Running head: bZIP Transcription Factors in Rice

Corresponding author:
Dr. Jitendra P. Khurana
Professor and Coordinator (UGC-SAP)
Department of Plant Molecular Biology
University of Delhi South Campus
Benito Juarez Road
New Delhi-110021, India
Telephone: 91-011-24115126
Fax: 91-01-24115270; 24119430
E-mail: khuranaj@genomeindia.org

Research area: Signal Transduction and Hormone Action
A Genomic Survey and Gene Expression Analysis of Basic Leucine Zipper (bZIP) Transcription Factor Family in Rice

Aashima Nijhawan, Mukesh Jain, Akhilesh K. Tyagi, and Jitendra P. Khurana*

Interdisciplinary Centre for Plant Genomics and Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi-110021, India

Corresponding author: Dr. Jitendra P. Khurana, Professor and Coordinator (UGC-SAP), Department of Plant Molecular Biology, University of Delhi South Campus, Benito Juarez Road, New Delhi-110021, India
Telephone: 91-011-24115126
Fax: 91-01-24115270; 24119430
E-mail: khuranaj@genomeindia.org
Footnote: This work was financially supported by the Department of Biotechnology, Government of India, the University Grants Commission, and the Council of Scientific and Industrial Research, New Delhi (research fellowship to A.N.).
The basic leucine zipper (bZIP) proteins comprise a family of transcriptional regulators present exclusively in eukaryotes. The bZIP proteins characteristically harbour a bZIP domain composed of two structural features: a DNA binding basic region and the leucine zipper dimerization region. They have been shown to regulate diverse plant-specific phenomena, including seed maturation and germination, floral induction and development, and photomorphogenesis, and are also involved in stress and hormone signaling. We have identified 89 bZIP transcription factor encoding genes in rice genome. Their chromosomal distribution and sequence analyses suggest that bZIP transcription factor family has evolved via gene duplication. The phylogenetic relationship among rice bZIP domains as well as with bZIP domains from other plant bZIP factors suggests that homologous bZIP domains exist in plants. Similar intron/exon structural patterns were observed in the basic-hinge regions of their bZIP domains. A detailed sequence analysis has been done to identify additional conserved motifs outside the bZIP domain and to predict their DNA-binding site specificity as well as dimerization properties which has helped classify them into different groups and sub-families, respectively. The expression of bZIP transcription factor encoding genes has been analyzed by full-length cDNA and EST-based expression profiling. This expression profiling was complemented by microarray analysis. The results indicate specific or co-expression patterns of rice bZIP transcription factors starting from floral transition to various stages of panicle and seed development. The bZIP transcription factor encoding genes in rice also displayed differential expression patterns in rice seedlings in response to abiotic stress and light irradiation. An effort has been made to link the structure and expression pattern of bZIP transcription factor encoding genes in rice to their function, based on the information obtained from our analyses and earlier known results. This information will be important for functional characterization of bZIP transcription factors in rice.
Several families of transcription factors have been identified both in animals as well as plants. The bZIP transcription factor family is among the largest and diverse dimerizing transcription factor families. The bZIP transcription factors owe their name to their highly conserved bZIP domain composed of a basic region and a leucine zipper (Hurst, 1994). The bZIP domain is 60-80 amino acids in length. The basic region consists of about 16 amino acid residues characterized by the presence of an invariant N-x7-R/K motif, while the leucine zipper is composed of heptad repeats of leucines or other bulky hydrophobic amino acids (isoleucine, valine, phenylalanine, or methionine) positioned exactly nine amino acids towards the C-terminus. The basic and leucine zipper regions are functionally distinct. The basic region is responsible for sequence-specific DNA-binding and is highly conserved. The leucine zipper is an amphipathic sequence in the form of coiled-coil which confers dimerization specificity and is less conserved. This leucine zipper sequence mediates homo- and/or hetero-dimerization in bZIP proteins. The bZIP monomers are in the form of long α-helices and, at the time of DNA binding, the N-terminal half binds in the major groove to double-stranded DNA while the C-terminal half mediates dimerization to form a superimposed coiled-coil structure called the leucine zipper (Landschulz et al., 1988; Ellenberger et al., 1992). The bZIP transcription factor encoding genes have been identified exclusively in eukaryotes. Their number is 17 in Saccharomyces cerevisiae, 31 in Caenorhabditis elegans, 27 in Drosophila, 56 in humans and 67 in Arabidopsis genomes (Riechmann et al., 2000; Fassler et al., 2002; Vinson et al., 2002; Deppmann et al., 2004).

