JIPB Journal of Integrative Plant Biology

Journal of Integrative Plant Biology 2011, **53** (3): 212–231

Research Article

Genome-wide Expansion and Expression Divergence of the Basic Leucine Zipper Transcription Factors in Higher Plants with an Emphasis on Sorghum

Jizhou Wang[1], Junxia Zhou[1], Baolan Zhang[1], Jeevanandam Vanitha[2], Srinivasan Ramachandran[1,2*] and Shu-Ye Jiang[1,2*]

[1]Institute of Botany and Temasek Life Sciences Laboratory Joint Research & Development Laboratory, Institute of Botany, the Chinese Academy of Sciences, Beijing 100093, China

[2]Temasek Life Sciences Laboratory, 1 Research Link, the National University of Singapore, Singapore 117604, Singapore

*Corresponding author

Tel: +65 6872 7489; Fax: +65 6872 7007; E-mail: sri@tll.org.sg and shuye@tll.org.sg

Articles can be viewed online without a subscription.

Available online on 7 December 2010 at www.jipb.net and www.onlinelibrary.wiley.com

doi: 10.1111/j.1744-7909.2010.01017.x

Abstract

Plant bZIP transcription factors play crucial roles in multiple biological processes. However, little is known about the sorghum *bZIP* gene family although the sorghum genome has been completely sequenced. In this study, we have carried out a genome-wide identification and characterization of this gene family in sorghum. Our data show that the genome encodes at least 92 bZIP transcription factors. These *bZIP* genes have been expanded mainly by segmental duplication. Such an expansion mechanism has also been observed in rice, arabidopsis and many other plant organisms, suggesting a common expansion mode of this gene family in plants. Further investigation shows that most of the *bZIP* members have been present in the most recent common ancestor of sorghum and rice and the major expansion would occur before the sorghum-rice split era. Although these *bZIP* genes have been duplicated with a long history, they exhibited limited functional divergence as shown by nonsynonymous substitutions (*Ka*)/synonymous substitutions (*Ks*) analyses. Their retention was mainly due to the high percentages of expression divergence. Our data also showed that this gene family might play a role in multiple developmental stages and tissues and might be regarded as important regulators of various abiotic stresses and sugar signaling.

Wang J, Zhou J, Zhang B, Vanitha J, Ramachandran S, Jiang SY (2011) Genome-wide expansion and expression divergence of the basic leucine zipper transcription factors in higher plants with an emphasis on sorghum. *J. Integr. Plant Biol*. **53**(3), 212–231.

Introduction

Transcription factors (TFs) are sequence-specific binding proteins that bind to the promoter regions of target genes to regulate their transcription. Total TFs encoded by a eukaryotic genome account for 3.5–7.0% of estimated total numbers of genes in a genome (Riechmann et al. 2000; Riechmann and Ratcliff 2000). Based on the reported data, the arabidopsis, rice, sorghum and maize genomes encode 1510–1581, 1611, 2448 and 3337 TFs, respectively (Riechmann et al. 2000; Gong et al. 2004; Guo et al. 2005; Iida et al. 2005; Xiong et al. 2005; Yilmaz et al. 2009). These TFs can be classified into 40–60 families based on their primary and/or three-dimensional structure similarities in the DNA-binding and multimerization domains (Wingender et al. 2001; Warren 2002; Yilmaz et al. 2009). Among them, the basic leucine zipper (bZIP) family consists of considerable numbers of TFs with two structural features located on a contiguous alpha-helix (Jakoby et al. 2002). The basic region is highly conserved and consists of around 16 amino acid residues with an invariant N-x7-R/K motif

for nuclear localization and DNA binding. The leucine zipper is a dimerization motif and is less conserved, consisting of a heptad repeat of leucines or other bulky hydrophobic amino acids. Plant bZIP proteins preferentially bind to DNA sequences with an ACGT core. Binding specificity is regulated by flanking nucleotides. Currently, the bZIP family members have been identified or predicted in multiple eukaryotic genomes including plants, animals and yeasts (Riechmann et al. 2000; Jakoby et al. 2002; Iida et al. 2005; Nijhawan et al. 2008; Yilmaz et al. 2009). Arabidopsis bZIP TFs were classified into 10 groups based on sequence similarity of the basic region and the presence of additional conserved motifs in arabidopsis (Jakoby et al. 2002). An additional one and three groups have been identified by Nijhawan et al. (2008) in rice and by Corrêa et al. (2008) in multiple green plants, respectively, indicating a unified classification of angiosperm bZIP genes.

Although the identification or prediction of many bZIP TFs from different genomes, only a small part of them have been functionally characterized. Data show that bZIP TFs have been involved in various biological processes under normal and stressed growth conditions. Generally, bZIP TFs play important roles in organ and tissue differentiation (Walsh et al. 1998; Chuang et al. 1999; Abe et al. 2005; Silveira et al. 2007; Shen et al. 2007), cell elongation (Yin et al. 1997; Fukazawa et al. 2000), nitrogen/carbon and energy metabolism (Ciceri et al. 1999; Weltmeier et al. 2006; Baena-Gonzalez et al. 2007), unfolded protein response (Iwata and Koizumi 2005; Liu et al. 2007), seed storage protein gene regulation (Lara et al. 2003), somatic embryogenesis (Guan et al. 2009) and so on. On the other hand, accumulated data show that bZIP TFs have also been regarded as important regulators in response to various biotic/abiotic stresses and signaling such as pathogen defense (Zhang et al. 1993; Despres et al. 2000; Niggeweg et al. 2000b; Pontier et al. 2001; Thurow et al. 2005; Kaminaka et al. 2006), hormone and sugar signaling (Finkelstein and Lynch 2000; Niggeweg et al. 2000a; Uno et al. 2000; Nieva et al. 2005), light response (Wellmer et al. 1999; Osterlund et al. 2000; Ulm et al. 2004), osmotic stresses (Satoh et al. 2004; Weltmeier et al. 2006), and so on. Recently, reported data have emphasized the biological functions of these TFs in drought, high salinity and cold stresses. These data are from arabidopsis (Uno et al. 2000; Kim et al. 2004; Liu et al. 2008; Weltmeier et al. 2009; Yang et al. 2009; Yoshida et al. 2010), rice (Aguan et al. 1993; Gupta et al. 1998; Shimizu et al. 2005; Mukherjee et al. 2006; Xiang et al. 2008; Zou et al. 2008; Lu et al. 2009; Hossain et al. 2010a, 2010b; Yun et al. 2010), wheat (Kobayashi et al. 2008), tomato (Yáñez et al. 2009; Hsieh et al. 2010), soybean (Kim et al. 2001; Liao et al. 2008a, 2010b), pepper (Lee et al. 2006), bean (Rodriguez-Uribe and O'Connell 2006), barley (Xue and Loveridge 2004), maize (Kusano et al. 1995) and so on. In addition, recent data also highlighted the roles of bZIP TFs in sugar signaling pathway (Rook et al. 1998; Wiese et al. 2004; Hanson et al. 2008; Usadel et al. 2008; Rahmani et al. 2009; Weltmeier et al. 2009; Kang et al. 2010).

Although *bZIP* genes were frequently reported to be involved in various stress responses, little is known about the genome-wide expression patterns of this gene family in most of plant genomes except for the work from the rice *bZIP* gene family (Nijhawan et al. 2008). In addition, although large amounts of data have been reported on the molecular, biochemical and physiological analyses as mentioned above, less data are available about the expansion mechanism and evolutionary history of this gene family.

On the other hand, the sorghum genome has been sequenced using grain sorghum AT×623/BT×623 as a genome donor (Bedell et al. 2005; Paterson et al. 2009). Its small genome (around 730 Mb) makes itself as a model for functional genomics of C4 grasses (Paterson et al. 2009). Sorghum has a remarkable ability to endure both drought conditions and water-logging and it grows well on marginal lands (Bedell et al. 2005). Furthermore, sweet sorghum has been regarded as a biofuel crop of growing importance for ethanol production due to its high biomass yield and sugar content. However, in sorghum, to our knowledge, only one *bZIP* gene has been isolated, which is an *Opaque2* (maize) homolog gene (Vettore et al. 1998) and no other data are available in sorghum so far, which illustrates the expression regulation of this gene family under drought, high salinity and cold stresses as well as the role in sugar-mediated signaling. In addition, only limited data are available on genome-wide identification and their characterizations in the sorghum genome. Therefore, it is important for us to genome-widely identify and to annotate the sorghum *bZIP* gene family with relation to the regulation of abiotic stresses and/or sugar metabolism. In this study, we first report the identification of the *bZIP* family in sorghum based on its complete genome sequence analysis by comparing with the members from rice and *Arabidopsis*. Subsequently, we have analyzed their gene expansion mechanisms and substitution rates to shed light on their evolution. Consequently, we analyzed expression patterns of these family members in all three eukaryotes based on our experimental data or the publicly available expression data, at tissue levels and under various stress conditions to address their putative biological functions.

Results

Genome-wide identification, classification and gene structure of the *bZIP* gene family in the sorghum genome

Currently, genome-wide identification of bZIP TFs has been carried out in at least six green plant genomes including *Arabidopsis thaliana* (Jakoby et al. 2002) and *Oryza sativa*

(Nijhawan et al. 2008) as well as another four genomes including *Chlamydomonas reinhardtii*, *Ostreococcus tauri*, *Physcomitrella patens*, *Populus trichocarpa* (Corrêa et al. 2008). We have used both BLAST and profile hidden Markov model (HMM, see Methods) searches to achieve all genes encoding bZIP TFs in the sorghum genome. To better explore their expansion mechanisms, evolutionary history and expression divergence, two other genomes were also selected for comparative analyses including rice (model plant for C3 monocot) and arabidopsis (model plant for dicot) genomes. We have used key members for the searches (Methods) and the obtained new members were also used for new searches. Thus, after multiple cycles of searches, totals of 112, 110 and 93 putative bZIP genes have been detected in sorghum, rice and arabidopsis, respectively. These members were then subjected to the Pfam (Finn et al. 2006; http://pfam.sanger.ac.uk/) and SMART (Letunic et al. 2009; http://smart.embl-heidelberg.de/) domain searches to confirm the presence of the bZIP-related domains. Some of the members contain no typical domain structure and have no expression evidence with the characters of pseudogenes. Due to the low feasibility of phylogenetic analyses by integrating these partial fragments, we removed these members from our analyses although we may underestimate the rate of gene duplication. Thus, our analyses reveal that sorghum, rice and arabidopsis genomes encode a total of 92, 88 and 72 members of the *bZIP* gene family, respectively. Their locus name, physical position and annotated protein sequences were presented in Supporting Table S1 (sorghum), S2 (rice) and S3 (arabidopsis).

Since the rice *bZIP* gene family was previously identified on a genome-wide level (Corrêa et al. 2008; Nijhawan et al. 2008), we made a detailed comparison with the sorghum *bZIP* family. Nijhawan et al. (2008) reported the presence of 89 *bZIP* genes and two of them *LOC_Os02g33560* and *LOC_Os12g09270* were not annotated in this study. The former contains no bZIP domain by the Pfam domain search with an e-value at 0.01 as the cutoff level and the latter was not maintained at the latest version (release 6.1) of the Michigan State University (MSU)/The Institute of Genome Research (TIGR) rice genome annotation database (http://rice.plantbiology.msu.edu/; Yuan et al. 2005; Ouyang et al. 2007). Thus, these two members were not included for further analysis. Furthermore, we have identified an additional member *LOC_Os12g09250* with a typical bZIP domain structure in its protein sequences. Corrêa et al. (2008) also reported 89 bZIP TFs from the *japonica* genome. However, four members including *LOC_Os03g59460*, *LOC_Os06g50480*, *LOC_Os06g50830*, and *LOC_Os12g09250* were not identified in the report, which was included in this study and by Nijhawan et al. 2008. Besides *LOC_Os02g33560* and LOC_*Os12g09270*, Corrêa et al. (2008) have identified three more members including *LOC_Os02g22280*, *LOC_Os11g04390*, and *LOC_Os12g06010*, which were not included in this study due to the lack of complete bZIP domain structure based on the Pfam and SMART database searches. More detailed comparison has been presented in Supporting Table S2.

In arabidopsis, a total of 75 and 77 *bZIP* gene family members have been identified by Jakoby et al. (2002) and Corrêa et al. (2008), respectively. However, some of them were detected with the incomplete bZIP domain or lacked this domain based on the Pfam or SMART domain searches with e-value less than 0.01. Thus, these members were excluded for further analysis. As a result, only 72 members were detected in this study and detailed information has been listed in the Supporting Table S3.

