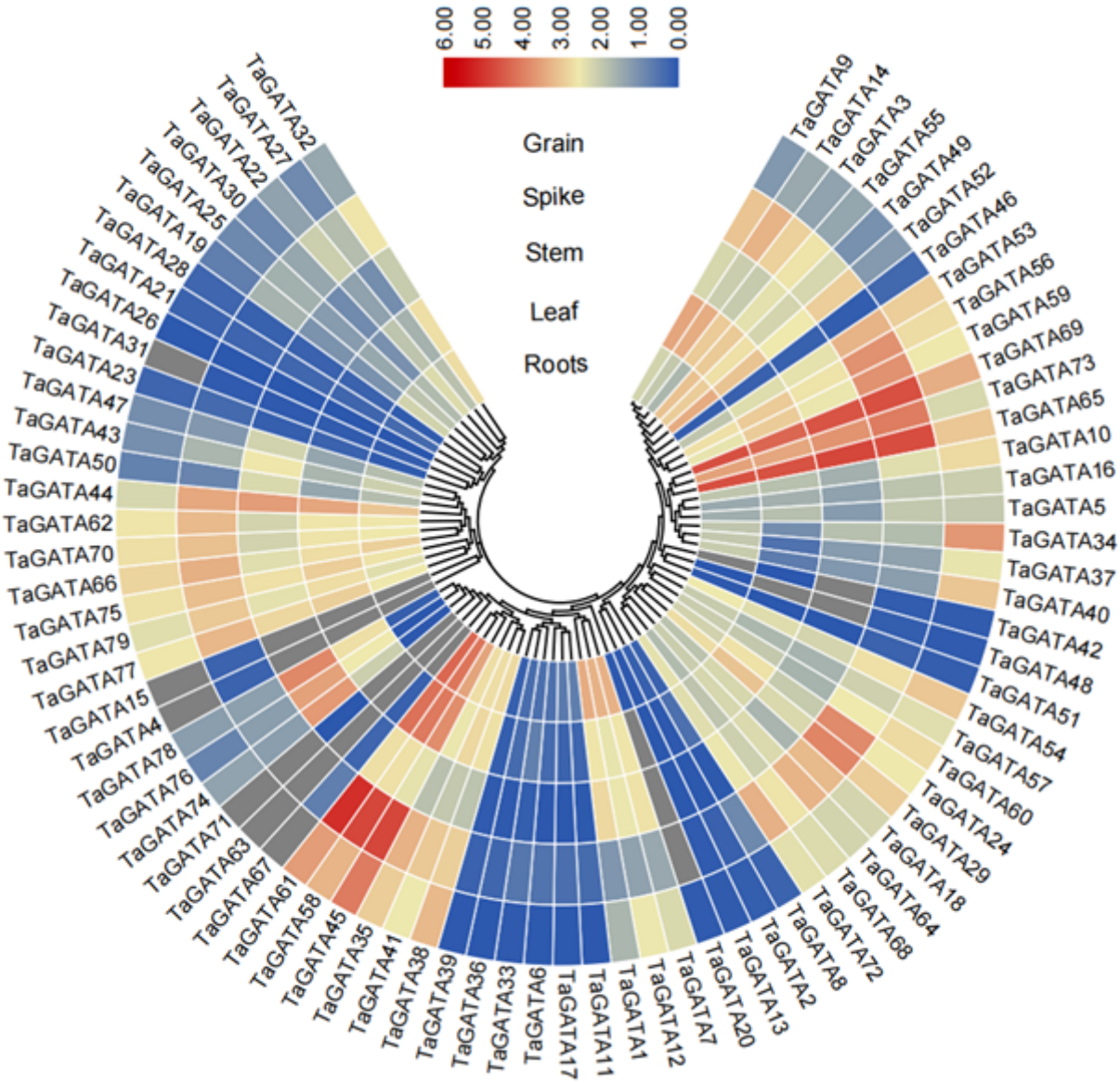

Figure 8

Expression profiles of the *TaGATA* genes in different tissues. Expression data were processed with log2 normalization. The color scale represents relative expression levels from high (red) to low (blue).

