Abiotic Stresses Cause Differential Regulation of Alternative Splice Forms of GATA Transcription Factor in Rice

Priyanka Gupta 1, Kamlesh K. Nutan 1, Sneh L. Singla-Pareek 2 and Ashwani Pareek 1*

1 Stress Physiology and Molecular Biology Laboratory, School of Life Sciences, Jawaharlal Nehru University, New Delhi, India, 2 Plant Stress Biology, International Centre for Genetic Engineering and Biotechnology, New Delhi, India

The GATA gene family is one of the most conserved families of transcription factors, playing a significant role in different aspects of cellular processes, in organisms ranging from fungi to angiosperms. GATA transcription factors are DNA-binding proteins, having a class IV zinc-finger motif $\text{CX}_2\text{CX}_{17-20}\text{CX}_2\text{C}$ followed by a highly basic region and are known to bind a consensus sequence $\text{WGATAR}$. In plants, GATAs are known to be involved in light-dependent gene regulation and nitrate assimilation. However, a comprehensive analysis of these GATA gene members has not yet been highlighted in rice when subjected to environmental stresses. In this study, we present an overview of the GATA gene family in rice ($\text{OsGATA}$) in terms of, their chromosomal distribution, domain architecture, and phylogeny. Our study has revealed the presence of 28 genes, encoding 35 putative GATA transcription factors belonging to seven subfamilies in the rice genome. Transcript abundance analysis in contrasting genotypes of rice—IR64 (salt sensitive) and Pokkali (salt tolerant), for individual GATA members indicated their differential expression in response to various abiotic stresses such as salinity, drought, and exogenous ABA. One of the members of subfamily VII—$\text{OsGATA23a}$, emerged as a multi-stress responsive transcription factor giving elevated expression levels in response to salinity and drought. ABA also induces expression of $\text{OsGATA23a}$ by 35 and 55-folds in IR64 and Pokkali respectively. However, $\text{OsGATA23b}$, an alternative splice variant of $\text{OsGATA23}$ did not respond to above-mentioned stresses. Developmental regulation of the $\text{OsGATA}$ genes based on a publicly available microarray database showed distinct expression patterns for most of the GATA members throughout different stages of rice development. Altogether, our results suggest inherent roles of diverse $\text{OsGATA}$ factors in abiotic stress signaling and also throw some light on the tight regulation of the spliced variants of $\text{OsGATA}$ genes in response to different environmental conditions.

Keywords: rice, OsGATA, abiotic stress, gene family, ABA, alternative splice variants, transcription factor

INTRODUCTION

Genetic material present in all living system as DNA eventually encodes and governs almost all the fundamental processes in live forms. Selective upregulation and downregulation of the set of genes encoded by DNA, allows an organism to respond to distinct stimuli. Transcription factors (TFs) act as synchronizing elements between action (stimuli) and
reaction (gene expression). Plants being sessile, require more efficiently regulated gene expression to cope with the plethora of environmental stresses. TFs are one such key regulators governing gene expression by specifically binding to the promoter/enhancer sequences of the gene. TFs can be grouped into different categories based on the ability to bind the cis-acting elements in the promoter region. These transcription factors have so far been named as MADS, WRKY, MYB, bZIP (basic leucine zipper), PHD (plant homeodomain), zinc-finger, NAC (NAM, ATAF1/2, and CUC1/2), and AP2/EERE (Apetala2/ethylene responsive element binding protein), depending upon the presence of DNA binding motifs (Shore and Sharrocks, 1995; Krishna et al., 2003; Mizoi et al., 2012; Alves et al., 2013; Ambawat et al., 2013; Sun et al., 2013). Some of these TFs are specific only to plants viz. NAC and AP2/EERE (Mizoi et al., 2012; Shao et al., 2015). Extensive studies have been carried out to understand the role of these transcription factors in biotic and abiotic stress as well as in crosstalk between these stresses in several crop plants (Cheong et al., 2002; Pandey and Somssich, 2009; Chen et al., 2010; Lindemose et al., 2013; Gupta et al., 2016; Nutan et al., 2017).

Based on the conserved domain structure, members of the zinc-finger TFs are further classified into different families (Takatsuji, 1998). GATA transcription factors, as the name suggest, are characterized by their ability to bind W-G-A-T-A-R (W = T/A, R = G/A) sequence in the promoter region (Merika and Orkin, 1993). These are type-IV zinc-finger motif with the consensus CX_2CX_{17–20}CX_2C sequence followed by a basic region facilitating DNA binding. GATA zinc-finger with 17–18 residues in the binding loop is a characteristic feature of animal and fungal GATA TFs. While, the plant GATA factors possess 17–20 residues in the zinc-finger loop (Reyes et al., 2004; Behringer and Schwechheimer, 2015). The DNA binding domain of these GATA transcription factors have been well-studied by NMR structures in chicken GATA1 and AreA zinc-finger protein of *Aspergillus nidulans* (Omithinski et al., 1993; Starich et al., 1998). Based on their studies, it is now known that the interaction between zinc-finger loop and specific DNA element is facilitated by hydrophobic interactions with the nitrogenous bases present in the major groove of the DNA.

The role of these GATA transcription factors have been widely studied in fungi and animals (Tsai et al., 1994; Marzluf, 1997; Scazzocchio, 2000; Tong et al., 2000; Marzluf, 2004; Fikkarainen et al., 2004). First identified in chicken, the GATA TFs have been reported to be involved in haematopoiesis (Omithinski et al., 1993). Apart from their active involvement in cell differentiation, GATA TFs are also documented to be involved in regulation of various stress signaling and metabolic pathways (Crespo et al., 2001; Xu and Kim, 2012). Fungal GATA TFs are the combination of both plant and animal GATA transcription factors in terms of the amino acid residues present in the zinc-finger loop. Diverse roles governed by GATA transcription factors in fungal cells include controlling nitrogen metabolism, circadian regulation, and siderophore production (Teakle and Kay, 1995; Arguello-Astorga and Herrera-Estrella, 1998; Haas et al., 1999; García-Sáucedo et al., 2006; Chi et al., 2013).

In plants, GATA TFs have not yet been studied extensively and the knowledge about this class of transcription factors remains elusive. The first plant GATA transcription factor NTL1 identified from tobacco is a homolog of NIT-2 from *Neurospora crassa* that functions in nitrogen metabolism (Daniel-Vedele and Caboche, 1993). Recent studies carried out in other plants have revealed the involvement of GATA transcription factors in regulation of various stress responsive genes, nitrogen metabolism, flowering, developmental related genes, and in hormone signaling such as GA, auxin, and cytokinin (Richter et al., 2010, 2013; Chiang et al., 2012; Hudson et al., 2013; Behringer et al., 2014; Behringer and Schwechheimer, 2015; Zhang et al., 2015). Furthermore, active involvement of *Arabidopsis* GATA TFs in the prevention of photooxidative damage via tetrapyrrole biosynthesis (TPB) has been documented very well (Kobayashi and Masuda, 2016). *Arabidopsis* class-B GATA TFs with C-terminal LLM domain, have been widely studied and characterized as regulator of vegetative growth and development (Behringer et al., 2014). In an interesting finding by Kobayashi et al. (2017), it has been observed that GNC-LIKE (GNL-class B GATA) TF functioned downstream to type B ARRs in *Arabidopsis* and hence interplayed at the junction of auxin and cytokinin signaling.

With the progression of genomic tools and availability of huge genomic data, studies related to whole genome mining have become more precise and informative. In rice, 28 gene loci encoding GATA proteins have been reported more than a decade ago (Reyes et al., 2004). Since the database is evolving continuously and with the availability of refined tools, we have carried out the search for GATA family members again to identify the new proteins. In the present work, we report 28 rice loci encoding 35 putative GATA transcription factors. Taking alternative splice variants into consideration along with gene structure, number and position of GATA domain as well as the presence of accessory domains other than GATA, we have categorized them into seven subfamilies. Further, a detailed analysis of the conserved GATA domain, sequence similarity between the genes, and the phylogenetic relationship has been performed. To investigate the role of GATA transcription factors in abiotic stress signaling, we have carried out expression analysis of OsGATA genes in two contrasting rice genotypes i.e., salt sensitive variety—IR64 and salt tolerant landrace—Pokkali. Transcript abundant analysis suggests the differential expression of alternative splice products of OsGATA genes under environmental signals. Taken together, our results may open a new path to explore the potential role of GATA transcription factors in abiotic stresses in model crop plant rice.