The amino acid sequences of the basic region of all sorghum bZIP TFs were then submitted to the phylogenetic analysis (see Methods) for clustering the sorghum family members (**Figure 1**). Based on the analysis, a total of seven classes of bZIP TFs have been identified. In arabidopsis, a total of 10 groups of bZIP TFs have been classified and they were named as group A to I and S (Jakoby et al. 2002) and corresponding groups in sorghum have also been labeled in **Figure 1**. Compared with the nomenclature in arabidopsis, both group I and E have been combined into the class 1; class 2 consists of group B, D, F and H; and group S has been divided into classes 6 and 7 (**Figure 1**).

In order to obtain some insight into gene structures of the 92 sorghum *bZIP* genes, their exon-intron organization was also investigated (Supporting Figure S1). We have detected a total of 17 *bZIP* genes with no intron, accounting for 18.5% of total *bZIP* genes. Most of these intronless genes were clustered into the class 6 (Supporting Figure S1). Similar cases have also been observed in arabidopsis and in rice (Jakoby et al. 2002; Nijhawan et al. 2008), suggesting the evolutionary conservation. Among the intron-containing *bZIP* genes, the number of introns in their open reading frames varied from 1 to 14. They were distributed into different classes of the *bZIP* family. In arabidopsis and rice, a maximum of 12 introns have been detected (Jakoby et al. 2002; Nijhawan et al. 2008). On the other hand, to analyze the evolution of exon-intron structures, we further carried out the investigation of patterns of intron positions in this family. The intron positions were obtained from their corresponding protein alignment. Only introns in the bZIP domain region were exploited since the remaining regions were more variable. A total of 92 sorghum *bZIP* genes were submitted to such analysis and their intron positions were presented in Supporting Figure S2. A maximum of three introns have been detected in the total of 92 sorghum bZIP domains. These introns were mainly distributed on three hot regions as shown in Supporting Figure S3A. Based on the intron positions in these three hot regions, a total of eight different patterns of intron positions have been detected (Supporting Figure S3B). Except for the 24 intronless *bZIP* genes in the domain region (pattern

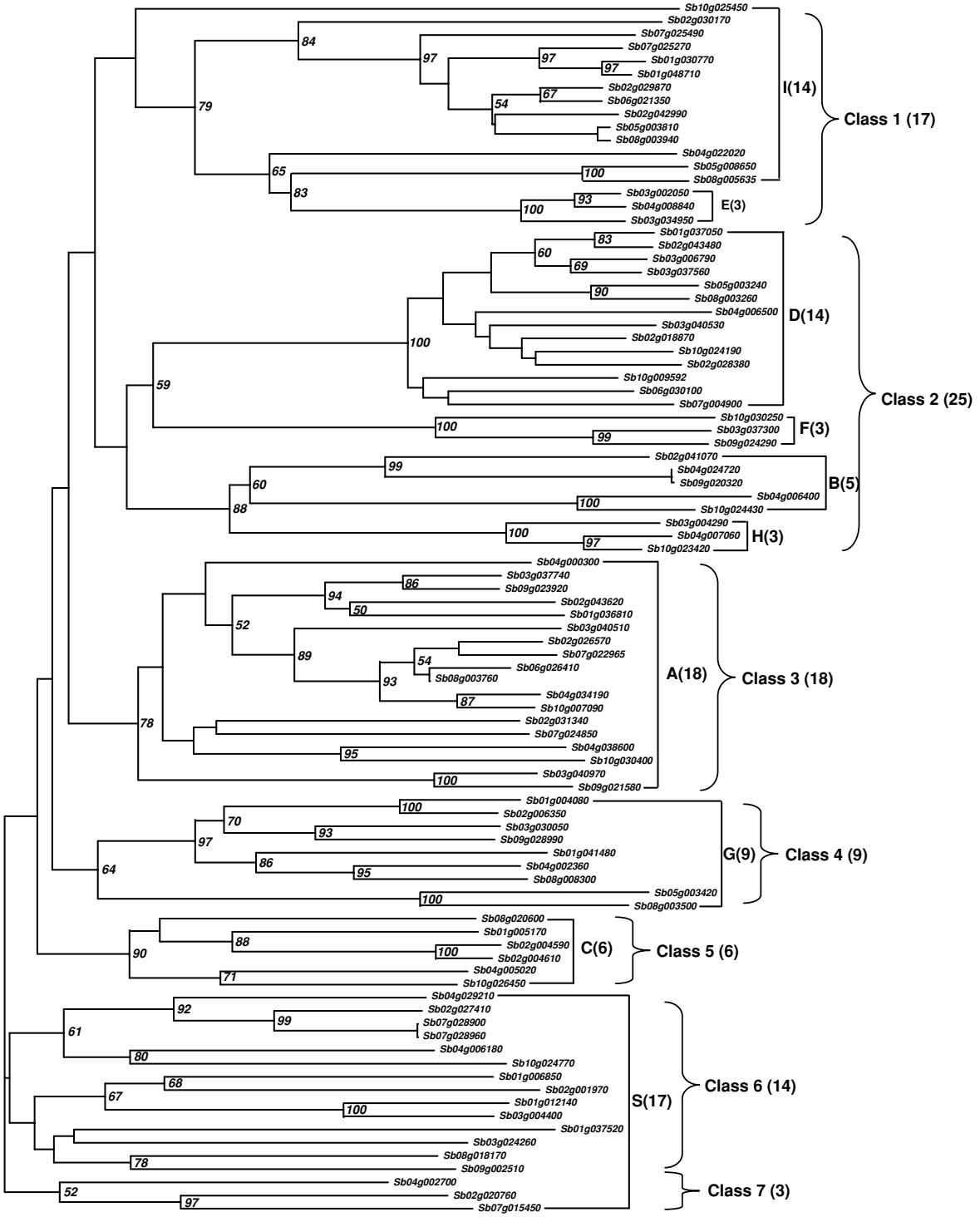


Figure 1. Classification of the *bZIP* gene family in sorghum.

Phylogenetic analysis was carried out using the bZIP transcription factor (TF) domain amino acid sequences. Trees were generated using the bootstrap method with a heuristic search by the PAUP 4.0b8 program. Generated trees were similar to those from Bayesian analyses. A total of seven classes of *bZIP* genes have been clustered. The classification system was compared with the published system where 10 groups of bZIP TFs were defined according to the report by Jakoby et al. (2002). In each group, additional conserved motifs have been detected using the MEME tools (http://meme.sdsc.edu/meme4_4_0/cgi-bin/meme.cgi).

8), most of the bZIP domain contains two introns (patterns 2, 3 and 4) and the remaining bZIP domains consist of one or three introns in their corresponding genome sequences. The genes with shared intron positions might have been evolved from a common ancestor (Bagavathi and Malathi 1996). Thus, sorghum *bZIP* genes might have been evolved from multiple ancestral units after divergence from its most recent common ancestor (MRCA). However, intron gain and loss might also contribute to the divergence of intron positions (Rogozin et al. 2005).

Expansion patterns and mechanisms of the *bZIP* family

To understand the expansion mechanism in detail, we analyzed the contribution of tandem/segmental duplications to the expansion of this gene family. The tandem cluster is defined as a region containing *bZIP* members with less than 10 genes apart between neighboring members as suggested by Shiu et al. (2004). We first physically mapped all sorghum members in its chromosomes and then determined members with tandem duplication. Those tandemly duplicated members were highlighted with red fonts as shown in **Figure 2A**. Totally, we detected only three pairs of tandem duplicates, indicating the limited contribution of tandem duplication to the gene family expansion. A similar result has been observed in the rice and arabidopsis genomes (data not shown). We then carried out a genome-wide identification of segmentally duplicated *bZIP* genes in sorghum (see Methods). We have detected 49 sorghum *bZIP* genes being involved in segmental duplication, accounting for around 53% (49/92) of total *bZIP* genes (**Figure 2A**). Among these genes, some of them segmentally duplicated once and some of them duplicated twice or thrice. The duplication might occur within a chromosome or between chromosomes. These segmentally duplicated genes have contributed to the expansion of multiple classes of *bZIP* genes (**Figure 2B**). For the class 7, no member has been expanded by segmental duplication and only 21% of class 6 of *bZIP* genes were segmentally duplicated. For the remaining five classes of *bZIP* genes, 59–67% of them have been involved in segmental duplication.

Similar contribution rates of segmental duplication to the family expansion have also been observed in rice and arabidopsis. In rice, we have detected 52 out of 88 *bZIP* genes (59%) with relation to segmental duplication and one pair of genes *LOC_Os01g64000* and *LOC_Os01g64020* have been involved in tandem duplication. In arabidopsis, the ratio has been reduced to 54% (39 out of 72 *bZIP* genes) for segmental duplication and we have also detected two pairs of tandemly duplicated genes including *At2g21230* and *At2g21235* as well as *At5g06950* and *At5g06960*. Thus, segmental duplication has been regarded as a major driver to contribute to the expansion of this gene family in the sorghum, rice and arabidopsis genomes.

To explore other mechanisms to contribute to the expansion of this gene family, we also investigated the transposon/retrotransposon-mediated duplication of the *bZIP* genes in sorghum. For each gene, we have analyzed an around 100 Kb sequence region from 50 Kb upstream to 50 Kb downstream of a *bZIP* gene. We first surveyed whether a *bZIP* gene was flanked by a transposon or a retrotransposon. The examined transposable elements included retrotransposon (LTR- or non-LTR-related) and DNA transposon (*MULE*, *CACTA*, *hAT* and *Heliton*). We examined these elements according to our previous description (Jiang et al. 2009). This analysis showed that no such element has been detected to contain a *bZIP* gene. Similar results have also been observed in rice and arabidopsis. Therefore, no significant contribution of transposable elements to the *bZIP* gene expansion has been observed and segmental duplication could be regarded as the major mechanism to the expansion of this gene family in multiple organisms.

Collinearity among sorghum, rice and *Arabidopsis* and expansion era

Our data showed that the bZIP TFs in sorghum, rice and arabidopsis exhibited similar expansion patterns. Thus, a question may arise that the gene family might practice a large scale of expansion during monocot-dicot or rice-sorghum split. We first analyzed the collinearity between sorghum and arabidopsis. The data showed that only six arabidopsis *bZIP* genes can be detected with seven orthologs in the rice genome (blue lines in **Figure 3**). A similar result was observed when we aligned the arabidopsis *bZIP* genes to the rice genome (red lines in **Figure 3**). Therefore, the large scale of expansion may not occur before monocot-dicot split. We then analyzed the collinearity between sorghum and rice. Our data showed that 72 sorghum *bZIP* genes could find their corresponding 66 rice orthologs (black lines in **Figure 3**). The data suggest that the large scale of expansion might occur before the era of sorghum-rice split.