In plants, like other transcription factor genes, members of bZIP transcription factor family are also either expressed constitutively or in organ-specific (Schindler et al., 1992a; Rodriguez-Uribe and O’Connell, 2006), stimulus-responsive (de Vetten and Ferl, 1995), development-dependent (Chern et al., 1996) and cell-cycle-specific (Minami et al., 1993) manner. So far, over 120 bZIP transcription factors have been identified in different plants (Supplemental Table S1). In many plants, bZIP transcription factors have been reported to function in the regulation of seed maturation and germination (Izawa et al., 1994; Chern et al., 1996). The available functional information suggests their role in not only the integration of spatial and temporal information during floral induction but also in flower development (Chuang et al., 1999; Walsh and Freeling, 1999; Strathmann
et al., 2001; Abe et al., 2005; Thurow et al., 2005; Wigge et al., 2005; Muszynski et al., 2006). Genetic and molecular studies have revealed their regulatory role in photomorphogenesis and light signaling (Jakoby et al., 2002; Holm et al., 2002; Mallappa et al., 2006). They also participate in defence against pathogens (Zhang et al., 2003; Thurow et al., 2005). Many bZIP transcription factors function in abscisic acid (ABA) and/or stress signaling (de Vetten and Ferl, 1995; Lopez-Molina et al., 2002; Fujita et al., 2005), and a few of them have been shown to be hormone responsive (Fukazawa et al., 2000; Niggeweg et al., 2000).

Fourteen bZIP transcription factors have been identified or functionally characterized from rice. Amongst these, OSBZ8 and TRAB1 are possibly involved in the regulation of transcription by ABA-mediated signaling (Nakagawa et al., 1996; Hobo et al., 1999a). OsZIP-1a has a putative role in the induction of Em gene promoter by ABA (Nantel and Quatrano, 1996), and LIP19 and OsOBF1 are involved in cold signaling (Aguan et al., 1993; Shimizu et al., 2005). RISBZ1 is required for endosperm-specific expression of storage protein genes and is thought to be involved in regulation of some metabolic enzymes expressed in developing seeds (Onodera et al., 2001; Yamamoto et al., 2006). RITA-1 is assumed to be involved in the regulation of rice genes expressed during seed development (Izawa et al., 1994), and REB is a major DNA-binding protein for alpha-globulin gene promoter in rice endosperm, later shown to be a transcriptional activator in rice grains (Nakase et al., 1997; Yang et al., 2001). OsZIP-2a and OsZIP-2b are expected to have putative roles in selective inactivation of some G-box binding factors (Nantel and Quatrano, 1996). RF2a and RF2b function in vascular development as well as play a role in the symptom development of rice tungro disease (Yin et al., 1997b; Dai et al., 2003; Dai et al., 2004). Two other rice bZIP proteins, RISBZ4 and RISBZ5, have unknown functions (Onodera et al., 2001).

In this study, we report the identification and characterization of 89 genes encoding bZIP transcription factors in the rice genome. These bZIP transcription factors in rice have been classified on the basis of their putative DNA-binding site specificity and dimerization properties. Their domain organization and gene structure have been analyzed. The duplication events likely to contribute to the expansion of the bZIP family in rice were also identified. Phylogenetic analyses have been done to reveal the
relationship pattern among the bZIP transcription factor genes in rice as well as with other known plant bZIP proteins. An elaborate analysis of the spatial and temporal expression pattern of bZIP transcription factors, during various developmental stages, including vegetative growth, floral transition, and panicle and seed development has been performed. Moreover, it has been demonstrated that the expression of bZIP transcription factor encoding genes is regulated by various environmental cues such as light and abiotic stress.

RESULTS AND DISCUSSION

Identification and Nomenclature of the bZIP Transcription Factors in Rice

In order to identify bZIP transcription factor genes in rice, BLAST searches of the rice (Oryza sativa subsp. japonica cv Nipponbare) genome were performed using Hidden Markov Model (HMM) profile of bZIP domain as query. Further analysis based on the presence of bZIP domain with confidence (E value less than 1.0) as given by SMART revealed 89 potential non-redundant bZIP transcription factor genes which were designated as OsbZIP genes. The OsbZIP genes were given a number designation from 1 to 89 to provide a unique identifier for each bZIP transcription factor gene as proposed for bZIP transcription factors in Arabidopsis (Jakoby et al., 2002). The nomenclature was based on their position on rice chromosomes 1 to 12 and from top to bottom. The OsbZIP gene number, The Institute of Genomic Research (TIGR) locus, open reading frame length, protein length, and chromosomal location of all the 89 OsbZIP genes are listed in Supplemental Table S2. Domain analysis revealed that all but five of the potential bZIP transcription factors, had a typical bZIP domain having an invariant N-x7-R/K motif in the basic region and a heptad repeat of leucines or other bulky hydrophobic amino acids positioned exactly nine amino acids upstream of R/K towards the C-terminus. Out of the remaining five, in case of OsbZIP34, the heptad repeat of leucines was positioned 30 amino acids instead of the usual nine amino acids towards the C-terminus, while in case of OsbZIP80 the conserved arginine/lysine in the basic region was replaced by an isoleucine residue. This is similar to the earlier reported rice bZIP transcription factor OsZIP-1a (Nantel and Quatrano, 1996). On the other hand, among the remaining three unusual OsbZIP transcription factors both OsbZIP21 and 82 have a lysine, and OsbZIP85
has a phenylalanine in place of the conserved asparagine in the basic region. Such an unusual basic domain has also been reported in case of yeast Met4 bZIP transcription factor (Thomas et al., 1992).