**MATERIALS AND METHODS**

**Characterization and Nomenclature of the GATA Gene Family in Rice**

To identify all the putative GATA gene members of rice, MSU rice genome annotation project (RGAP) release 7 database (http://rice.plantbiology.msu.edu/) was scanned with GATA pfam (http://pfam.xfam.org/) domain ID PF00320. List of
genes retrieved from MSU rice genome database was further confirmed by BLASTP (protein BLAST) in three different databases: NCBI (https://blast.ncbi.nlm.nih.gov/), phytozome V.11 (https://phytozome.jgi.doe.gov/pz/portal.html), and plant genome database release 187 (http://www.plantgdb.org/), using *Arabidopsis* GATA1 protein sequence as the reference. Redundant sequences were omitted manually. Functional domains in the full-length protein sequence were identified using pfam, Interpro (https://www.ebi.ac.uk/interpro/), and SMART (http://smart.embl-heidelberg.de/) databases. To avoid ambiguity because of multiple names, nomenclature of these GATA genes was retained same as that of Reyes et al. (2004). However, newly identified alternative spliced products were denoted as gene number extended with suffix “a” and “b” as suggested earlier by Pareek et al. (2006).

**Chromosomal Distribution of the GATA Gene Members**

For locating the GATA members on rice chromosomes, CDS coordinates were retrieved from MSU RGAP database version 7 for each GATA gene and were placed on each of the rice chromosomes according to the physical location of the gene. Plant genome duplication database (http://chibba.agtec.uga.edu/) was used to search for the duplication events (segmental duplication and tandem duplication) of the OsGATA genes. Duplicated genes have been connected by dotted lines.

**Multiple Sequence Alignment and Phylogenetic Tree Construction**

Multiple sequence alignment of protein sequences of only GATA domain was performed using ClustaW program available in MEGA 7. The Neighbor-Joining tree was generated based on the MUSCLE alignment of the full-length OsGATA protein sequences using Jones-Taylor-Thornton (JTT) model under default setting in the MEGA 7 program. To infer phylogeny, 1,000 bootstrap replicates were taken. Gene structure display server version 2.0 (http://gsds.cbi.pku.edu.cn) was used to analyze the gene structure and for calculating the number of exon and intron in the gene sequence. The cis-acting elements in the promoter region of OsGATA genes were deduced using PlantPan version 2 (http://plantpan2.itps.ncku.edu.tw/).

**Plant Material and Stress Treatment**

Seeds of *Oryza sativa* L. cv IR64 and landrace Pokkali were rinsed thoroughly in sterile water, germinated in hydroponic set-up and raised for 7 days on half-strength Yoshida medium, at 28 ± 2°C for 12 h light and dark cycles in a plant growth chamber. For stress treatment, 7 days old seedlings were transferred to half-strength Yoshida medium supplemented with either 200 mM NaCl (for salinity), 20%PEG (for drought) or 100 µM ABA and shoot samples were harvested after 4 and 24 h of stress application. Un-treated seedlings growing in half-strength Yoshida medium were taken as control. After harvesting, samples were frozen immediately in liquid nitrogen and stored at −80°C until further use.

**Extraction of Total RNA**

Total RNA was isolated from the shoots of seedlings using TRizol Reagent (Life Technologies, USA). For extraction, 100 mg tissue was homogenized to a fine powder with liquid nitrogen using pre-chilled mortar and pestle. RNA was extracted as described earlier by Soda et al. (2013). Purity and integrity of the total RNA was analyzed using spectrophotometry (Thermo Scientific, USA) and denaturing agarose gel electrophoresis respectively. The quality of RNA was checked by A260/A280 ratio and samples having A260/A280 > 1.8 were used for further analysis.

**Synthesis of First Strand cDNA**

First-strand cDNA was synthesized using first strand cDNA synthesis kit (Fermentas Life Sciences, USA) as per manufacturer’s instructions. For removing genomic DNA contamination, DNase (Epicenter, USA) treatment was done before proceeding for the cDNA synthesis as described earlier (Soda et al., 2013).

**Primer Designing and Quality Check**

Primers for OsGATA genes were designed from the region corresponding to the junction of unique 3’ UTR region and CDS sequence, using Primer Express 3.0 software (Applied Biosystems, USA). In the case of alternative splice variants, primers were designed from a unique region within the CDS sequence. The uniqueness of each primer pair to amplify a selected gene was confirmed by BLASTN using the RGAP database and NCBI databases.

**Real-Time Quantitative PCR Analysis**

The qRT-PCR analysis was performed with a Sequence Detection System ABI Prism 7500 (Applied Biosystems, USA). Reactions (final volume, 10 µl) were set up with the 2X SYBR Green PCR Master Mix (Applied Biosystems, USA), 3 µl cDNA sample and 0.5 mM of gene-specific forward and reverse primers. All the PCR reactions were performed under the following conditions: 2 min at 50°C, 10 min at 95°C and 40 cycles of 15 s at 95°C, 1 min at 58–62°C (annealing temperature range for different genes) and 30 s at 72°C in 96-well optical reaction plates (Applied Biosystems, USA). The specificity of the amplification was tested by dissociation curve analysis. Three technical replicates were analyzed for each sample and the data analysis was performed using SDS 1.4 software (Applied Biosystems, USA). For data normalization, the rice eukaryotic elongation factor 1 alpha (eEF-1α) was taken as internal control. Transcript abundance of the selected group of genes was analyzed by qRT-PCR using the 2−dCt and 2−ΔdCt method for the calculation of fold change and transcript abundance respectively (Livak and Schmittgen, 2001).

**Analysis of the Transcript Abundance of the OsGATA Genes at Different Developmental Stages of Rice**

To analyze the expression of the OsGATA genes at different developmental stages of rice, publicly available microarray database (https://www.genevestigator.com/gv/) was scanned with locus ID listed in Table 1. Expressions of these GATA genes
were analyzed at germination, seedling, tillering, stem elongation, booting, heading, flowering, milking, and dough stages of rice plant development.

**RESULTS**

**Members of the GATA Family Show Huge Diversity in Their Size, Gene Structure, and Isoelectric Point (pl)**

BLASTP search in NCBI using full-length protein sequence from *Arabidopsis* GATA1 as query identified 35 sequences which contain at least one GATA zinc-finger domain (Table 1). Further, MSU RGAP database version 7 was scanned for putative numbers of GATA genes using GATA domain ID PF00320 retrieved from pfam. Domain search also yielded 35 putative GATA transcription factors encoded by 28 gene loci. Our analysis yielded additional 7 OsGATA transcripts which were not reported earlier. Protein BLAST searches in RGAP database also yielded similar results. These 28 GATA genes were named as OsGATA1- OsGATA28 as described earlier by Reyes et al. (2004) (Table 1). The alternative spliced forms were named as "a" and "b" along with the GATA gene number (Table 1).