To explore the history of the family expansion, we then aligned the domain sequences of the bZIP members from sorghum, rice and arabidopsis. The alignments were used to generate the phylogenetic tree shown in Supporting Figure S4. We then broke down the phylogeny into ancestral units, which were present right before the monocot-dicot (sorghum/rice-arabidopsis) or sorghum-rice split according to the description by Shiu et al. (2004). Based on this analysis, we have identified 75 ancestral units between sorghum and rice (blue circles in Supporting Figure S4). The data suggested that most of *bZIP* members have been present in the MRCA of sorghum and rice and very limited expansion has occurred after the split of

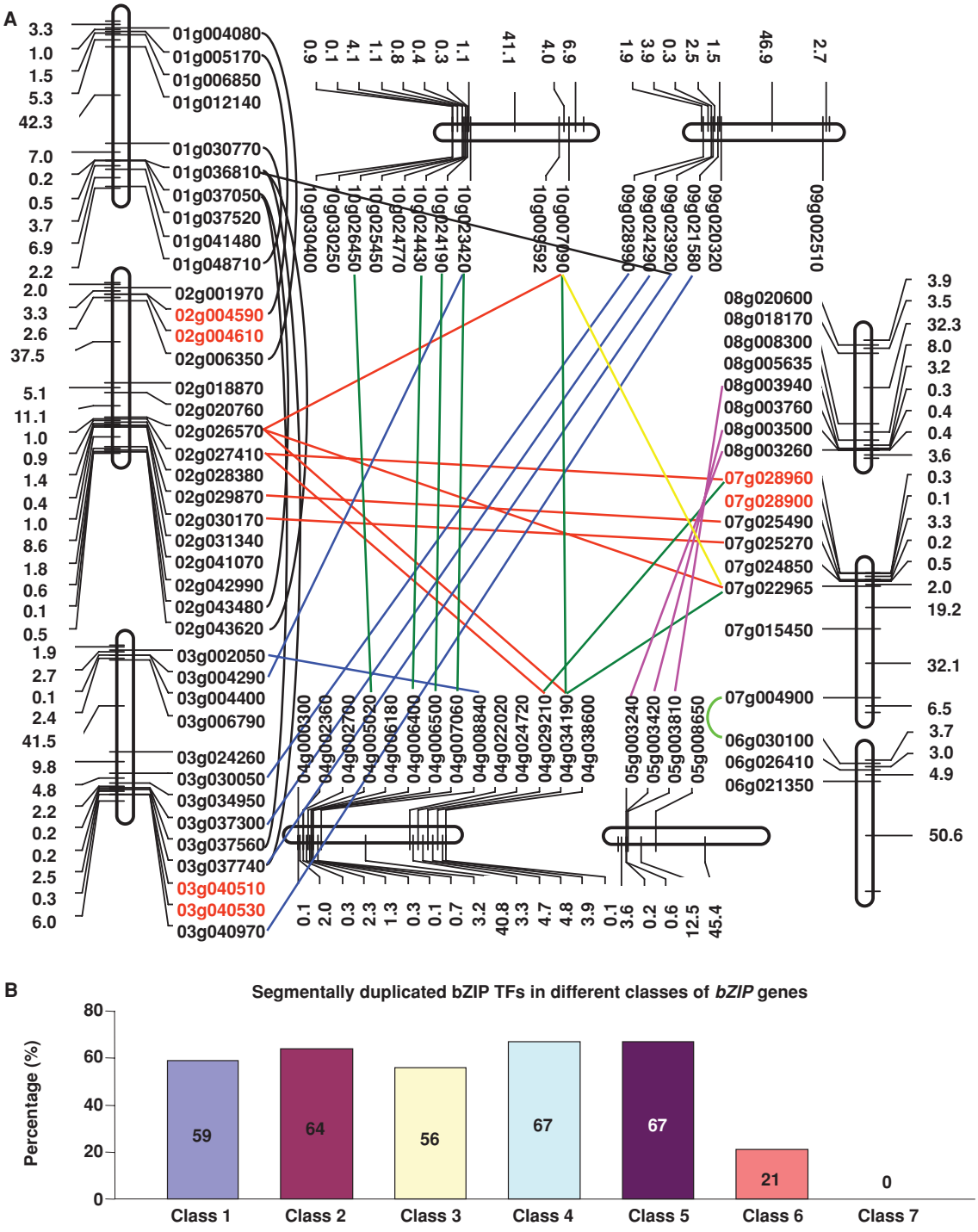


Figure 2. Chromosomal distribution and detection of duplicated genes in the sorghum *bZIP* gene family.

(A) Chromosomal distribution and detection of duplicated genes. Chromosomal mapping was based on the physical position (Mb) of duplicated genes in 10 sorghum chromosomes. The last number in each chromosome represents the physical position (Mb) of mapped genes. The penultimate number shows the physical distance to the last gene. The rest could be reduced by analogy. Expanded sorghum *bZIP* genes from tandem duplications are marked with red fonts. The segmentally duplicated genes are marked with normal black fonts. Different colored lines indicate the intra-genome syntenic relationships by segmental duplication. The prefix "Sb" in each sorghum locus name has been omitted for convenience.

(B) The contribution rates of segmental duplication to various classes of the sorghum *bZIP* genes.

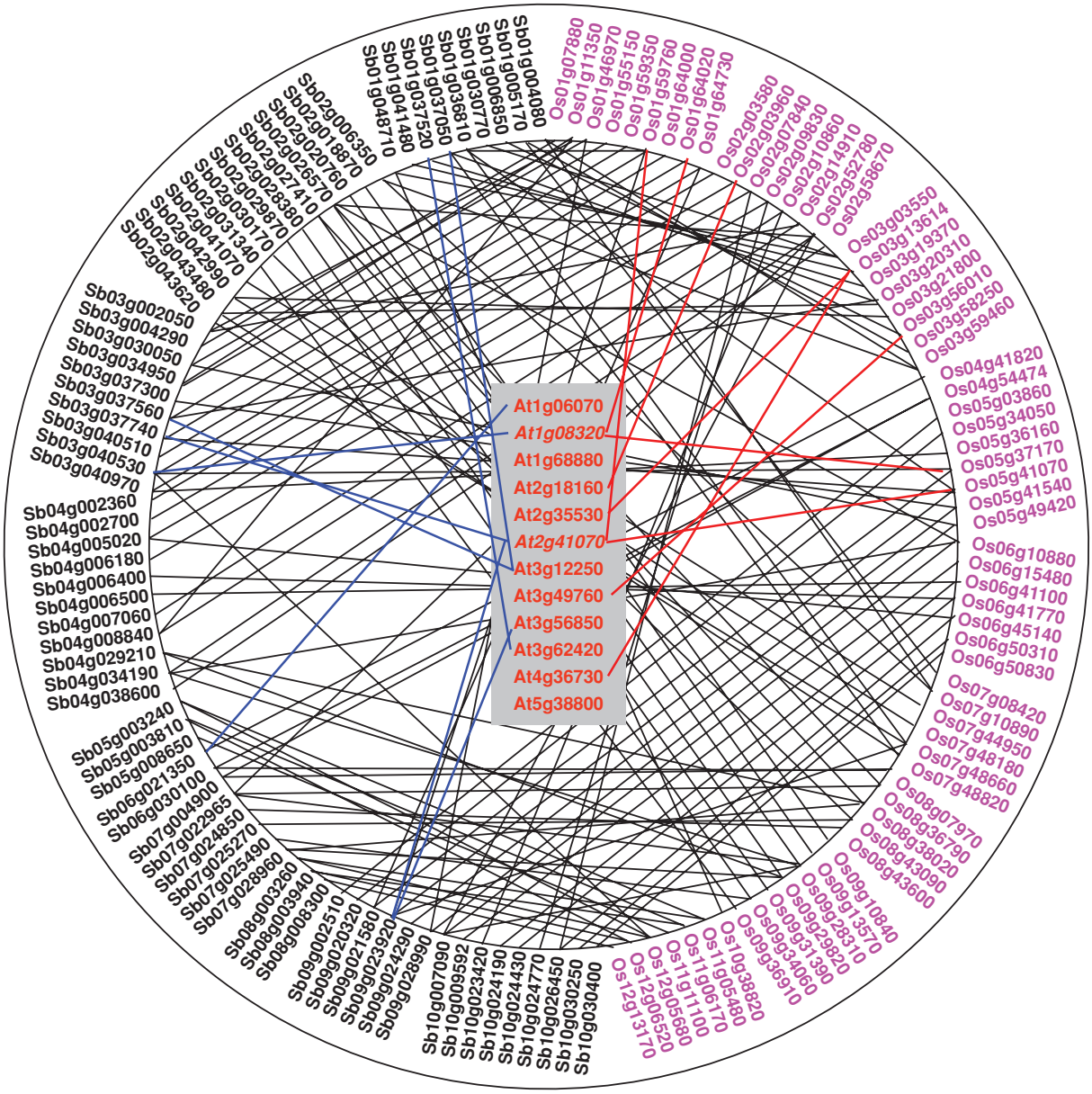


Figure 3. Collinearity relationships of *bZIP* genes among sorghum, rice and arabidopsis.

Orthologous *bZIP* genes among sorghum, rice and arabidopsis were identified by the program OrthoMCL (Li et al. 2003) according to the OrthoMCL standard settings (Blastp e-value < 1e-05). Locus names from sorghum, rice and arabidopsis are marked with black, pink and red fonts, respectively. The prefix "LOC_" in each rice locus name has been omitted for convenience. Black, blue and red lines indicate the orthologous pairs between sorghum and rice, sorghum and arabidopsis as well as rice and arabidopsis, respectively.

sorghum from rice. On the other hand, only 28 bZIP TFs were identified in the monocot-dicot MRCA (Supporting Figure S4), suggesting that the major expansion of this gene family would not occur before the divergence between monocot and dicot plants.

Expression patterns of the sorghum classes 5, 6 and 7 of *bZIP* genes in various developmental stages of tissues

It is imperative to further improve sweet sorghum cultivars in their sugar content and abiotic stress tolerance for their

plantation as a bio-ethanol crop. Group C and S of *bZIP* genes have been frequently reported to be involved in sugar signaling and abiotic stress regulation (Jakoby et al. 2002; Hanson and Smeekens 2009; Smeekens et al. 2010). Therefore, we further investigated the expression patterns and stress regulation of these groups of members. In sorghum, a total of six group C (class 5) and 17 group S (class 6 and 7) *bZIP* genes have been classified (**Figure 1**). However, two *bZIP* genes *Sb07g028900* and *Sb07g028960* from class 6 were located at different genomic positions but they showed 100% identities in their genome sequences, indicating a recent duplication event. Thus, only one set of primers was designed for their expression analysis. We analyzed the transcript profiles of these genes under nine different tissues from multiple developmental stages using the reverse-transcription polymerase chain reaction (RT-PCR) technique. These tissues include young and mature leaves, roots, panicles and seeds as well as stems. Among the six *bZIP* genes from group C (class 5), only one member *Sb02g004590* showed expression in all tested tissues and the remaining five members exhibited no or very week transcript in young/mature leaves (**Figure 4A**), suggesting that most of group C (class 5) *bZIP* genes may function in non-leaf tissues. Two of them also showed no expression in stem tissues including *Sb02g004610* and *Sb08g020600*.

Among the 16 tested *bZIP* genes from group S (classes 6 and 7), only seven of them can be detected with expression signals among nine different tissues tested (**Figure 4A**). Two of them *Sb01g006850* and *Sb04g006180* were expressed in all tested tissues. Three of them, including *Sb03g004400*, *Sb04g029210* and *Sb07g015450*, showed no or very week expression in leaf tissues. The remaining two including *Sb02g020760* and *Sb10g024770* exhibited no expression in the stem tissue. To further investigate the expression patterns of *bZIP* genes in different tissues, quantitative RT-PCR (qRT-PCR) was carried out. The analysis revealed three more *bZIP* genes from class 6 with significant expression signals (data not shown). These genes include *Sb03g024260* with the highest expression at immature panicles, *Sb08g018170* with higher expression abundance at seeds and *Sb09g002510* with higher expression abundance at leaves.

Expression regulation of the sorghum classes 5, 6 and 7 of *bZIP* genes (group C and S) under various abiotic stresses and sugar treatments

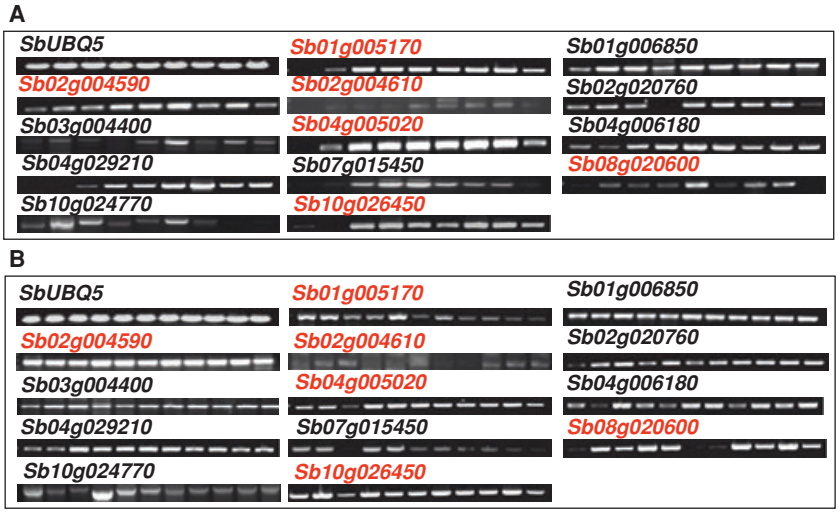
We further carried out the expression analysis of these three classes of *bZIP* genes under drought, high salinity and cold stresses as well as under glucose and sucrose treatments. We used PEG (polyethylene glycol, 6000) to mimic drought stress. We first used RT-PCR to examine the expression patterns of these genes under these stresses/treatments. The result showed that some *bZIP* genes from these three classes were regulated by single to multiple stresses and some of them exhibited no detectable difference in their expression abundance under these stresses/treatments (**Figure 4B**). For example, the gene *Sb01g005170* from group C (class 5) was statistically downregulated by PEG (2 h), NaCl (2 h), cold (2 and 8 h), glucose (2 and 6 h) and sucrose (2 and 6 h); the gene *Sb02g020760* from group S (class 7) was upregulated by all stresses/treatments at all treated time points. To further explore their expression regulation under these stresses/treatments, qRT-PCRs were carried out to detect their transcript abundance under these stresses/treatments. The *t*-test analysis was used for detecting differentially expressed genes under these stresses/treatments. Based on the analysis, 10 out of a total 16 expressed genes were detected to be down- or upregulated either by drought/high salinity/cold or by glucose/sucrose treatment (**Figure 4C**). Most of them were upregulated and only two *bZIP* genes were downregulated. The gene *Sb07g015450* was downregulated by both glucose and sucrose treatments and the gene *Sb10g024770* was downregulated only under high salinity stress.