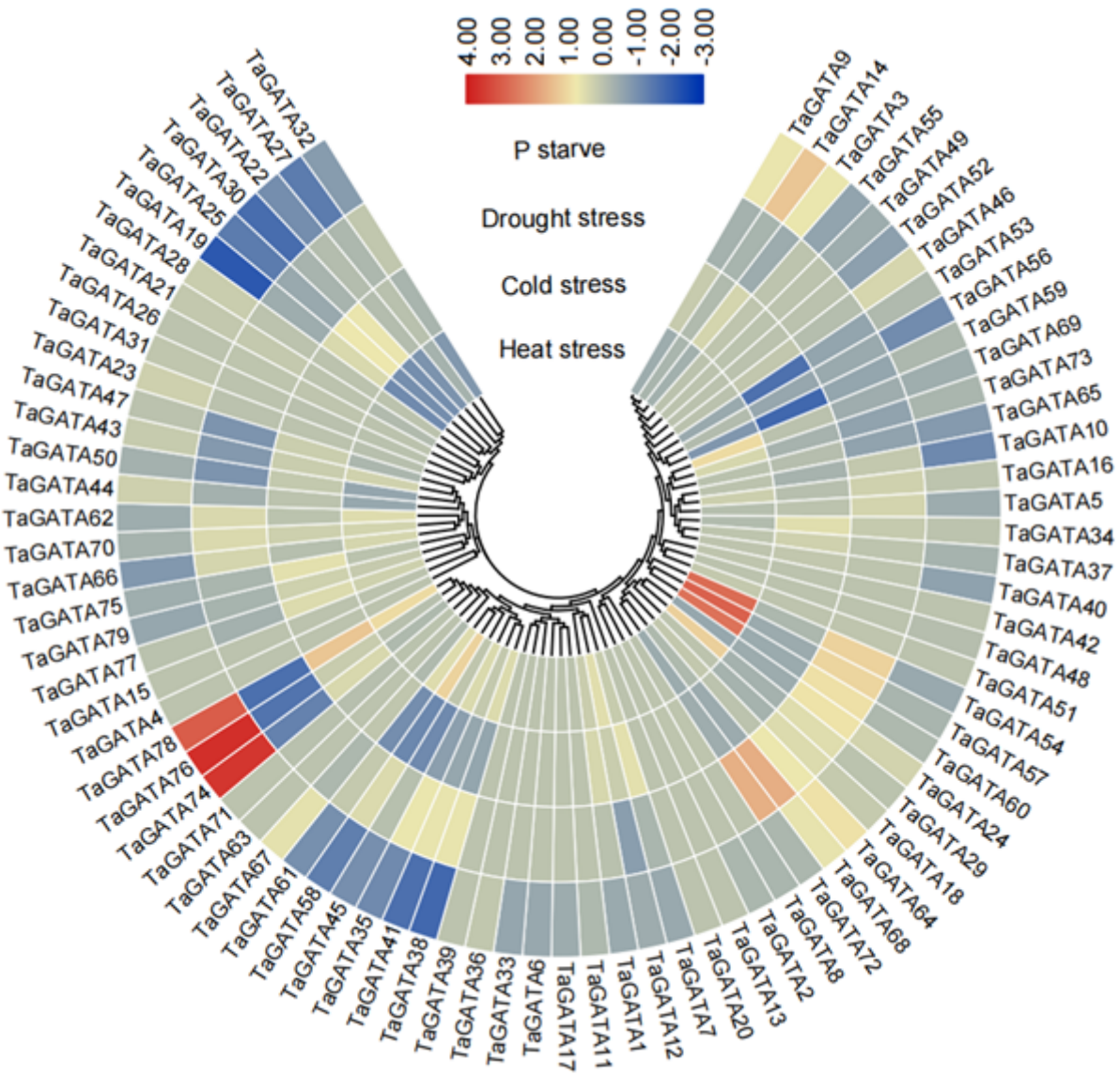

Figure 9

Expression profiles of the *TaGATA* genes under different abiotic stresses. Expression data were the ratio to control values. The color scale represents expression levels from upregulation (red) to downregulation (blue).

Supplementary Files

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

- Fengetal.TableS1S12.xlsx