All the 35 GATA proteins contain at least one conserved GATA domain with a typical CX\_2CX\_18-20CX\_2 zinc-finger motif except OsGATA8b which has partially truncated zinc-finger loop. Protein domain analysis using pfam, SMART, INTERPRO databases confirmed that two of the GATA genes have more than one GATA domains in the encoded protein sequences. OsGATA26 has two, and OsGATA24 has three and one truncated zinc-finger loop in their encoded proteins. Protein sequences encoded by 15 GATA genes that contained accessory domains other than GATA might play additional roles in different physiological responses (Table 1). Among them, six members, OsGATA17, OsGATA18, OsGATA19, OsGATA20, OsGATA22, and OsGATA23 possess CX\_2CX\_20CX\_2 zinc-finger loop in GATA domain. While rest of the GATA members contain a CX\_2CX\_13CX\_2 type of domain structure. All the rice GATA members are listed in Table 1 along with the gene nomenclature, domain details, and amino acid length. However, we found many differences in the amino acid length as well as in exon/intron structure of GATA TFs from earlier reported information. Predicted amino acid length of OsGATA1, OsGATA2, OsGATA5, OsGATA6, OsGATA10, OsGATA13, OsGATA16, OsGATA20, OsGATA21, OsGATA22, OsGATA23, and OsGATA26 are 387, 431, 376, 386, 142, 225, 390, 292, 450, 732, 742, and 383 respectively but earlier reports by Reyes et al. (2004) showed the amino acid length as 386, 387, 390, 387, 140, 155, 348, 332, 303, 778, 786, and 415 respectively. In addition to amino acid length, we found differences in the number of exons. Predicted exon numbers in the OsGATA5, OsGATA7, OsGATA10, OsGATA13, OsGATA16, OsGATA19, OsGATA20, OsGATA22, and OsGATA23 are 2, 2, 2, 5, 3, 8, 6, 5, and 5 respectively which were previously reported as 3, 3, 3, 3, 4, 9, 8, 7, and 6. This variation could be because of rapidly evolving genomic data and availability of refined annotation tools in the rice genome database version 7.

The GATA TFs vary in amino acid length from 101 to 742 with a predicted isoelectric point (pl) ranging from 4.56 to 10.03 and molecular weight ranging from 10.98 to 84.43 kDa. OsGATA8b was found to be the smallest protein having amino acid length 101 and molecular weight 10.98 kDa. On the other hand, OsGATA23a was the largest protein with an amino acid length of 742 and molecular weight of 84.43 kDa.

**Chromosomal Location and Phylogenetic Relationships among the GATA Family Members Reveal Their Random Distribution in Rice Genome**

The OsGATA family members are randomly distributed on all the rice chromosomes, except VIII and IX (Figure 1). Maximum GATA genes i.e., six have been found to be present on chromosome III. On the other hand, only one each GATA gene has been annotated each on chromosomes VII and XI. The number of GATA genes vary from two to four on other rice chromosomes. OsGATA5 and OsGATA21 were clustered on chromosome IV between 27 and 27.2 Mb segments. OsGATA14 and OsGATA15 were present on chromosome V between 28.2 and 28.9 Mb region (Figure 1). Gene duplication has always been one of the well-known basis for the expansion of a gene family. Duplication can be either tandem; if duplicated genes are located on the same chromosome and closely linked or segmental; if duplicated genes are located on different chromosomes. We have observed eight segmental duplication events between OsGATA gene members (Figure 1) and one tandem duplication between OsGATA18 and OsGATA19 located between 27.5 and 30 Mb region of chromosome III (Figure 1).

Further, to compute the evolutionary distance between the genes, a Neighbour-joining tree was constructed in the Mega 7 program using the Jones Taylor Thornton (JTT) model. In this analysis, proteins with similar kind of domains got clustered in one group (Figure 2). In the course of setting up a new structural classification criteria for monocots, we have re-categorized the GATA proteins on the basis of their gene structure, the number of GATA domains, the position of GATA domain, and accessory domains (Figure 2). On the basis of homology in the GATA domain as well as the presence of accessory domain other than GATA, all the GATA genes have been subdivided into seven subfamilies (Figure 2). Typical domain structures of these TFs belonging to diverse subfamilies are presented in Figure 3. Subfamily-I has eight gene members including OsGATA1, OsGATA2, OsGATA3, OsGATA4, OsGATA5, OsGATA6, OsGATA7, and OsGATA25 (Figure 2). All these GATA proteins carry a single GATA domain at the C-terminal end (Figure 3). Among them, OsGATA1, OsGATA3, OsGATA6, OsGATA7, and OsGATA25 show the highest homology within the GATA domain (Figure 5). Subfamily-II is the largest and comprises of nine GATA genes, OsGATA8, OsGATA9, OsGATA10, OsGATA11, OsGATA12, OsGATA13, OsGATA14, OsGATA15, and OsGATA16 with the GATA domain being centrally located. OsGATA8b, one of the alternative splice variant of OsGATA8, contains a partially truncated GATA domain.
### Table 1: List of the OsGATA genes