Expression patterns of arabidopsis and rice *bZIP* genes under normal and stressed conditions

Since a large amount of full-length cDNA or expressed sequence tags (ESTs) is available in arabidopsis and rice, we investigated the expression of all *bZIP* genes in these two organisms by examining the presence of a full-length cDNA or EST. The result showed that up to 97.2% of arabidopsis *bZIP* genes contained their corresponding full-length cDNAs or ESTs (**Figure 5A**). In rice, 82.8% of *bZIP* genes were expressed with either full-length cDNA or EST as evidence (**Figure 5A**). In contrast, only 59.8% of sorghum *bZIP* genes have their corresponding full-length cDNAs or ESTs (**Figure 5A**).

In addition, we have surveyed the expression regulation under various abiotic and biotic stresses in arabidopsis. We have detected a total of 29 *bZIP* genes with differential expression abundance under drought, high salinity and/or cold stresses (**Figure 5B**). Among them, six genes were regulated by all three kinds of stresses, 14 genes were regulated by both drought and high salinity, seven genes were regulated by only high salinity and the remaining one gene was differentially expressed under cold stress. However, we detected only five genes with differentially regulated expression patterns under biotic stresses (**Figure 5C**). One of them, *At3g30530* was regulated by both fungus pathogens *Botrytis cinerea* (*Bcin*) and *Erysiphe orontii* (*EOr*). The remaining four genes were regulated either by *EQr* for *At1g68880* and *At2g41070* or by *Bcin* for *At2g42380* and *At3g58120*.

Besides arabidopsis, we have also investigated the expression profile of rice *bZIP* genes under both abiotic and biotic stresses. Under drought, high salinity and cold stresses, a



C

Locus	PEG	NaCl	Cold	Glucose	Sucrose
Sb01g005170	–	–	–	–	–
Sb01g006850	ND	ND	ND	ND	ND
Sb02g004590	ND	ND	ND	ND	ND
Sb02g004610	+	ND	ND	+	+
Sb02g020760	+	+	+	+	+
Sb03g004400	+	ND	ND	+	+
Sb03g024260	ND	ND	ND	ND	ND
Sb04g005020	ND	+	+	+	+
Sb04g006180	+	ND	+	+	+
Sb04g029210	+	+	+	ND	ND
Sb07g015450	ND	ND	ND	–	–
Sb08g018170	ND	ND	ND	ND	ND
Sb08g020600	+	+	ND	+	+
Sb09g002510	ND	ND	ND	ND	ND
Sb10g024770	ND	+	ND	ND	ND
Sb10g026450	ND	ND	ND	ND	ND

Figure 4. Expression profiling of sorghum *bZIP* genes from group C and S under normal and stressed conditions.

The expression analysis was carried out by reverse transcription-polymerase chain reaction (RT-PCR) and was verified by quantitative RT-PCR (qRT-PCR). The red and black locus names indicate the *bZIP* genes from groups C and S, respectively. The amplification of the sorghum gene *SbUBQ5* was used as control for RT-PCR or to normalize the data for qRT-PCR analysis. **(A)** Expression patterns of group C and S *bZIP* genes among different tissues shown by RT-PCR. For each gene, amplified bands from left to right represent products from young leaves, mature leaves, young roots, mature roots, young panicles, mature panicles, young seeds, mature seeds and stems, respectively. **(B)** Transcription abundance of group C and S *bZIP* genes under abiotic or sugar treatments in sorghum. The amplified bands for each gene from left to right represent products from the following tissues: 1, control (un-treated tissue); 2 and 3, stressed by 30% polyethylene glycol (PEG) for 0.5 and 2.0 h, respectively; 4 and 5, stressed by 250 mM NaCl for 2 and 8 h, respectively; 6 and 7, stressed by cold at 4°C for 2 and 8 h, respectively; 8 and 9, treated by 5% glucose for 2 and 6 h, respectively; 10 and 11, treated by 5% sucrose for 2 and 6 h, respectively. **(C)** A summary of RT-PCR and qRT-PCR analysis of the sorghum group C and S *bZIP* genes under various abiotic and sugar treatments. "ND" indicates no significant difference in their expression shown by *t* test under various stresses; "+" indicates significant upregulation in their expression; "–" indicates significant downregulation in their expression.

total of 43 *bZIP* genes were differentially expressed by one, two or three of these stresses (**Figure 5D**). Among them, the transcript abundance of 11, three and six genes were differentially regulated by cold, drought and high salinity, respectively. Thirteen genes were significantly regulated by cold and drought stresses, four genes by both cold and high salinity and only one gene by both drought and NaCl. In addition, a total of five genes was regulated by all three stresses. Under biotic stresses, a

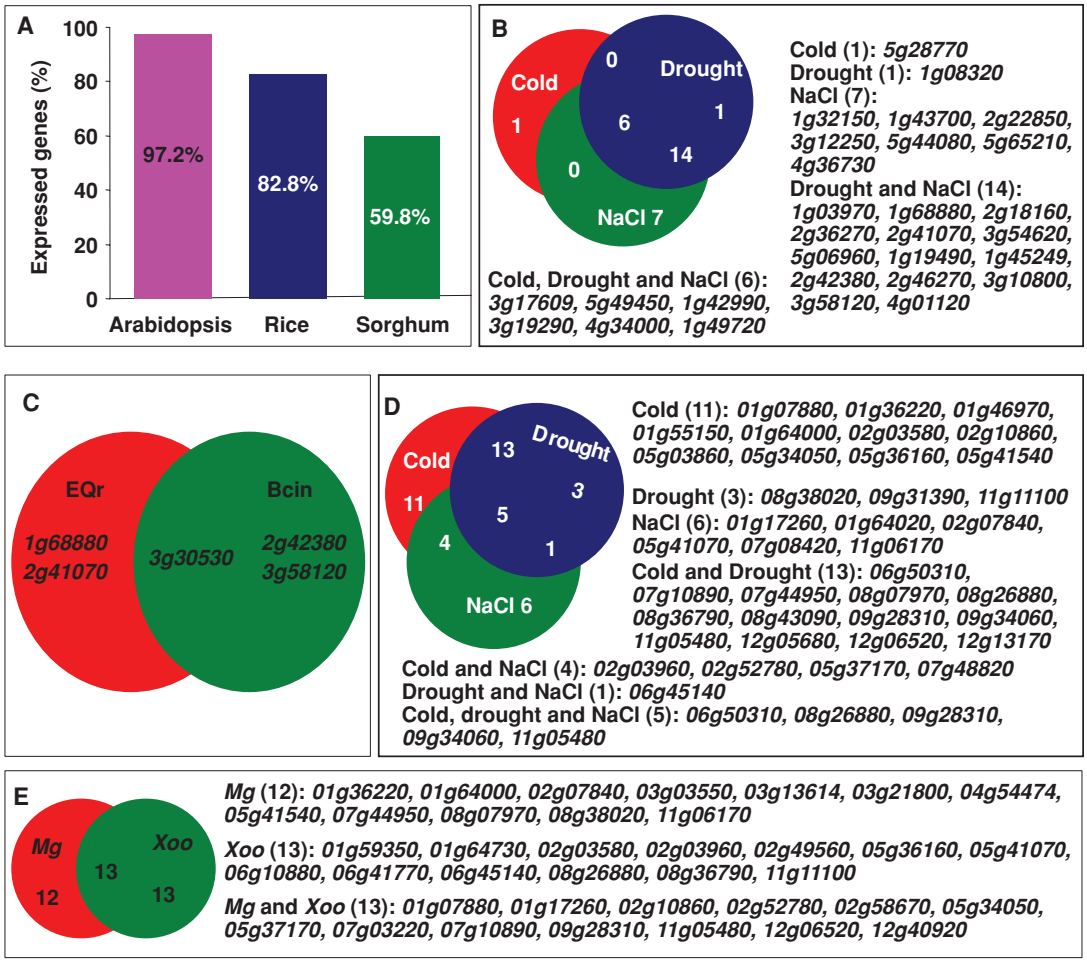


Figure 5. Expression evidence and stress regulation of *bZIP* genes.

(A) Expression evidence from full-length cDNA or expressed sequence tags (ESTs) in rice, arabidopsis and sorghum.
(B) and **(C)** Venn diagram of the expression of *bZIP* genes under various biotic and abiotic stresses in arabidopsis and rice, respectively. The arabidopsis expression data were achieved from the website http://www.Arabidopsis.org and from the report by Matsui et al. 2008. The rice expression data were downloaded from the rice MPSS database (http://mpss.udel.eud/rice/; Nobuta et al. 2007).

total of 38 *bZIP* genes were differentially expressed in their transcript abundance under the fungus pathogen *Megnaporthe grisea* (*Mg*) and/or the bacterium pathogen *Xanthomonas oryzae* pv *oryzae* (*Xoo*) (**Figure 5E**). Among them, 13 genes were regulated by both biotic stresses and the expression abundance of the remaining 25 genes was in response to either *Mg* or Xoo.

Non-synonymous substitutions (*Ka*)/synonymous substitutions (*Ks*) analyses and expression divergence

The *bZIP* gene family in sorghum, rice and arabidopsis has been expanded largely by segmental/tandem duplication. To explore how these duplicated genes survived after duplication, we first analyzed non-synonymous substitutions per site (*Ka*) and synonymous substitutions per site (*Ks*) as well as their *Ka*/*Ks* ratio. Each segmental/tandem duplication pair was aligned using their amino acid sequence, which formed the basis on calculating the *Ka*/*Ks* ratio. In rice, a total of 38 segmental duplication pairs and 1 tandem pairs were analyzed. The data showed that the *Ka*/*Ks* ratios from 37 segmental duplication pairs (small circles) and one tandem duplication pair (big circle) were less than 0.5 (**Figure 6A**). Only one segmental pair showed the *Ka*/*Ks* ratio with slightly larger than 0.5 (**Figure 6A**). The result suggested that no significant functional divergence has occurred after segmental/tandem duplications. We then further analyzed the expression divergence of these duplicated pairs. We surveyed the expression regulation under cold, drought and high salinity stresses as well as under *Mg* and *Xoo* stresses between duplicated pairs. Among 39