| Chr | Locus ID | Nomenclature | Subfamily | Domains | Gene start | Gene end | Strand | Exons | Introns | Amino acid length | Predicted pl | Molecular weight (kDa) |
|-----|----------|--------------|-----------|---------|------------|----------|--------|-------|---------|---------------------|-------------|------------------------|
| Chr1 | LOC_Os01g24070.1 | OsGATA8a | II | GATA, LLM | 13570067 | 13573322 | + | 2 | 1 | 131 | 9.4 | 14.13801 |
| Chr1 | LOC_Os01g24070.2 | OsGATA8b | II | GATA, LLM | 13570067 | 13573125 | + | 3 | 2 | 101 | 9.57 | 10.98943 |
| Chr1 | LOC_Os01g47360.1 | OsGATA9 | II | GATA, HAN | 27056960 | 27056849 | - | 2 | 1 | 242 | 6.63 | 24.92279 |
| Chr1 | LOC_Os01g42100.1 | OsGATA1 | I | GATA | 31180196 | 31182456 | - | 2 | 1 | 131 | 9.4 | 14.13801 |
| Chr1 | LOC_Os01g54210.1 | OsGATA1 | I | GATA | 31180196 | 31182456 | - | 2 | 1 | 131 | 9.4 | 14.13801 |
| Chr2 | LOC_Os02g05510.1 | OsGATA17a | VI | TIFY, CCT, GATA | 2664370 | 2668170 | + | 7 | 6 | 328 | 5.0 | 34.85652 |
| Chr2 | LOC_Os02g05510.3 | OsGATA17b | VI | TIFY, CCT, GATA | 2664370 | 2668170 | + | 6 | 5 | 304 | 4.99 | 32.38363 |
| Chr2 | LOC_Os02g12790.1 | OsGATA11 | II | GATA, LLM | 6705604 | 6707734 | - | 3 | 2 | 353 | 9.39 | 37.63795 |
| Chr2 | LOC_Os02g43150.1 | OsGATA2a | I | GATA | 25996914 | 25999176 | + | 3 | 2 | 431 | 8.67 | 40.61048 |
| Chr2 | LOC_Os02g43150.2 | OsGATA2b | I | GATA | 25996914 | 25999176 | + | 3 | 2 | 431 | 8.67 | 40.61048 |
| Chr3 | LOC_Os03g05160.1 | OsGATA4 | I | GATA | 2506064 | 2507173 | + | 2 | 1 | 219 | 9.04 | 23.36916 |
| Chr3 | LOC_Os03g03850.1 | OsGATA27 | III | GATA | 1733298 | 1734113 | - | 1 | 0 | 271 | 9.4 | 28.96747 |
| Chr3 | LOC_Os03g08370.1 | OsGATA22 | VII | GATA, FAR1, MULE, SWIM | 4277280 | 4277474 | + | 5 | 4 | 732 | 7.56 | 83.39103 |
| Chr3 | LOC_Os03g61570.2 | OsGATA12 | II | GATA, LLM | 34913404 | 34914861 | + | 3 | 2 | 353 | 9.39 | 37.63795 |
| Chr3 | LOC_Os03g47970.1 | OsGATA18a | VII | TIFY, CCT, GATA | 27269705 | 27274117 | - | 8 | 7 | 319 | 4.56 | 34.1237 |
| Chr3 | LOC_Os03g47970.2 | OsGATA18b | VII | TIFY, CCT, GATA | 27269705 | 27274117 | - | 8 | 7 | 319 | 4.56 | 34.1237 |
| Chr4 | LOC_Os04g45650.2 | OsGATA5 | I | GATA | 27000814 | 27003617 | + | 2 | 1 | 219 | 9.04 | 23.36916 |
| Chr4 | LOC_Os04g46020.1 | OsGATA21a | V | GATA | 27260209 | 27264560 | - | 6 | 5 | 450 | 7.17 | 49.3862 |
| Chr4 | LOC_Os04g46020.2 | OsGATA21b | V | GATA | 27260209 | 27264560 | - | 6 | 5 | 450 | 7.17 | 49.3862 |
| Chr5 | LOC_Os05g44400.1 | OsGATA6 | I | GATA | 25832469 | 25834622 | + | 2 | 1 | 386 | 6.32 | 39.44034 |
| Chr5 | LOC_Os05g40630.1 | OsGATA13 | I | GATA, LLM | 3245204 | 3247503 | - | 5 | 4 | 225 | 8.75 | 24.6188 |
| Chr5 | LOC_Os05g50270.1 | OsGATA15 | II | GATA, HAN | 28817605 | 2881871 | + | 2 | 1 | 279 | 6.74 | 28.29209 |
| Chr5 | LOC_Os05g42980.1 | OsGATA14 | II | GATA, HAN | 28271822 | 28272850 | - | 2 | 1 | 250 | 8.51 | 25.60681 |
| Chr6 | LOC_Os06g37450.1 | OsGATA20 | VI | TIFY, CCT, GATA | 29365981 | 29370366 | - | 6 | 6 | 292 | 4.96 | 31.5719 |
| Chr7 | LOC_Os07g42400.1 | OsGATA22a | VII | GATA, FAR1, MULE, SWIM | 25371994 | 25376165 | + | 5 | 4 | 742 | 7.98 | 84.4357 |
| Chr7 | LOC_Os07g42400.2 | OsGATA22b | VII | GATA, FAR1, MULE, SWIM | 25371994 | 25376165 | + | 4 | 4 | 732 | 8.17 | 83.18586 |
| Chr10 | LOC_Os10g04810.1 | OsGATA7 | I | GATA | 21943291 | 21945186 | + | 2 | 1 | 387 | 7.4 | 39.62075 |
| Chr10 | LOC_Os10g02070.1 | OsGATA24 | V | GATA, GATA | 16847405 | 16849151 | - | 3 | 2 | 528 | 8.09 | 57.3008 |
| Chr10 | LOC_Os11g08410.1 | OsGATA28 | III | GATA | 4432776 | 4434071 | - | 1 | 0 | 431 | 6.76 | 44.55654 |
| Chr12 | LOC_Os12g42970.1 | OsGATA25 | I | GATA | 26695519 | 26696516 | + | 2 | 1 | 309 | 7.54 | 34.8304 |
| Chr12 | LOC_Os12g71120.1 | OsGATA26 | V | GATA, GATA | 3496286 | 3497437 | + | 1 | 0 | 383 | 10.03 | 40.2071 |
Members of the subfamily—II are complex in terms of their domain architecture (Figure 3) and have been functionally sub-categorized as B-class of GATA transcription factors in rice and Arabidopsis (Behringer et al., 2014; Behringer and Schwechheimer, 2015). OsGATA9, OsGATA14, and OsGATA15 possess a HAN (HANABA TARANU) domain at the N-terminal of the protein (Figure 3). On the other hand OsGATA8, OsGATA10, OsGATA11, OsGATA12, OsGATA13, and OsGATA16 possess a highly conserved LLM (leucine-leucine-methionine) domain at the C-terminal of the protein (Figure 3). Members of the subfamily-III include OsGATA27 and OsGATA28 (Figure 2) which contain an N-terminal located GATA domain. Though they possess similar domain structure but differ in their gene structure (Figure 4). These are intronless genes (Figure 4). The lone member of subfamily-IV, OsGATA21 came out as an outlier with an extreme N-terminal GATA domain (Figure 3). OsGATA21 possesses six exons and five introns (Figure 4). The two members of subfamily-V, OsGATA26 and OsGATA24 contain unique 2 and 3 & 1 truncated GATA domains respectively, in the encoded protein. Subfamily-VI comprises of GATA genes which encode for GATA protein having GATA domain along with two accessory domains namely TIFY and CCT (Figure 2). OsGATA17, OsGATA18, OsGATA19, and OsGATA20 belong to subfamily-V. The alternative splice variants of the genes OsGATA17 (a and b), OsGATA18 (a and b), and OsGATA19 (a and b) possess typical GATA domain with CX2CX20CX2 zinc-finger loop. Gene structure of the members of this subfamily is complex, having 6–7 exons in the coding sequence (Figure 4). The GATA subfamily-VII comprises of only two members, OsGATA22 and OsGATA23 (Figure 2). The coding sequence for these genes is interrupted by four introns and hence possess five exons (Figure 4). Typical domain structure of this subfamily includes GATA, FAR1, MULE, and SWIM domain (Figure 3). Both FAR1 (FAR Red Impaired Response1) and MULE (Mutator-like transposases) domains show sequence homology and possess C2H2 zinc-finger-like
motif and SWIM domain (found in SWI2/SNF and MuDR transposases). Further, to analyze the conserved amino acid residues in the GATA zinc-finger loop; we have carried out multiple sequence alignment of the GATA domains from all the peptide sequences (Figure 5). In the case of OsGATA26, both the GATA domains were kept in analysis and numbered as OsGATA26_1 and OsGATA26_2. Similarly, all the three GATA domains of OsGATA24; OsGATA24_1, OsGATA24_2, and OsGATA24_3 were aligned along with the other GATA domains. Partially truncated GATA domains of OsGATA8b and OsGATA24_4 were not included in this analysis. Close inspection of the aligned protein sequences revealed that apart from the conserved Cys residues at Cys-1, Cys-4, Cys-25, and Cys-28 in the zinc-finger loop, few amino acid residues in between the Cys-4 and Cys-28 are also conserved. The residues Thr-11, Pro-12, Gly-17, Pro-18, Lys-24, Asn-26, and Ala-27 (Figure 5), contribute to the formation of \( \alpha \)-helix in the zinc-finger loop; suggesting their role in maintaining the structural integrity of the domain.

The OsGATA Family Members Are Differentially Regulated in Contrasting Rice Genotypes in Response to Salinity, Drought, and ABA

To comment on the possible roles of OsGATA family members in abiotic stress response, transcript abundance of OsGATA genes were analyzed in two contrasting rice genotypes, IR64 and Pokkali in response to distinct abiotic stresses such as salinity, drought, and stress responsive phytohormone ABA (Figure 6). To check the expression of all the 35 transcripts, unique primer combinations (Table S2), from the junction of 3’ UTR and CDS sequence were designed. However, OsGATA6, OsGATA7, OsGATA9, OsGATA14, OsGATA15, OsGATA19b, OsGATA21a, OsGATA24, OsGATA27, and OsGATA28 could not be amplified from any of the cDNA used in this analysis; therefore these were kept out of the expression analysis. Expression data of remaining OsGATA genes has been presented in the form of heat map (Figures 6A–F).
FIGURE 3 | Typical domain structures of the GATA subfamily members in rice. Schematic representation of conserved domains of OsGATA subfamilies. Each domain is represented as the colored shape with the name listed at the right. The scale at the bottom of the figure represents the amino acid length of the longest OsGATA protein of particular subfamily.

FIGURE 4 | Structure of the OsGATA genes in the rice genome. Exon/intron structure was deduced using gene display server. Black filled boxes represents exon, black line represents introns and white boxes denotes 3′ and 5′ UTR. The scale at the bottom represents the bases in kilobase pairs.

The qRT-PCR analysis revealed unique findings for the OsGATA family expression. Differential accumulation of the OsGATA transcripts was observed in IR64 and Pokkali (Figures 6A–I). Basal level expression of OsGATA2b, OsGATA8b, OsGATA11, OsGATA16, OsGATA17b, OsGATA20, OsGATA22, OsGATA23b, and OsGATA25 were comparatively higher in both the rice genotypes (Figures 6A–F). However, expression of some of the genes was found to be genotype specific. GATA members such as OsGATA17a, OsGATA18a, OsGATA18b, OsGATA19a, and OsGATA21b were highly expressed in IR64 under control conditions. On the other hand, expression of OsGATA1 and OsGATA10 were higher only in Pokkali.
FIGURE 5 | Multiple sequence alignment of the GATA domain of the OsGATA family members. Analysis of the GATA domain of 28 different OsGATA genes of rice showing characteristic CX2-CX4-CX18-CX2 conserved residue in the Zn-finger loop. Multiple sequence alignment was made using ClustalW program in MEGA 7. Conserved residues in all the OsGATA proteins viz. CCTPGPLCNAC are labeled with a star at bottom of the alignment. The scale at the top of the alignment denotes the amino acid length in the conserved zinc-finger loop.

|        | 5 | 10 | 15 | 20 | 25 | 30 | 35 |
|--------|---|----|----|----|----|----|----|
| GATA22 |   |    |    |    |    |    |    |
| GATA23a|   |    |    |    |    |    |    |
| GATA23b|   |    |    |    |    |    |    |
| GATA19a|   |    |    |    |    |    |    |
| GATA19b|   |    |    |    |    |    |    |
| GATA18a|   |    |    |    |    |    |    |
| GATA18b|   |    |    |    |    |    |    |
| GATA20  |   |    |    |    |    |    |    |
| GATA20  |   |    |    |    |    |    |    |
| GATA21a|   |    |    |    |    |    |    |
| GATA21b|   |    |    |    |    |    |    |
| GATA9   |   |    |    |    |    |    |    |
| GATA14  |   |    |    |    |    |    |    |
| GATA15  |   |    |    |    |    |    |    |
| GATA8a  |   |    |    |    |    |    |    |
| GATA13  |   |    |    |    |    |    |    |
| GATA11  |   |    |    |    |    |    |    |
| GATA16  |   |    |    |    |    |    |    |
| GATA10  |   |    |    |    |    |    |    |
| GATA12  |   |    |    |    |    |    |    |
| GATA24_2|   |    |    |    |    |    |    |
| GATA24_3|   |    |    |    |    |    |    |
| GATA26_1|   |    |    |    |    |    |    |
| GATA24_1|   |    |    |    |    |    |    |
| GATA27  |   |    |    |    |    |    |    |
| GATA28  |   |    |    |    |    |    |    |
| GATA26_2|   |    |    |    |    |    |    |
| GATA4   |   |    |    |    |    |    |    |
| GATA2a  |   |    |    |    |    |    |    |
| GATA2b  |   |    |    |    |    |    |    |
| GATA5   |   |    |    |    |    |    |    |
| GATA1   |   |    |    |    |    |    |    |
| GATA6   |   |    |    |    |    |    |    |
| GATA3   |   |    |    |    |    |    |    |
| GATA7   |   |    |    |    |    |    |    |
| GATA25  |   |    |    |    |    |    |    |

OsGATA3 was upregulated in response to exogenous ABA in both the genotypes (Figures 6C,F). However, induction was more pronounced in Pokkali. OsGATA26 accumulated in response to salinity and drought in both the genotypes.

genotype under the control conditions (Figures 6D–F). Also, the transcript levels of the OsGATA2a and OsGATA13 in IR64 and Pokkali were not affected by any of the aforesaid stresses (Figures 6A–F).
but accumulation pattern varied with duration of stress given (Figures 6A,B,D,E). In Pokkali, the gene was upregulated after 4 h of salinity and drought stress. However, in IR64, under drought stress, transcripts accumulated at 24 h of stress while under salinity stress, induction is marked at 4 h (Figures 6A,B).

In IR64, members of subfamily-VI such as OsGATA17a, OsGATA17b, OsGATA18a, OsGATA18b, OsGATA19a, and OsGATA20 as well as OsGATA22, a member of subfamily-VII, maintained higher transcript level at 24 h of all the applied stresses (Figures 6A–F). OsGATA23a also maintained higher transcript level at 24 h but only under drought in IR64. Contrarily, in Pokkali, the expression pattern of the gene members from subfamily VI and VII varied with the duration of stress (Figures 6D–F). Though induction was observed with the onset of stress (4 h) in response to salinity, drought as well as ABA (Figures 6D–F); but the transcripts level declined as the stress continued for 24 h, in salinity and ABA (Figures 6D,F); finally maintaining a level higher than the control. Under drought, transcript levels maintained a similar accumulation at the end of 24 h as that of 4 h level (Figure 6).

In this study, differential regulation of the alternative spliced variants of some of the OsGATA genes was also identified (Figures 6A–I). The transcript level of OsGATA2a, a spliced variant of OsGATA2, remained unchanged in all the applied stresses for both the rice genotypes (Figures 6A–F). Interestingly, expression of OsGATA2b was induced in response to salinity, drought, and ABA in both the genotypes at early stress duration (4 h). However, responses varied from one genotype to another during the later duration of stresses. In case of IR64, OsGATA2b showed higher transcript levels at 24 h post salinity and drought while the expression was downregulated in the ABA treated samples (Figures 6A–C). On the other hand, downregulation was seen in the Pokkali samples for salinity as well as ABA stress treatment at the end of 24 h while a higher level was seen in case of drought stress imposed samples for the same duration of time (Figures 6C–F).

Expression of OsGATA8b, the alternative spliced variant of the gene OsGATA8, was higher than the OsGATA8a in both the genotypes under all the three conditions tested (Figures 6A–F). Similarly, splice variants of OsGATA23 also varied in their expression pattern in both the genotypes in response to...
abiotic stresses (Figures 6A–F). In IR64 as well as in Pokkali, OsGATA23a showed an induction at 4 h stress duration and continued to maintain higher transcript till the end of 24 h of salinity and drought but it was downregulated in response to ABA at 24 h in both the genotypes (Figures 6A–F). On the other hand, transcripts of OsGATA23b were relatively low as compared to the OsGATA23a in both the genotypes. It can be said that, though the gene showed induction at the onset of stress but ultimately its transcript levels declined as the stress prolonged for 24 h in both the genotypes (Figures 6A–F).

In terms of fold change, maximum induction i.e., 40-folds was observed for OsGATA8a and OsGATA26 in IR64 under salinity stress (Figure 6G). Furthermore, in Pokkali salinity induced transcript levels of OsGATA8b, OsGATA18a, and OsGATA23a by more than 100-folds (Figure 6G). Interestingly in IR64, drought stress and ABA application lead to induction of OsGATA23a by more than 150- and 35-folds respectively (Figures 6H, I).

Similarly in Pokkali, it was seen that drought and ABA modulate the expression of a set of genes that included OsGATA4, OsGATA8a, and OsGATA21b upregulating them by more than 100- and 35-folds with respect to control under drought and ABA treatment respectively (Figures 6H, I). Similarly, OsGATA11 and OsGATA26 show more than 200- and 100-fold change respectively in transcript level under the drought stress. On the other hand, OsGATA18a showed ABA-dependent upregulation up to 300-folds and OsGATA23a showed a 250-folds change in response to ABA (Figure 6I).

The expression of the OsGATA23a was also found to be higher as compared to the OsGATA23b. In our study, OsGATA23a was identified as multi-stress responsive gene as it showed maximum upregulation for salinity, drought as well as ABA stress. In both the genotypes, this gene was induced at 4 h of stress. In the case of IR64, induction was 40-fold while in Pokkali more than 60-fold induction was observed. Since ABA is a stress hormone, it not only governs stomatal opening but also acts as a master regulator of abiotic stress signaling. Our data suggested that some of the OsGATA genes are highly responsive to ABA, hence these OsGATA genes might also be interplaying an important role at the junction of stress signaling cascade.