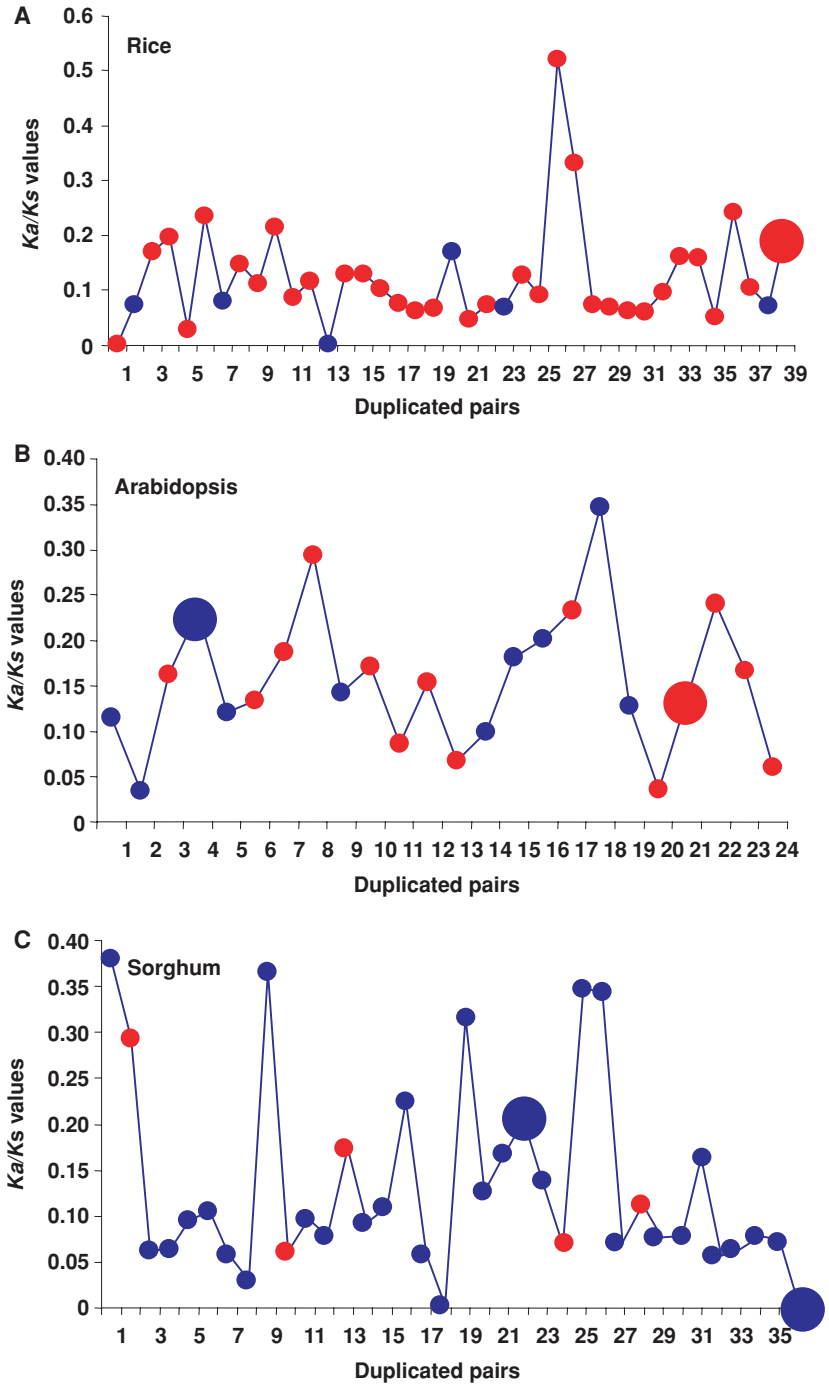


Figure 6. Expression and functional divergence of *bZIP* genes in rice, arabidopsis and sorghum.

Both tandem and segmental pairs were employed for detecting expression and functional divergence after duplication. Small and larger circles indicate pairs from segmental and tandem duplication. The expression divergence under drought, high salinity, cold stresses was evaluated in all three organisms including rice **(A)**, arabidopsis **(B)** and sorghum **(C)**. However, the expression divergence was evaluated under the pathogen *Mg* and *X*oo for rice (A), under *Bcin* and *Xoo* for arabidopsis (B) and under glucose and sucrose treatment for sorghum (C). The functional divergence was estimated by *Ka*/*Ks* ratios as described in Methods.

segmental/tandem pairs, 33 pairs (84.6%) exhibited significant difference in their transcript abundance under these abiotic and biotic stresses (red circles in **Figure 6A**). The data suggested that expression divergence should be regarded as the major driver to retain these genes survival after duplication.

A similar result has been observed in arabidopsis. We have analyzed total of 24 segmentally/tandemly duplicated pairs. The *Ka*/*Ks* ratios were all less than 0.35, suggesting a limited functional divergence after duplication. In contrast, up to 14 pairs (58.3%) showed significant expression divergence under both abiotic and biotic stresses. On the other hand, in sorghum, a total of 35 segmentally/tandemly duplicated pairs have been analyzed and all of the *Ka*/*Ks* ratios were less than 0.4, confirming that limited functional divergence has occurred after duplication. However, in sorghum, relatively less data are available for estimating the expression divergence. We have analyzed five segmental duplication pairs and all of them showed expression divergence (red circle in **Figure 6C**). All of these data suggested that the expression divergence but not the functional divergence should be regarded as the major mechanism for the family segmental/tandem expansion followed by survival retention.

Discussion

bZIP TFs ubiquitously exist in eukaryotes and prokaryotes

In this study, we have genome-widely investigated the *bZIP* gene family in sorghum, rice and arabidopsis. Besides these plant species, we have also surveyed the presence of this gene family in another 15 plant species including *Manihot esculenta, Ricinus comunis, Populus trichocarpa, Medicago truncatula, Glycine max, Cucumis sativus, Arabidopsis lyrata, Carica papaya, Vitis vinifera, Mimulus guttatus, Zea mays, Brachypodium distachyon, Selaginella moellendorffii, Physcomitrella patens,* and *Chlamydomonas reinhardtii*. Their draft genome sequences have been released by the Phytozome database (http://www.phytozome.net/). Our data showed that bZIP TFs exist not only in higher plants but also in green algae, moss and fern. In addition to this, our data showed that bZIP TFs could also be detected in other organisms such as mammal (*Mus musculus*), non-mammal (*Gallus gallus*), insect (*Drosophila melanogaster*), nematode (*Caenorhabditis elegans*), echinoderms (*Strongylocentrotus purpuratus*), urochordata (*Ciona intestinalis*), yeast (*Saccharomyces cerevisiae*), pathogen (*Ustilago maydis*), protozoa (*Dictyostelium discoideum* and *Leishmania braziliensis*), bacteria (*Bacillus licheniformis* and *Polaribacter dokdonensis*) and archaea (*Hyperthermus butylicus*) and so on. In addition, we have detected bZIP TF sequences from at least 30 additional plant species by BLAST searches against the National Centre for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/). All of these data suggested that bZIP TFs should be ubiquitous in eukaryotes and prokaryotes and they are encoded by an ancient gene family.

The evolutionary history of bZIP TFs

Our data showed that bZIP TFs could be detected in most of the tested organisms. In some Archaea species, only one member can be detected with the bZIP domain PF00170. The protein is usually annotated as proteasome-activating nucleotidase with the AAA domain (ATPase family associated with various cellular activities). Thus, the bZIP domain-containing protein in some Archaea species may have evolved into additional functions or may have lost the TF function. On the other hand, since only one member could be detected in some species from Archaea, the common ancestor among eukaryotes and prokaryotes may have only a single *bZIP* gene. Furthermore, the primitive eukaryote *Giardia lamblia* also encodes only one bZIP member (Deppmann et al. 2006), suggesting a single *bZIP* sequence in the common ancestor of eukaryotes. In green plants, the single celled chlorophyte *Chlamydomonas reinhardtii* (algae) genome encodes around seven *bZIP* genes (Corrêa et al. 2008). The fern *Selaginella moellendorffii* is among the few surviving members of the lycophytes, an ancient group of plants whose origins can be traced back as far as 400 million years ago. Its genome also encodes around eight bZIP members based on our genome-wide identification. However, the moss *Physcomitrella patens* genome encodes up to 40 members of bZIP TFs (Corrêa et al. 2008), suggesting a large expansion after the divergence of this species from green algae. Our data showed that the MRCA of sorghum and rice was estimated to encode at least 75 *bZIP* sequences and the MRCA of sorghum, rice and arabidopsis encoded only 28 TFs. The data suggested that the large scale of expansion occurred after the divergence of monocots from dicots. Since both *Chlamydomonas reinhardtii* and *Selaginella moellendorffii* encode less than 10 members, it is suggested that another large scale of expansion would have happened after/during the divergence of Spermatophyla.

Expansion and retention of bZIP TFs

In the present study, we have identified all *bZIP* genes in sorghum and they could be clustered into seven classes. Some of them have evolved into larger classes such as classes 1, 2, 3 and 6 with more than 10 members in each class (**Figure 1**). On the contrary, some of them have practiced limited expansion, such as class 7 with only three members retained. Similar results were also observed in the arabidopsis, rice and poplar genomes (Jakoby et al. 2002; Corrêa et al. 2008; Nijhawan et al. 2008). Thus, the result suggested that this gene family

might expand itself with a common mechanism. On the other hand, our data show that genomes from different species encode various numbers of bZIP TFs. The detailed analysis from sorghum, rice and arabidopsis suggested that segmental duplication could be regarded as the major mechanism to drive the expansion of this gene family. To demonstrate if it is a commonly used mechanism for the *bZIP* gene expansion in other plants, we randomly selected three more completely sequenced genomes for such analyses. We first genome-widely identified all genes encoding bZIP TFs in these three genomes. We then genome-widely surveyed if the majority of these genes were expanded from segmental duplication. We first analyzed the soybean (*Glycine max*) genome. We have identified at least 78 *bZIP* genes in the genome and 62 of them (79.5%) have been involved in segmental duplication. A similar result has been observed in the *Populus trichocarpa* and *Brachypodium distachyon* genomes. In these two genomes, up to 79% (83/105) and 61% (51/84) of *bZIP* genes have been involved in segmental duplication, respectively. Thus, it seems that plant genomes duplicated *bZIP* genes with a common mechanism.

After duplication, one copy may disappear or evolve into a pseudogene due to the lack of selective constraints (Jiang et al. 2008). The duplicated gene may also be retained by obtaining complementary or new biological functions (Jiang et al. 2008). To explore the retention mechanism of duplicated genes, we analyzed *Ka*/*Ks* ratios of duplicated pairs. To our surprise, the *Ka*/*Ks* value of only one duplicated pair is larger than 0.5 and all the remaining *Ka*/*Ks* ratios are less than 0.5. The fact suggested that high percentages of duplicated genes should have been lost due to the very limited functional divergence. However, many of them have been retained. We then analyzed their expression divergence between duplicated pairs. Our data indicated that high percentages of duplicated genes showed the difference in their expression patterns under various abiotic and biotic stresses in rice, arabidopsis and sorghum (**Figure 6**). The percentages should be underestimated since the expression divergence among different tissues is not yet calculated. Thus, the expression divergence should be regarded as the major mechanism for the retention of duplicated *bZIP* genes in sorghum, rice and arabidopsis. In fact, expression divergence followed by segmental duplication have been observed not only in plants (Ganko et al. 2007; Li et al. 2009), but also in animals (Blekhman et al. 2009).

Biological functions of bZIP TFs

Plant bZIP TFs play vital roles in multiple biological processes including different stages of plant development and the responses to various abiotic and biotic stresses/signals as mentioned in the Introduction. Based on the expression data in sorghum (**Figure 4**), *bZIP* genes may play a role in the whole developmental stages. Some of the *bZIP* genes were expressed in all tested tissues and some of them were absent in leaf, root, seed or stem tissues (**Figure 4**). However, all expressed *bZIP* genes showed the transcript signal in panicles, suggesting the role of this gene family in flower development. Totally, we have detected 16 expressed genes out of 22 tested genes including all six members of group C (class 5) *bZIPs*. No signal could be detected in all nine tested tissues for the remaining six *bZIP* genes from group S (class 6 and 7). These six genes could also not be activated either by drought, high salinity and cold stresses or by glucose and sucrose treatments. In addition, another set of primers was designed for the expression analysis and similarly, no signal could be detected. Several reasons may be provided to explain why we can not detect any expression signal. For example, these genes may have been silenced or their expression is too weak to detect in this study. Other reasons may be that these genes are cell type-specific or are responsive to other abiotic/biotic factors or are wrongly annotated genes.