To understand the regulation of these GATA transcription factors, we analyzed the cis-acting elements in the promoter region of OsGATA genes (Table S1). Binding sites of various stress responsive transcription factors such as AP2 (Apetala2), ERF (Ethylene Response Factors), bHLH (basic helix loop helix),
Expression of the OsGATA Genes Is Developmentally Regulated

To comment on the role of OsGATA transcription factors in rice developmental processes, we analyzed the expression data from publicly available microarray database, Genevestigator (Table S3). Transcript abundance of distinct members of the OsGATA gene subfamilies at various developmental stages like germination, seedling, tillering, stem elongation, booting, heading, flowering, milking, and dough stage was checked (Figure 7). Interestingly, transcripts of members of the subfamily IV and VI were found to be comparatively abundant throughout all the developmental stages of the rice plant. On the other hand, members of the subfamily II showed huge variation in terms of fold change at the different stages from seedling to maturity. The members of subfamily I and II showed mixed expression pattern. Our analysis showed that the OsGATA12 gene was induced the most during seedling stage amongst all OsGATA genes. Expression of OsGATA22, members of subfamily VII, was observed to be low as compared to other GATA gene members (Figure 7). On the other hand OsGATA23 falls under higher expression group (Figure 7).

DISCUSSION

Regulated expression of a gene is essential for distinct physiological and biochemical processes in a living system. Transcription factors play a key role in governing gene regulation and exhibiting differential expressions under different physiological and environmental conditions. In this study, we present a detailed analysis of GATA transcription factors in rice. Total 28 OsGATA genes were identified in corroboration with the previous finding by Reyes et al. (2004). Besides, we have reported 35 putative OsGATA TFs encoded from 28 OsGATA genes. Newly identified alternative spliced products of the OsGATA genes—OsGATA8, OsGATA12, OsGATA17, OsGATA18, OsGATA19, OsGATA21, and OsGATA23, contribute for the expansion of complex rice GATA TFs family. Alternative splicing is the hallmark of the complex transcriptome in eukaryotes, as the splice variants can drive the diverse functions of a gene (Park and Gravelle, 2007; Syed et al., 2012). In our analysis, we have observed that the splice variants of the OsGATA2, OsGATA8, OsGATA18, and OsGATA23 respond differently to diverse environmental conditions.

Taking into account, the distinct domain architecture of the GATA proteins, their complex gene structure and phylogenetic analysis, the rice GATA genes were categorized into seven subfamilies. Subfamily II is the largest with nine GATA members viz. OsGATA8, OsGATA9, OsGATA10, OsGATA11, OsGATA12, OsGATA13, OsGATA14, OsGATA15, and OsGATA16. In this case, the GATA binding domain is centrally located and the GATA domain coding region is split into two halves by an intron sequence. However, members of the subfamily-II have been well characterized and functionally categorized into B-class of GATA gene family (Behringer and Schwechheimer, 2015). Additional domains like LLM and HAN present in the members of subfamily-II, have been well-studied and identified as functional component of plant growth in Arabidopsis, tomato, Brachypodium, and barley (Behringer et al., 2014). They are involved in regulation of various physiological as well as structural transitions in plants such as germination, hypocotyl elongation, embryo development, flower development, and senescence (Behringer et al., 2014; Behringer and Schwechheimer, 2015). These additional domains present in rice GATA TFs might also be involved in various stages of rice plant development. With the eight members grouping together, subfamily I is the second largest. Members of the subfamily I possess C-terminal GATA domain with a typical CXC2CX12CX2 zinc-finger loop. Subfamily III possesses two GATA members wherein the GATA domain is located at the N-terminal. OsGATA21 is the only member of subfamily IV with unique six exons in the gene sequence and extreme N-terminal zinc-finger loop. Subfamily V includes OsGATA26 and OsGATA24 with 2 and 3 GATA domains respectively. Members of the subfamily VI and VII are the most peculiar as they possess CXC2CX12CX2 like zinc-finger loop for DNA binding. Apart from GATA domain, members of subfamily VI have TIFY and CCT domains. On the other hand, members of the subfamily VII contain FAR1, MULE, and SWIM domains. Previous studies related to GATA transcription factors in animals suggested that the C-terminal GATA finger proteins are involved in recognizing DNA elements in the promoter region, while N-finger GATA either assist this binding by stabilizing the DNA-protein complex or are involved in other physiological processes (Ko and Engel, 1993; Pedone et al., 1997). Like in animal system, extra GATA domain present in rice OsGATA24 and OsGATA26 might also be playing crucial roles in diverse cellular processes. Expression analysis carried out under salinity and drought stress resulted in more than 50-fold change in OsGATA26 transcripts. Furthermore, exogenous application of ABA also upregulated expression of OsGATA26 by more than 10-folds. These findings indicate that the additional GATA domains present in OsGATA26 might be regulating its role in abiotic stress signaling.

GATA TFs have been shown to play an integral role in light-mediated signaling (Putterill et al., 1995; Chen et al., 2009). In our genome-wide investigation, we have found that members of subfamily VI possess a unique CCT domain which is also found in TOC1 (Timing of Cab Expression 1) and CO (CONSTANS) proteins. Both of these proteins are important components of light signaling, circadian clock, and flowering (Robson et al., 2001; Más et al., 2003; Wenkel et al., 2006; Gendron et al., 2012). Our analysis also revealed the presence of FAR1 domain along with GATA domain in members of subfamily VII. The FAR1 domain is found in proteins involved in phytochrome signaling (Hudson et al., 1999; Lin et al., 2008; Li et al., 2011). Although, no clear evidence for the involvement of rice GATA TFs in light signaling has been established yet, the domain analysis presented here indicates that OsGATA22 and
OsGATA23 bearing FAR1 domain might be crucial components of phytochrome signaling and the circadian clock. In higher plants, proteins with FAR1 and SWIM domains are involved in phytochrome signaling. This suggested that GATA TFs having accessory domains like FAR1, SWIM and MULE domain might play a role in light-regulated signaling in plants. Luo et al. (2010) have demonstrated that one GATA TF, GATA2 from Arabidopsis functions at the junction of brassinosteroid and phytochrome signaling. This indicates that GATA transcription factors with similar domain structures may be functioning via a similar pathway.

To comment on the role of these GATA TFs in abiotic stresses, we have analyzed the relative transcripts level in response to salinity, drought, and multi-stress responsive phytohormone ABA. We found that under non-stress conditions, the transcript level of some of the GATA genes was higher in salt tolerant genotype Pokkali as compared to salt sensitive variety IR64. OsGATA1 and OsGATA10 maintained higher transcript levels in non-stress conditions in Pokkali. Earlier, from our lab, it has been reported that salt tolerant Pokkali maintains higher constitutive level of stress related genes which are otherwise induced in salt sensitive IR64 (Karan et al., 2009; Kumari et al., 2009; Soda et al., 2013; Nutan et al., 2017). Recent study on expression of Saltol QTL localized transcription factors revealed that Pokkali has an abundance of these transcripts which are otherwise induced under stress conditions only in IR64 genotype (Nutan et al., 2017). Our findings also indicate that above mentioned GATA TFs, OsGATA1, and OsGATA10 might also be acting as mediators of abiotic stress signaling and response.

Furthermore, expression of some of the OsGATA genes are stress specific but not genotype-specific. OsGATA3 is specifically induced in response to ABA and OsGATA26 in response to salinity and drought in both the genotypes. On the other hand, expression of members of subfamily VI and subfamily VII were higher in all the applied stresses. These findings suggest that OsGATA proteins with accessory domains may be function via a cross talk between different abiotic stress signaling pathways. Moreover, no change in the expression of OsGATA2a and OsGATA13 was observed with respect to any of the applied stresses in both the genotypes. This shows that both of these genes might not be part of abiotic stress signaling rather might be functioning in the rice developmental pathways. There are reports confirming the role of GATA TFs in distinct plant development and physiological processes (Liu et al., 2005; Lu et al., 2017).

In addition to this, differential regulation of alternative spliced forms of an OsGATA gene is also observed in response to abiotic stresses. OsGATA23a, a spliced variant of OsGATA23 is highly expressed in all the three stresses while relative expression of OsGATA23b was very low under salinity, drought, and ABA. Numerous reports are available justifying the role of alternative spliced forms in different environmental circumstances (Mastrangelo et al., 2012). Recent studies carried out by Jiang et al. (2017) reveal that heat stress alters the expression of some RNA binding proteins which in turn promotes the alternative splicing in grape. Our data also shows stress specific expression of the alternative spliced variants of OsGATA genes suggesting a similar mechanism operating in rice for the regulation of OsGATA genes under abiotic stresses.

To further gain insight into the regulation of OsGATA genes, we looked for the various cis-acting elements in the promoter region of OsGATA genes (Table S1). Binding sites for various stress responsive TFs were studied. Interestingly, OsGATA23 is the only gene having MADS box binding site in the promoter region. In rice, it has been reported that members of OsMADS gene family are regulators of abiotic stress signaling (Arora et al., 2007). Recently, it was reported that many of the OsMADS TFs are localized in Saltol QTL and are differentially regulated in contrasting rice genotypes (Nutan et al., 2017). As such it can be said that the multistress responsive nature of the OsGATA23 may be because of the binding and regulation via MADS-box TFs.

We have also examined the expression of OsGATA genes at different stages of rice plant development. In this analysis, we observed that members of subfamily IV and VI vary in their expression pattern from a medium to high level as the development proceeds from seedling to dough stage. In contrast, OsGATA gene members from subfamily I, II, and III showed varied expression level ranging from a low to high. At seedling stage, expression of OsGATA12 was observed to be highest. In an interesting study carried out in Arabidopsis, it has been reported that BME3 which is a GATA transcription factor, plays a significant role in seed germination (Liu et al., 2005). Transgenic Arabidopsis seeds under-expressing BME3 showed delayed germination. Our analysis based on genome-wide expression analysis clearly shows that OsGATA genes might be playing a crucial role in rice developmental pathways. Behringer et al. (2014) have documented that mutation in LLM domain of B-GATA transcription factor of Arabidopsis affects plant growth and hypocotyl elongation. Besides this, distinct GATA factors are identified as functional component of shoot apical meristem (SAM) development, chloroplast development, flowering, growth, and cell division in Arabidopsis (Zhao et al., 2004; Chiang et al., 2012). Therefore, it is clearly evident that the GATA factors are one of the important messengers for plant cell transitions and differentiation.

Our in silico analysis and expression data markedly sheds some light on the complex circuitry of rice GATA transcription factors and their potential role in various physiological processes as well as in abiotic stress signaling. Additionally, it also opens a path for the future exploration and characterization of OsGATA genes to further understand the molecular regulatory network of transcription factors in rice.

**CONCLUSIONS**

In this study, whole genome analysis of OsGATA gene family in rice was done to identify putative OsGATA transcription factors encoded by rice genome. Their gene, as well as protein structure, phylogeny, chromosomal location, was deduced and expression pattern under environmental stresses as well as at
various rice developmental stages was analyzed. A total of 35 OsGATA TFs encoded from 28 loci were found to be randomly distributed on rice chromosomes and were categorized into seven subfamilies. Members of the OsGATA gene family were differentially expressed under abiotic stresses. OsGATA23a is multi-stress responsive as it showed high transcript levels induced by salinity, drought as well as ABA treatment. Overall, the present work in aimed at providing not only an insight into the diversity of OsGATA TFs, guides toward functional cataloging of OsGATA TFs in response to environmental signals, but to establish leads for understanding the mechanism governing abiotic stress adaptability in rice via GATA TFs.

AUTHOR CONTRIBUTIONS

PG carried out in silico analysis. PG and KN did transcript abundance analysis. PG and AP drafted the figures, tables, and manuscript. AP and SS-P conceived and designed the experiments. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

Authors would like to thank research funds received from The Department of Science and Technology, Ministry of Science and Technology under Indo-Sri Lanka Project. Award of Research Fellowship from Council of Scientific and Industrial Research (CSIR-India) to PG and SERR-National Post Doctoral Fellowship (N-PDF) to KN is gratefully acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2017.01944/full#supplementary-material

REFERENCES

Alves, M. S., Dadalto, S. P., Gonçalves, A. B., De Souza, G. B., Barros, V. A., and Fietto, L. G. (2013). Plant bZIP transcription factors shaped to pathogens: a review. Int. J. Mol. Sci. 14, 7815–7828. doi: 10.3390/ijms14047815

Ambawat, S., Sharma, P., Yadav, N. R., and Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: an overview. Physiol. Mol. Biol. Plants 19, 307–321. doi: 10.1007/s12298-013-0179-1

Arguello-Astorga, G., and Herrera-Estrella, L. (1998). Evolution of light-regulated plant promotors. Annu. Rev. Plant Biol. 49, 525–555. doi: 10.1146/annurev.arplant.49.1.525

Arora, R., Agarwal, P., Ray, S., Singh, A. K., Singh, V. P., Tyagi, A. K., et al. (2007). MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. BMC Genomics 8:242. doi: 10.1186/1471-2164-8-242

Behringer, C., and Schwechheimer, C. (2015). B-GATA transcription factors – insights into their structure, regulation, and role in plant development. Front. Plant Sci. 6:99. doi: 10.3389/fpls.2015.00090

Behringer, C., Bastakis, E., Ranftl, Q. L., Mayer, K. F., and Schwechheimer, C. (2014). Functional diversification among the family of B-GATA transcription factor through the leucine-leucine-methionine domain. Plant Physiol. 166, 293–305. doi: 10.1104/pp.114.246660

Chen, C. H., Ringelberg, C. S., Gross, R. H., Dunlap, J. C., and Loros, J. J. (2009). Genome-wide analysis of light-inducible responses reveals hierarchical light signaling in Neurospora. EMBO J. 28, 1029–1042. doi: 10.1038/emboj.2009.54

Chen, H., Lai, Z., Shi, J., Xiao, Y., Chen, Z., and Xu, X. (2010). Roles of Arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscise acid and abiotic stress. BMC Plant Biol. 10:281. doi: 10.1186/1471-2229-10-281

Cheong, Y. H., Chang, H. S., Gupta, R., Wang, X., Zhu, T., and Luan, S. (2016). Signaling cross talk between biotic and abiotic stress responses in soybean. Abiotic Biotic Stresses Soybean Product. 1, 27–52. doi: 10.1016/B978-0-12-801536-0.00002-5

Chiang, Y. H., Zubo, Y. O., Tapken, W., Kim, H. J., Lavanway, A. M., Howard, L., et al. (2012). Functional characterization of the GATA transcription factors GNC and CGA1 reveals their key role in chloroplast development, growth, and division in Arabidopsis. Plant Physiol. 160, 332–348. doi: 10.1104/pp.112.198705

Crespo, J. L., Daicho, K., Ushimaru, T., and Hall, M. N. (2001). The GATA transcription factors GLN3 and GATI link TOR to salt stress in Saccharomyces cerevisiae. J. Biol. Chem. 276, 34441–34444. doi: 10.1074/jbc.M110 3601200

Daniel-Vedele, F., and Caboche, M. (1993). A tobacco cDNA clone encoding a GATA-1 zinc finger protein homologous to regulators of nitrogen metabolism in fungi. Mol. Gen. Genet. 240, 365–373.