On the other hand, some of group C and S *bZIP* genes have been experimentally proved to participating in sugar signaling (Jakoby et al. 2002; Hanson and Smeekens 2009; Smeekens et al. 2010). Among the six sorghum group C (class 5) *bZIPs*, one gene was downregulated and three genes were upregulated by both glucose and sucrose (**Figure 4C**). Among 16 tested group S (class 6 and 7) *bZIP* genes, we also detected four members being involved in glucose/sucrose-mediated signaling pathway (**Figure 4C**). These genes may be potentially useful for the further improvement of sorghum sugar content by molecular breeding. In arabidopsis, genome-wide microarray expression analyses under glucose/sucrose treatments have also been carried out. We have retrieved the expression data from the ArrayExpress under the accession number E-MEXP-475 (http://www.ebi.ac.uk/gxa/experiment/E-MEXP-475) for glucose treatment (Li et al. 2006) and the NCBI GEO datasets (http://www.ncbi.nlm.nih.gov/geo/) with accession numbers GDS1734 for sucrose treatment (Gonzali et al. 2006). Based on our analysis, three *bZIP* genes, *At1g68880*, *At1g77920* and *At4g34590*, were upregulated and six genes, *At1g43700*, *At4g36730*, *At5g15830*, *At5g24800*, *At5g28770* and *At5g49450*, were downregulated by glucose treatment. Under sucrose treatment, five genes including *At1g22070*, *At1g49720*, *At1g75390*, *At2g17770* and *At4g34590* were up-regulated and three genes including *At2g22850*, *At4g36730* and *At5g24800* were downregulated. In rice, suspension cells were treated by sucrose starvation, a total of nine differentially expressed *bZIP* genes was detected including *LOC_Os01g64730, LOC_Os02g07840, LOC_Os02g16680, LOC_Os02g52780, LOC_Os04g41820, LOC_Os05g03860, LOC_Os05g49420, LOC_Os06g10880, LOC_Os06g41770* (Wang et al. 2007). Only one gene *LOC_Os04g41820* was downregulated and the remaining eight genes were

upregulated. These data confirmed that *bZIP* genes should have been involved in sugar-mediated signaling pathway in plants. On the other hand, except for *bZIP* genes from group C and S, some of the family members from other groups were also differentially regulated by sugar treatments, suggesting that other groups of *bZIP* genes might also participate in this signaling pathway.

bZIP genes play roles not only in sugar signaling but also in abiotic stress-mediated signaling pathways. In this study, we have detected 10 out of 22 tested *bZIP* genes from group C (class 5) and S (class 6 and 7) with differential expression under drought, high salinity and/or cold stresses in sorghum. In arabidopsis and rice, a total of 29 (40%) and 43 (49%) *bZIP* genes were regulated by these abiotic stresses, respectively. Nijhawan et al. (2008) reported 37 drought/salinity/cold-regulated *bZIP* genes in rice. We have detected six more differentially expressed genes, which may be due to the different detection methods for gene expression. In general, expression analyses revealed high percentages of *bZIP* genes in a genome with differential expression regulation under stresses, suggesting an extensive involvement of this gene family in abiotic stress-related biological processes.

Materials and Methods

Plant materials and growth conditions

Grain sorghum (*Sorghum bicolor* (L.) Moench) cultivar BT×623 was used for all experiments. Seeds were germinated in water and were then planted in pots filled with the commercially available soil from SPA Flora Singapore. Plants were grown in a greenhouse under natural light and temperature conditions. Totally, nine different tissues were used for total RNA preparation. Young leaves and roots were taken from 10-d-old seedlings. Mature leaves and roots were from 3-month-old plants. Stems were from 40-d-old plants. Young and mature panicles were from un-opened and flowering panicles, respectively. Young and mature seeds were taken from milky and harvest stages of plants, respectively.

Abiotic stress and sugar treatments

Two-week-old sorghum seedlings were collected from pots and the soil was carefully washed away from roots. After washing, the whole seedlings were used for all abiotic and sugar treatments. For drought stress, seedlings were treated with 30% PEG solution and the whole plants were collected at different time intervals (0.0, 0.5 and 2.0 h) and then frozen in liquid nitrogen for total RNA isolation. For high salinity and cold stresses, the similar stage of seedlings was subjected to a 250 mM NaCl solution or ice-containing water, respectively. The whole seedlings at different time intervals (0, 2 and 8 h) were used for total RNA isolation. For glucose and sucrose treatments, 2-week-old seedlings were subjected to 5% glucose and sucrose solutions, respectively. Samples were then collected at 0, 2, and 8 h intervals for glucose treatment and 0, 2, and 6 h intervals for sucrose treatment.

RT-PCR and qRT-PCR analysis

Total RNA samples were isolated from various stages of tissues or from abiotic stress/sugar-treated samples using a QIAGEN (Hilden, Germany) RNeasy Mini Kit. The first-strand cDNAs were synthesized using an Invitrogen kit. The cDNA samples were used for detecting expression profiles of *bZIP* genes under normal and stressed conditions. PCRs were performed in 20 μL reaction mixtures with 20 ng cDNA, 200 μM dNTPs, 2.5 mM $MaCl_2$, 0.5 μM primers and 1 unit taq DNA polymerase in 1× PCR buffer provided by QIAGEN. The temperature profile for PCR is as follows: 94 °C for 2 min followed by 35 cycles at 94 °C for 10 s, 55–65 °C for 10 s (annealing temperature varies from different primer sets) and 72 °C for 1 min. The reaction was stopped by a 5 min extension step at 72 °C. PCR products were visualized by ethidium bromide staining in an agarose gel. The q-RT-PCR analysis was carried out according to the description by Jiang et al. (2007).

All primers used for RT-PCR and qRT-PCR were designed by Applied Biosystems Primer Express software. The designed primer sets were then subjected to the phytozome sorghum database (http://www.phytozome.net/sorghum) for a BLAST search to eliminate the non-specific primers. Supporting Table S4 lists all the primers used for both RT-PCR and qRT-PCR.

Processing of publicly available expression data from sorghum, rice and arabidopsis

Sorghum ESTs were collected from the KEGG Organisms/EST Datasets (http://www.genome.jp/kegg/catalog/org_list2.html) and the NCBI EST database. Rice ESTs were collected from the MSU/TIGR rice genome annotation database (http://rice.plantbiology.msu.edu/) and the NCBI EST database. Arabidopsis ESTs were obtained from *The Arabidopsis Information Resource* (TAIR, http://www.Arabidopsis.org/). We identified if a *bZIP* gene contained ESTs by BLAST searches with at least 95% identity over at least 90% of the length of ESTs.

The rice Massively Parallel Signature Sequencing (MPSS) database (http://mpss.udel.eud/rice/; Nobuta et al. 2007) was used to evaluate differentially regulated *bZIP* genes under abiotic (drought, high salinity and cold) and biotic (*Mg* and *Xoo*) stresses according to our previous description (Jiang et al. 2009). For arabidopsis, the expression data under

abiotic (drought, high salinity and cold) and biotic (*EOr* and *Bcin*) stresses were downloaded from the TAIR database (www.arabidopsis.org) and differentially expressed *bZIP* genes were identified according to the description (Matsui et al. 2008). Sucrose regulated rice and arabidopsis *bZIP* genes were collected from previous reports by Wang et al. (2007) and Gonzali et al. (2006), respectively. Glucose regulated *bZIP* genes in arabidopsis were obtained according to the description by Li et al. (2006).

Database searches for genes encoding bZIP TFs in higher plants

Representative bZIP TFs were obtained from the Pfam database (http://pfam.sanger.ac.uk/). In the database, a total of 22 seed bZIP members was selected from plants, animals and yeasts. The domain amino acid sequences were retrieved from these sequences using the Pfam and the SMART domain search programs with E-value cutoff of 0.01. The domain sequences were then aligned using ClustalX 2.0 (Thompson et al. 1997). The aligned amino acid sequences were used to generate HMM profiles for HMM searches with E-value cutoff of 1.0 against annotated protein databases from multiple organisms. The latest version of the sorghum annotation database (release Sbi1.4) was used to retrieve all annotated genes and their putative protein sequences. The release 6.1 of TIGR/MSU Rice Genome Annotation database (http://rice.plantbiology.msu.edu/; Yuan et al. 2005; Ouyang et al. 2007) and the release TAIR9 of arabidopsis genome annotation (http://www.Arabidopsis.org) were employed for gene and protein annotation in rice and arabidopsis, respectively. Besides the HMM searches, BLASTP searches were also carried out using the seed bZIP domain sequences as queries to figure out more members in multiple genomes. In addition, the newly obtained sequences from both HMM and BLASTP searches were also used as queries for BLASTP searches to try to pick up additional members in multiple genomes. These members were then submitted to domain searches to confirm the presence of the bZIP domain by the Pfam and SMART conserved domain programs with E-value = 0.01 as cutoff level. Proteins confirmed by domain searches were regarded as putative bZIP TFs (referred to bZIP TFs for convenience). Otherwise, they were excluded from our dataset.

Sequence alignment and phylogenetic analysis

The domain amino acid sequences were used for alignment using ClustalX 2.0 (Thompson et al. 1997). The aligned domain sequences were used for the construction of phylogenetic tree according to the description by Jiang and Ramachandran (2006).

Chromosome localization and detecting of duplicated genes

Chromosomal distributions of *bZIP* genes were performed by searching their physical positions of their corresponding locus numbers in the Phytozome database for sorghum, the TIGR/MSU database for rice and the TAIR database for arabidopsis.

Tandemly duplicated *bZIP* genes in sorghum, rice and arabidopsis were determined if the duplicated genes are less than or equal to 10 genes apart and they belong to the same family. In addition, tandem pairs should be within 350 kb for both sorghum and rice and within 100 Kb for arabidopsis as suggested by Lehti-Shiu et al. (2009).

Segmentally duplicated chromosome blocks have been previously identified in a genome-wide level in arabidopsis (Simillion et al. 2002) and rice (Vandepoele et al. 2003; Lin et al. 2006). Thus, segmentally duplicated *bZIP* genes were obtained by comparing positions of *bZIP* genes with known duplicated chromosomal blocks. For sorghum, soybean and poplar, we identified duplicated blocks using the flanking regions (50 kb upstream and downstream) of *bZIP* genes according to the method by Kong et al. 2007.

Detection of transposable element (TE)-related *bZIP* genes

We have investigated the contribution of both retrotransposons and DNA transposons to the expansion of *bZIP* genes. To detect possible retrogenes in the *bZIP* family, amino acid sequences from single exon *bZIP* genes were subjected to BLASTP searches against all of the remaining bZIP TFs with two or more exon-containing coding sequences in their corresponding genes. Retrogenes were identified according to the criteria (Wang et al. 2006). To determine the contribution of DNA transposons to the expansion of the *bZIP* family, the flanking genomic sequences of the 50 kb upstream and downstream of *bZIP* genes were used for the identification of four major transposon family members including mutator-like transposable element (*MULE*), *hAT*, *CACTA* and *Helitron* families according to the methods (Jiang et al. 2010).

Estimation of synonymous (*Ka*) and synonymous (*Ks*) substitutions per site and their ratios

Pairs from segmental or tandem duplication were used for the estimation of *Ka*, *Ks* and their ratios. Amino acid sequences from segmentally or tandemly duplicated pairs were aligned first and the aligned sequences were subsequently transferred into original cDNA sequences using the PAL2NAL program (http://www.bork.embl.de/pal2nal/; Suyama et al. 2006). The aligned cDNA sequences were then used for calculating *Ka*/*Ks*

ratios using the yn00 program of PAML package as described (Yang and Nielsen 2000).

Acknowledgements

We are grateful for the assistance from Dr Ritu Bhalla in editing and polishing this document. We would like to thank Dr Yunhua Chi and Ms Yansong Cheng for their technical assistance in the RT-PCR analysis. This work was supported by Strategic Research Program from Temasek Life Sciences Laboratory in Singapore.

Received 26 Aug. 2010 Accepted 23 Nov. 2010

References

Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* **309**, 1052–1056.

Aguan K, Sugawara K, Suzuki N, Kusano T (1993) Low-temperature-dependent expression of a rice gene encoding a protein with a leucine-zipper motif. *Mol. Gen. Genet.* **240**, 1–8.

Baena-Gonzalez E, Rolland F, Thevelein JM, Sheen J (2007) A central integrator of transcription networks in plant stress and energy signalling. *Nature* **448**, 938–943.

Bagavathi S, Malathi R (1996) Introns and protein revolution–an analysis of the exon/intron organisation of actin genes. *FEBS Lett.* **392**, 63–65.

Bedell JA, Budiman MA, Nunberg A, Citek RW, Robbins D, Jones J, Flick E, Rholfing T, Fries J, Bradford K, McMenamy J, Smith M, Holeman H, Roe BA, Wiley G, Korf IF, Rabinowicz PD, Lakey N, McCombie WR, Jeddeloh JA, Martienssen RA (2005) Sorghum genome sequencing by methylation filtration. *PLoS Biol.* **3**, e13.

Blekhman R, Oshlack A, Gilad Y (2009) Segmental duplications contribute to gene expression differences between humans and chimpanzees. *Genetics* **182**, 627–630.

Chuang CF, Running MP, Williams RW, Meyerowitz EM (1999) The PERIANTHIA gene encodes a bZIP protein involved in the determination of floral organ number in *Arabidopsis thaliana*. *Genes Dev.* **13**, 334–344.

Ciceri P, Locatelli F, Genga A, Viotti A, Schmidt RJ (1999) The activity of the maize Opaque2 transcriptional activator is regulated diurnally. *Plant Physiol.* **121**, 1321–1328.

Corrêa LG, Riaño-Pachón DM, Schrago CG, dos Santos RV, Mueller-Roeber B, Vincentz M (2008) The role of bZIP transcription factors in green plant evolution: adaptive features emerging from four founder genes. *PLoS One* **3**, e2944.

Deppmann CD, Alvania RS, Taparowsky EJ (2006) Cross-species annotation of basic leucine zipper factor interactions: Insight into the evolution of closed interaction networks. *Mol. Biol. Evol.* **23**, 1480–1492.

Despres C, DeLong C, Glaze S, Liu E, Fobert PR (2000) The *Arabidopsis* NPR1/NIM1 protein enhances the DNA binding activity of a subgroup of the TGA family of bZIP transcription factors. *Plant Cell* **12**, 279–290.

Finkelstein RR, Lynch TJ (2000) Abscisic acid inhibition of radicle emergence but not seedling growth is suppressed by sugars. *Plant Physiol.* **122**, 1179–1186.

Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A (2006) Pfam: clans, web tools and services. *Nucleic Acids Res.* **34**, D247–251.

Fukazawa J, Sakai T, Ishida S, Yamaguchi I, Kamiya Y, Takahashi Y (2000) Repression of shoot growth, a bZIP transcriptional activator, regulates cell elongation by controlling the level of gibberellins. *Plant Cell* **12**, 901–915.

Ganko EW, Meyers BC, Vision TJ (2007) Divergence in expression between duplicated genes in *Arabidopsis*. *Mol. Biol. Evol*. **24**, 2298–2309.

Gong W, Shen YP, Ma LG, Pan Y, Du YL, Wang DH, Yang JY, Hu LD, Liu XF, Dong CX, Ma L, Chen YH, Yang XY, Gao Y, Zhu D, Tan X, Mu JY, Zhang DB, Liu YL, Dinesh-Kumar SP, Li Y, Wang XP, Gu HY, Qu LJ, Bai SN, Lu YT, Li JY, Zhao JD, Zuo J, Huang H, Deng XW, Zhu YX (2004) Genome-wide ORFeome cloning and analysis of *Arabidopsis* transcription factor genes. *Plant Physiol.* **135**, 773–782.

Gonzali S, Loreti E, Solfanelli C, Novi G, Alpi A, Perata P (2006) Identification of sugar-modulated genes and evidence for *in vivo* sugar sensing in *Arabidopsis*. *J. Plant Res.* **119**, 115–123.

Guan Y, Ren H, Xie H, Ma Z, Chen F (2009) Identification and characterization of bZIP-type transcription factors involved in carrot (*Daucus carota* L.) somatic embryogenesis. *Plant J.* **60**, 207–217.

Guo A, He K, Liu D, Bai S, Gu X, Wei L, Luo J (2005) DATF: A database of Arabidopsis transcription factors. *Bioinformatics* **21**, 2568–2569.

Gupta S, Chattopadhyay MK, Chatterjee P, Ghosh B, SenGupta DN (1998) Expression of abscisic acid-responsive element-binding protein in salt-tolerant *indica* rice (*Oryza sativa* L. cv. Pokkali). *Plant Mol. Biol.* **37**, 629–637.

Hanson J, Hanssen M, Wiese A, Hendriks MM, Smeekens S (2008) The sucrose regulated transcription factor bZIP11 affects amino acid metabolism by regulating the expression of ASPARAGINE SYNTHETASE1 and PROLINE DEHYDROGENASE2. *Plant J.* **53**, 935–949.

Hanson J, Smeekens S (2009) Sugar perception and signaling–an update. *Curr. Opin. Plant Biol.* **12**, 562–567.

Hossain MA, Cho JI, Han M, Ahn CH, Jeon JS, An G, Park PB (2010a) The ABRE-binding bZIP transcription factor OsABF2 is a positive regulator of abiotic stress and ABA signaling in rice. *J. Plant Physiol.* **167**, 1512–1520.

Hossain MA, Lee Y, Cho JI, Ahn CH, Lee SK, Jeon JS, Kang H, Lee CH, An G, Park PB (2010b) The bZIP transcription factor OsABF1 is an ABA responsive element binding factor that enhances abiotic stress signaling in rice. *Plant Mol. Biol.* **72**, 557–566.

Hsieh TH, Li CW, Su RC, Cheng CP, Sanjaya, **Tsai YC, Chan MT** (2010) A tomato bZIP transcription factor, SlAREB, is involved in water deficit and salt stress response. *Planta* **231**, 1459–1473.

Iida K, Seki M, Sakurai T, Satou M, Akiyama K, Toyoda T, Konagaya A, Shinozaki K (2005) RARTF: database and tools for complete sets of *Arabidopsis* transcription factors. *DNA Res.* **12**, 247–256.

Iwata Y, Koizumi N (2005) An *Arabidopsis* transcription factor, AtbZIP60, regulates the endoplasmic reticulum stress response in a manner unique to plants. *Proc. Natl. Acad. Sci. USA* **102**, 5280–5285.

Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, Parcy F (2002) bZIP transcription factors in *Arabidopsis*. *Trends Plant Sci.* **7**, 106–111.

Jiang SY, Bachmann D, La H, Ma Z, Venkatesh PN, Ramamoorthy R, Ramachandran S (2007) *Ds* insertion mutagenesis as an efficient tool to produce diverse variations for rice breeding. *Plant Mol. Biol.* **65**, 385–402.

Jiang SY, Christoffels A, Ramamoorthy R, Ramachandran S (2009) Expansion mechanisms and functional annotations of hypothetical genes in the rice genome. *Plant Physiol.* **150**, 1997–2008.

Jiang SY, Ma Z, Ramachandran S (2010) Evolutionary history and stress regulation of the lectin superfamily in higher plants. *BMC Evol. Biol.* **10**, 79.

Jiang SY, Ramachandran S (2006) Comparative and evolutionary analysis of genes encoding small GTPases and their activating proteins in eukaryotic genomes. *Physiol. Genomics* **24**, 235–251.

Jiang SY, Rengasamy R, Ramachandran S (2008) Comparative transcriptional profiling and evolutionary analysis of the GRAM domain family in eukaryotes. *Dev. Biol*. **314**, 418–432.

Kaminaka H, Näke C, Epple P, Dittgen J, Schütze K, Chaban C, Holt BF 3rd, **Merkle T, Schäfer E, Harter K, Dangl JL** (2006) bZIP10- LSD1 antagonism modulates basal defense and cell death in *Arabidopsis* following infection. *EMBO J.* **25**, 4400–4411.

Kang SG, Price J, Lin PC, Hong JC, Jang JC (2010) The *Arabidopsis* bZIP1 transcription factor is involved in sugar signaling, protein networking, and DNA binding. *Mol. Plant* **3**, 361–373.

Kim JC, Lee SH, Cheong YH, Yoo CM, Lee SI, Chun HJ, Yun DJ, Hong JC, Lee SY, Lim CO, Cho MJ (2001) A novel cold-inducible zinc finger protein from soybean, SCOF-1, enhances cold tolerance in transgenic plants. *Plant J.* **25**, 247–259.

Kim S, Kang JY, Cho DI, Park JH, Kim SY (2004) ABF2, an ABRE-binding bZIP factor, is an essential component of glucose signaling and its overexpression affects multiple stress tolerance. *Plant J.* **40**, 75–87.

Kobayashi F, Maeta E, Terashima A, Takumi S (2008) Positive role of a wheat HvABI5 ortholog in abiotic stress response of seedlings. *Physiol. Plant.* **134**, 74–86.

Kong H, Landherr LL, Frohlich MW, Leebens-Mack J, Ma H, de Pamphilis CW (2007) Patterns of gene duplication in the plant SKP1 gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. *Plant J.* **50**, 873–885.

Kusano T, Berberich T, Harada M, Suzuki N, Sugawara K (1995) A maize DNA-binding factor with a bZIP motif is induced by low temperature. *Mol. Gen. Genet.* **248**, 507–517.

Lara P, Oñate-Sánchez L, Abraham Z, Ferrándiz C, Díaz I, Carbonero P, Vicente-Carbajosa J (2003) Synergistic activation of seed storage protein gene expression in *Arabidopsis* by ABI3 and two bZIPs related to OPAQUE2. *J. Biol. Chem.* **278**, 21003–21011.

Lee SC, Choi HW, Hwang IS, Choi du S, Hwang BK (2006) Functional roles of the pepper pathogen-induced bZIP transcription factor, CAbZIP1, in enhanced resistance to pathogen infection and environmental stresses. *Planta* **224**, 1209–1225.

Lehti-Shiu MD, Zou C, Hanada K, Shiu SH (2009) Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. *Plant Physiol.* **150**, 12–26.

Letunic I, Doerks T, Bork P (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res.* **37** (Database issue), D229–D232.

Li L, Stoeckert CJ Jr, **Roos DS** (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189.

Li Y, Lee KK, Walsh S, Smith C, Hadingham S, Sorefan K, Cawley G, Bevan MW (2006) Establishing glucose- and ABA-regulated transcription networks in *Arabidopsis* by microarray analysis and promoter classification using a relevance vector machine. *Genome Res.* **16**, 414–427.

Li Z, Zhang H, Ge S, Gu X, Gao G, Luo J (2009) Expression pattern divergence of duplicated genes in rice. *BMC Bioinform.* **10**(Suppl 6), S8.

Liao Y, Zhang JS, Chen SY, Zhang WK (2008a) Role of soybean GmbZIP132 under abscisic acid and salt stresses. *J. Integr. Plant Biol.* **50**, 221–230.

Liao Y, Zou HF, Wei W, Hao YJ, Tian AG, Huang J, Liu YF, Zhang JS, Chen SY (2008b) Soybean *GmbZIP44*, *GmbZIP62* and *GmbZIP78* genes function as negative regulator of ABA signaling and confer salt and freezing tolerance in transgenic *Arabidopsis*. *Planta* **228**, 225–240.

Lin H, Zhu W, Silva JC, Gu X, Buell CR (2006) Intron gain and loss in segmentally duplicated genes in rice. *Genome Biol.* **7**, R41.

Liu JX, Srivastava R, Che P, Howell SH (2007) Salt stress responses in *Arabidopsis* utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. *Plant J.* **51**, 897–909.

Liu JX, Srivastava R, Howell SH (2008) Stress-induced expression of an activated form of AtbZIP17 provides protection from salt stress in *Arabidopsis*. *Plant Cell Environ.* **31**, 1735–1743.

Lu G, Gao C, Zheng X, Han B (2009) Identification of OsbZIP72 as a positive regulator of ABA response and drought tolerance in rice. *Planta* **229**, 605–615.

Matsui A, Ishida J, Morosawa T, Mochizuki Y, Kaminuma E, Endo TA, Okamoto M, Nambara E, Nakajima M, Kawashima M, Satou M, Kim JM, Kobayashi N, Toyoda T, Shinozaki K, Seki M (2008) *Arabidopsis* transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. *Plant Cell Physiol.* **49**, 1135–1149.

Mukherjee K, Choudhury AR, Gupta B, Gupta S, Sengupta DN (2006) An ABRE-binding factor, OSBZ8, is highly expressed in salt tolerant cultivars than in salt sensitive cultivars of indica rice. *BMC Plant Biol.* **6**, 18.

Nieva C, Busk PK, Domínguez-Puigjaner E, Lumbreras V, Testillano PS, Risueño MC, Pagès M (2005) Isolation and functional characterisation of two new bZIP maize regulators of the ABA responsive gene *rab28*. *Plant Mol. Biol.* **58**, 899–914.

Niggeweg R, Thurow C, Kegler C, Gatz C (2000a) Tobacco transcription factor TGA2.2 is the main component of as-1-binding factor ASF-1 and is involved in salicylic acid- and auxin-inducible expression of as-1-containing target promoters. *J. Biol. Chem.* **275**, 19897–19905.