García-Salcedo, R., Casamayor, A., Ruiz, A., González, A., Prista, C., Loureiro-Dias, M. C., et al. (2006). Heterologous expression implicates a GATA factor in regulation of nutrient metabolic genes and ion homeostasis in the halotolerant yeast Debaryomyces hansenii. Eukaryotic Cell 5, 1388–1398. doi: 10.1128/EC.00154-06

Gendron, J. M., Pruneda-Paz, J. L., Doherty, C. J., Gross, A. M., Kang, S. E., and Kay, S. A. (2012). Arabidopsis circadian clock protein, TOC1, is a DNA-binding transcription factor. Proc. Natl. Acad. Sci. U.S.A. 109, 3167–3172. doi: 10.1073/pnas.1200355109

Gupta, P., Sharma, R., Sharma, M. K., Sharma, M. P., Satpute, G. K., Garg, S., et al. (2016). Signaling cross talk between biotic and abiotic stress responses in soybean. Abiotic Biotic Stresses Soybean Product. 1, 27–52. doi: 10.1016/B978-0-12-801536-0.00002-5

Haas, H., Zadra, I., Stößler, G., and Angermayr, K. (1999). The Aspergillus nidulans GATA factor SREA is involved in regulation of siderophore biosynthesis and control of iron uptake. J. Biol. Chem. 274, 4613–4619. doi: 10.1074/jbc.274.8.4613

Hudson, D., Guevara, D. R., Hand, A. J., Xu, Z., Hao, L., Chen, X., et al. (2013). Rice cytokinin GATA transcription Factor1 regulates chloroplast development and plant architecture. Plant Physiol. 162, 132–144. doi: 10.1104/pp.113. 217265

Hudson, M., Ringli, C., Boylan, M. T., and Quail, P. H. (1999). The F AR1 locus encodes a novel nuclear protein specific to phytochrome A signaling. Proc. Natl. Acad. Sci. U.S.A. 96, 160, 332–348.

Hudson, M., Li, Z., Shi, J., Xiao, Y., Chen, Z., and Xu, X. (2010). Roles of Arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscise acid and abiotic stress. BMC Plant Biol. 10:281. doi: 10.1186/1471-2229-10-281

Jiang, J., Liu, X., Liu, G., Liu, C., Li, S., and Wang, L. (2017). Integrating omics and alternative splicing reveals insights into grape response to high temperature. Plant Physiol. 175:01305. doi: 10.1104/pp.16.01305

Karlen, R., Singla-Pareek, S. L., and Pareek, A. (2009). Histidine kinase and response regulator genes as they relate to salinity tolerance in rice. Funct. Integr. Genomics 9, 411–417. doi: 10.1007/s10142-009-0119-x

Kim, H. J., Lavanway, A. M., Howard, L., et al. (2012). Functional characterization of the GATA transcription factors GNC and CGA1 reveals their key role in chloroplast development, growth, and division in Arabidopsis. Plant Physiol. 160, 332–348. doi: 10.1104/pp.112.198705

Ko, L. J., and Engel, J. D. (1993). DNA-binding specificities of the GATA transcription factor family. Mol. Cell. Biol. 13, 401–4022. doi: 10.1128/MCB.13.7.4011

Kobayashi, K., and Masuda, T. (2016). Transcriptional regulation of tetrapyrrole biosynthesis in Arabidopsis thaliana. Front. Plant Sci. 7:1811. doi: 10.3389/fpls.2016.01811

Kobayashi, K., Ohnishi, A., Sasaki, D., Fujii, S., Iwase, A., Sugimoto, K., et al. (2017). Shoot removal induces chloroplast development in roots via cytokinin signaling. Plant Physiol. 173, 2340–2355. doi: 10.1104/pp. 16.01368
Krishna, S. S., Majumdar, I., and Grishin, N. V. (2003). Structural classification of zinc fingers SURVEY AND SUMMARY. Nucleic Acids Res. 31, 532–550. doi: 10.1093/nar/gkg161

Kumari, S., Sabharwal, V. P., Kushwaha, H. R., Sopory, S. K., Singla-Pareek, S. L., and Pareek, A. (2009). Transcriptome map for seedling stage specific salinity stress response indicates a specific set of genes as candidate for saline tolerance in *Oryza sativa* L. Funct. Integr. Genomics 9:109. doi: 10.1007/s10725-008-9088-5

Li, J., Li, G., Wang, H., and Wang Deng, X. (2011). Phytochrome signaling mechanisms. *Arabidopsis Book* 9:e0148. doi: 10.1199/tab.0148

Lin, R., Teng, Y., Park, H. J., Ding, L., Black, C., Fang, P., et al. (2008). Discrete and essential roles of the multiple domains of Arabidopsis PHY3 in mediating phytochrome A signal transduction. *Plant Physiol.* 148, 981–992. doi: 10.1104/pp.108.120436

Lindemose, S., O’Shea, C., Jensen, M. K., and Skriver, K. (2013). Structure, function, and networks of transcription factors involved in abiotic stress responses. *Int. J. Mol. Sci.* 14, 5842–5838. doi: 10.3390/ijms14035842

Liu, P. P., Koizuka, N., Martin, R. C., and Nonogaki, H. (2005). The BME3 (Blue Microlyl End 3) GATA zinc finger transcription factor is a positive regulator of Arabidopsis seed germination. *Plant J.* 44, 960–971. doi: 10.1111/j.1365-3131.2005.02588.x

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. *Methods* 25, 402–408. doi: 10.1016/s1046-6566(01)00416-2

Luo, X. M., Lin, W. H., Zhu, S., Zhu, J. Y., Sun, Y., Fan, X. Y., et al. (2010). Integration of light-and brassinosteroid-signaling pathways by a GATA transcription factor in Arabidopsis. *Dev. Cell* 19, 872–883. doi: 10.1016/j.devcel.2010.02.023

Marzluf, G. A. (1997). Genetic regulation of nitrogen metabolism in the fungi. *Microbiol. Mol. Biol. Rev.* 61, 17–32.

Marzluf, G. A. (2004). “Multiple fungal GATA transcription factors and combinatorial gene regulation,” in *Biochemistry and Molecular Biology, The Mycota III*, eds R. Brambl, and G. A. Marzluf (Berlin Heidelberg: Springer), 111–119.

Más, P., Kim, W. Y., Somers, D. E., and Kay, S. A. (2013). A suite of new genes defining salinity stress tolerance in *Arabidopsis thaliana*. *Plant Cell Tissue Organ Cult.* 114, 532–550. doi: 10.1007/s12229-013-0332-3

Mizuochi, K., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2012). AP2/ERF family transcription factors. *Mol. Cell. Biol.* 32, 31163.x

Nakamura, T., and Takatsuji, H. (1998). Zinc-finger transcription factors in plants. *Trends Plant Sci.* 3, 371–377. doi: 10.1016/s1369-5266(98)01163-x

Peirel, A., Singh, M., and Bentsink, M. (2006). Whole-genome analysis of *Oryza sativa* reveals similar architecture of two-component signaling machinery with Arabidopsis. *Plant Physiol.* 142, 380–397. doi: 10.1104/pp.106.086371

Panda, S. P., and Somssich, I. E. (2009). The role of WRKY transcription factors in plant immunity. *Plant Physiol.* 150, 1648–1655. doi: 10.1104/pp.109.138990

Pandey, S. P., and Somssich, I. E. (2009). The role of WRKY transcription factors in plant immunity. *Plant Physiol.* 150, 1648–1655. doi: 10.1104/pp.109.138990

Pareek, A., Singh, A., Kumar, M., Kushwaha, H. R., Lynn, A. M., and Singla-Pareek, S. L. (2006). Complex alternative splicing. *Adv. Exp. Med. Biol.* 623, 50–63. doi: 10.1007/978-0-387-77374-2_4
Xu, X., and Kim, S. K. (2012). The GATA transcription factor egl-27 delays aging by promoting stress resistance in *Caenorhabditis elegans*. *PLoS Genet.* 8:e1003108. doi: 10.1371/journal.pgen.1003108

Zhang, C., Hou, Y., Hao, Q., Chen, H., Chen, L., Yuan, S., 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. doi: 10.1371/journal.pone.0125174

Zhao, Y., Medrano, L., Ohashi, K., Fletcher, J. C., Yu, H., Sakai, H., et al. (2004). HANABA TARANU is a GATA transcription factor that regulates shoot apical meristem and flower development in Arabidopsis. *Plant Cell* 16, 2586–2600. doi: 10.1105/tpc.104.024869

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

*Copyright © 2017 Gupta, Nutan, Singla-Pareek and Pareek. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*