Niggeweg R, Thurow C, Weigel R, Pfitzner U, Gatz C (2000b) Tobacco TGA factors differ with respect to interaction with NPR1, activation potential and DNA-binding properties. *Plant Mol. Biol.* **42**, 775–788.

Nijhawan A, Jain M, Tyagi AK, Khurana JP (2008) Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. *Plant Physiol.***146**, 333–350.

Nobuta K, Venu RC, Lu C, Beló A, Vemaraju K, Kulkarni K, Wang W, Pillay M, Green PJ, Wang GL, Meyers BC (2007) An expression atlas of rice mRNAs and small RNAs. *Nat. Biotechnol.* **25**, 473–477.

Osterlund MT, Hardtke CS, Wei N, Deng XW (2000) Targeted destabilization of HY5 during light-regulated development of *Arabidopsis*. *Nature* **405**, 462–466.

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR (2007) The TIGR Rice Genome Annotation Resource: Improvements and new features. *Nucleic Acids Res.* **35**, D883–D887.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* **457**, 551–556.

Pontier D, Miao ZH, Lam E (2001) Trans-dominant suppression of plant TGA factors reveals their negative and positive roles in plant defense responses. *Plant J.* **27**, 529–538.

Rahmani F, Hummel M, Schuurmans J, Wiese-Klinkenberg A, Smeekens S, Hanson J (2009) Sucrose control of translation mediated by an upstream open reading frame-encoded peptide. *Plant Physiol.* **150**, 1356–1367.

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among Eukaryotes. *Science* **290**, 2105–2110.

Riechmann JL, Ratcliffe OJ (2000) A genomic perspective on plant transcription factors. *Curr. Opin. Plant Biol.* **3**, 423–434.

Rodriguez-Uribe L, O'Connell MA (2006) A root-specific bZIP transcription factor is responsive to water deficit stress in tepary bean (*Phaseolus acutifolius*) and common bean (*P. vulgaris*). *J. Exp. Bot.* **57**, 1391–1398.

Rogozin IB, Sverdlov AV, Babenko VN, Koonin EV (2005) Analysis of evolution of exon-intron structure of eukaryotic genes. *Brief. Bioinform.* **6**, 118–134.

Rook F, Gerrits N, Kortstee A, van Kampen M, Borrias M, Weisbeek P, Smeekens S (1998) Sucrose-specific signalling represses translation of the *Arabidopsis* ATB2 bZIP transcription factor gene. *Plant J.* **15**, 253–263.

Satoh R, Fujita Y, Nakashima K, Shinozaki K, Yamaguchi-Shinozaki K (2004) A novel subgroup of bZIP proteins functions as transcriptional activators in hypoosmolarity-responsive expression of the *ProDH* gene in *Arabidopsis*. *Plant Cell Physiol.* **45**, 309–317.

Shen H, Cao K, Wang X (2007) A conserved proline residue in the leucine zipper region of AtbZIP34 and AtbZIP61 in *Arabidopsis thaliana* interferes with the formation of homodimer. *Biochem. Biophys. Res. Commun.* **362**, 425–430.

Shimizu H, Sato K, Berberich T, Miyazaki A, Ozaki R, Imai R, Kusano T (2005) LIP19, a basic region leucine zipper protein, is a Fos-like molecular switch in the cold signaling of rice plants. *Plant Cell Physiol.* **46**, 1623–1634.

Shiu SH, Karlowski WM, Pan R, Tzeng YH, Mayer KF, Li WH (2004) Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell* **16**, 1220–1234.

Silveira AB, Gauer L, Tomaz JP, Cardoso PR, Carmello-Guerreiro S, Vincentz M (2007) The *Arabidopsis* AtbZIP9 protein fused to the VP16 transcriptional activation domain alters leaf and vascular development. *Plant Sci.* **172**, 1148–1156.

Simillion C, Vandepoele K, Van Montagu MC, Zabeau M, van de Peer Y (2002) The hidden duplication past of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **99**, 13627–13632.

Smeekens S, Ma J, Hanson J, Rolland F (2010) Sugar signals and molecular networks controlling plant growth. *Curr. Opin. Plant Biol.* **13**, 274–279.

Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882.

Thurow C, Schiermeyer A, Krawczyk S, Butterbrodt T, Nickolov K, Gatz C (2005) Tobacco bZIP transcription factor TGA2.2 and related factor TGA2.1 have distinct roles in plant defense responses and plant development. *Plant J.* **44**, 100–113.

Ulm R, Baumann A, Oravecz A, Máté Z, Adám E, Oakeley EJ, Schäfer E, Nagy F (2004) Genome-wide analysis of gene expression reveals function of the bZIP transcription factor HY5 in the UV-B response of *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **101**, 1397–1402.

Uno Y, Furihata T, Abe H, Yoshida R, Shinozaki K, Yamaguchi-Shinozaki K (2000) Arabidopsis basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. *Proc. Natl. Acad. Sci. USA* **97**, 11632–11637.

Usadel B, Bläsing OE, Gibon Y, Retzlaff K, Höhne M, Günther M, Stitt M (2008) Global transcript levels respond to small changes of the carbon status during progressive exhaustion of carbohydrates in *Arabidopsis* rosettes. *Plant Physiol.* **146**, 1834–1861.

Vandepoele K, Simillion C, Van de Peer Y (2003) Evidence that rice and other cereals are ancient aneuploids. *Plant Cell* **15**, 2192–2202.

Vettore AL, Yunes JA, Cord Neto G, da Silva MJ, Arruda P, Leite A (1998) The molecular and functional characterization of an *Opaque2* homologue gene from Coix and a new classification of plant bZIP proteins. *Plant Mol. Biol.* **36**, 249–263.

Walsh J, Waters CA, Freeling M (1998) The maize gene liguleless2 encodes a basic leucine zipper protein involved in the establishment of the leaf bladesheath boundary. *Genes Dev.* **12**, 208–218.

Wang HJ, Wan AR, Hsu CM, Lee KW, Yu SM, Jauh GY (2007) Transcriptomic adaptations in rice suspension cells under sucrose starvation. *Plant Mol. Biol.* **63**, 441–463.

Wang W, Zheng H, Fan C, Li J, Shi J, Cai Z, Zhang G, Liu D, Zhang J, Vang S, Lu Z, Wong GK, Long M, Wang J (2006) High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* **18**, 1791–1802.

Warren AJ (2002) Eukaryotic transcription factors. *Curr. Opin. Struct. Biol.* **12**, 107–114.

Wellmer F, Kircher S, Rügner A, Frohnmeyer H, Schäfer E, Harter K (1999) Phosphorylation of the parsley bZIP transcription factor CPRF2 is regulated by light. *J. Biol. Chem.* **274**, 29476–29482.

Weltmeier F, Ehlert A, Mayer CS, Dietrich K, Wang X, Schütze K, Alonso R, Harter K, Vicente-Carbajosa J, Dröge-Laser W (2006) Combinatorial control of *Arabidopsis* proline dehydrogenase transcription by specific heterodimerisation of bZIP transcription factors. *EMBO J.* **25**, 3133–3143.

Weltmeier F, Rahmani F, Ehlert A, Dietrich K, Schütze K, Wang X, Chaban C, Hanson J, Teige M, Harter K, Vicente-Carbajosa J, Smeekens S, Dröge-Laser W (2009) Expression patterns within the *Arabidopsis* C/S1 bZIP transcription factor network: availability of heterodimerization partners controls gene expression during stress response and development. *Plant Mol. Biol.* **69**, 107–119.

Wiese A, Elzinga N, Wobbes B, Smeekens S (2004) A conserved upstream open reading frame mediates sucrose-induced repression of translation. *Plant Cell* **16**, 1717–1729.

Wingender E, Chen X, Fricke E, Geffers R, Hehl R, Liebich I, Krull M, Matys V, Michael H, Ohnhäuser R, Prüss M, Schacherer F, Thiele S, Urbach S (2001) The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.* **29**, 281–283.

Xiang Y, Tang N, Du H, Ye H, Xiong L (2008) Characterization of OsbZIP23 as a key player of the basic leucine zipper transcription factor family for conferring abscisic acid sensitivity and salinity and drought tolerance in rice. *Plant Physiol.* **148**, 1938–1952.

Xiong Y, Liu T, Tian C, Sun S, Li J, Chen M (2005) Transcription factors in rice: a genome-wide comparative analysis between monocots and eudicots. *Plant Mol. Biol.* **59**, 191–203.

Xue GP, Loveridge CW (2004) HvDRF1 is involved in abscisic acid-mediated gene regulation in barley and produces two forms of AP2 transcriptional activators, interacting preferably with a CT-rich element. *Plant J.* **37**, 326–339.

Yáñez M, Cáceres S, Orellana S, Bastías A, Verdugo I, Ruiz-Lara S, Casaretto JA (2009) An abiotic stress-responsive bZIP transcription factor from wild and cultivated tomatoes regulates stress-related genes. *Plant Cell Rep.* **28**, 1497–1507.

Yang O, Popova OV, Süthoff U, Lüking I, Dietz KJ, Golldack D (2009) The *Arabidopsis* basic leucine zipper transcription factor AtbZIP24 regulates complex transcriptional networks involved in abiotic stress resistance. *Gene* **436**, 45–55.

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under evolutionary models. *Mol. Biol. Evol.* **17**, 32–43.

Yilmaz A, Nishiyama MY Jr, **Fuentes BG, Souza GM, Janies D, Gray J, Grotewold E** (2009) GRASSIUS: a platform for comparative regulatory genomics across the grasses. *Plant Physiol.* **149**, 171–180.

Yin Y, Zhu Q, Dai S, Lamb C, Beachy RN (1997) RF2a, a bZIP transcriptional activator of the phloem-specific rice tungro bacilliform virus promoter, functions in vascular development. *EMBO J.* **16**, 5247–5259.

Yoshida T, Fujita Y, Sayama H, Kidokoro S, Maruyama K, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K (2010) AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. *Plant J.* **61**, 672–685.

Yuan Q, Ouyang S, Wang A, Zhu W, Maiti R, Lin H, Hamilton J, Haas B, Sultana R, Cheung F, Wortman J, Buell CR (2005) The institute for genomic research Osa1 rice genome annotation database. *Plant Physiol.* **138**, 18–26.

Yun KY, Park MR, Mohanty B, Herath V, Xu F, Mauleon R, Wijaya E, Bajic VB, Bruskiewich R, de Los Reyes BG (2010) Transcriptional regulatory network triggered by oxidative signals configures the early response mechanisms of japonica rice to chilling stress. *BMC Plant Biol.* **10**, 16.

Zhang B, Foley RC, Singh KB (1993) Isolation and characterization of two related *Arabidopsis* ocs-element bZIP binding proteins. *Plant J.* **4**, 711–716.

Zou M, Guan Y, Ren H, Zhang F, Chen F (2008) A bZIP transcription factor, OsABI5, is involved in rice fertility and stress tolerance. *Plant Mol. Biol.* **66**, 675–683.

(Co-Editor: Martin A. J. Parry)

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Exon-intron organization of seven classes of sorghum *bZIP* genes. The *bZIP* family was classified according to **Figure 1**. The value in parentheses indicate the number of corresponding classes of *bZIP* genes.

Figure S2. Corresponding intron positions among aligned 92 bZIP domain regions. The bZIP domain amino acid sequences were used for alignment using ClustalX 2.0 (Thompson et al. 1997). Red vertical lines indicate the insertion positions of corresponding introns.

Figure S3. Intron positions within the bZIP domain regions. (A) Three corresponding intron insertion regions of aligned bZIP amino acid sequences as shown by red triangles. (B) Total of eight patterns of intron distribution and their intron positions in the 92 *bZIP* sorghum genes.

Figure S4. Phylogenetic analysis of the bZIP gene families from sorghum, rice and arabidopsis. The tree was generated using the bootstrap method with a heuristic search of the PAUP 4.0b8 program using the bZIP domain amino acids sequences. The generated tree was similar to that from Bayesian analysis. Ancestral units were defined according to Shiu et al. (2004). Red circles represent the ancestral units among all the three organisms, and blue circles indicate the ancestral units between sorghum and rice.

Table S1. Genome-wide identification of the bZIP gene family in the sorghum genome

Table S2. Genome-wide identification of the bZIP gene family in the japonica rice genome

Table S3. Genome-wide identification of the bZIP gene family in the *Arabidopsis* genome

Table S4. Primer sequences used for reverse transcription-polymerase chain reaction (RT-PCR) and quantitative RT-PCR) (qRT-PCR)