**Structure of OsbZIP Genes**

In order to obtain some insight into the gene structure of the 89 OsbZIP genes, their exon-intron organization was analyzed. It was found that 17 (19.1%) of the total OsbZIP genes were intronless (Supplemental Table S2). This is in agreement with the total percentage (19.9%) of rice genes predicted to be intronless (Jain et al., 2007a). Amongst those having introns, the number of introns in the ORF varied from one to twelve. Further, the pattern of intron positions within the basic and hinge regions of the bZIP domain (since these regions are the most conserved in the bZIP domain) as well as intron phases with respect to codons were analyzed. Out of the 72 intron-containing OsbZIP genes, 62 had introns in the bZIP domain region. On the basis of intron presence, position and splicing phase, the OsbZIP genes were divided into seven patterns, a to g (Fig. 1; Supplemental Fig. S1; Supplemental Table S2). Patterns a and b were the most prevalent. Both had one intron each but the position and phase were different. Pattern a had an intron in phase 2 (P2) in the basic region while pattern b had an intron in phase 0 (P0) in the hinge region. Pattern c, having 2 introns each in phase 0, was seen in case of most of the members (13/16) of group VII only. Patterns d-f were uncommon and each was found in one OsbZIP gene only. Pattern g was devoid of any intron in the basic or hinge region. Pattern g was found in case of 27 OsbZIP genes out of which only 17 were intronless while the remaining 10 had introns in region other than their basic and hinge regions. It should also be noted that whenever an intron was present in the hinge region, both its phase and position were always conserved. However, introns present in the basic region were at variable positions either having splicing phase 0 (P0) or 2 (P2). Further, it was observed that there were more phase 0 introns and symmetric exons (with the same phase on both ends). Such a condition might aid in exon shuffling and recombinational fusion events in case of the OsbZIP gene family members (Patthy, 1987). The overall pattern of intron position acts as an index to the phylogenetic relationships in a gene
family. The splicing phase has remained well conserved during the course of evolution of OsbZIP genes.

Chromosomal Distribution of OsbZIP Genes

In order to determine the chromosomal location of all the 89 OsbZIP genes, the position (in bp) of each OsbZIP gene was determined on rice chromosome pseudomolecules available at TIGR (release 5). Figure 2 is a diagrammatical representation of the chromosomal distribution of OsbZIP genes on the 12 rice chromosomes (the exact position in bp of each OsbZIP gene on rice chromosome pseudomolecules is given in Supplemental Table S2). Although the OsbZIP genes are distributed on each of the 12 rice chromosomes, their distribution is not uniform. Their chromosomal distribution pattern reveals that certain chromosomes and chromosomal regions have a relatively high density of OsbZIP genes. For instance, 12 OsbZIP genes are located each on chromosomes 1 and 2, while there is a single OsbZIP gene present on chromosome 10.

In case of plants, during the course of its evolution, a gene family has generally undergone either tandem duplication or large scale segmental duplication to maintain high number of the family members (Cannon et al., 2004). To comprehend the mechanisms underlying the OsbZIP gene family evolution, both tandem and segmental duplication events were examined. Surprisingly, the duplication in case of OsbZIP genes was confined to only chromosomal block duplications in rice as none of the OsbZIP genes were found to be arranged in tandem. Forty nine OsbZIP genes were found to be located on the duplicated segmental regions of rice chromosomes mapped by TIGR (Fig. 2, Supplemental Table S3) at a maximum length distance permitted between collinear gene pairs of 500 kb. All the OsbZIP genes located on the duplicated segments of chromosomes belong to the same group, sub-divided on the basis of the DNA-binding properties as described later. Most of the members of groups VI (10 out of 14), VII (9 out of 16) and IX (8 out of 11) were found to be present on the duplicated segments of rice chromosomes. These results suggest that segmental genome duplication events have contributed largely to the expansion of the bZIP transcription factor gene family in rice.
Identification of Additional Structural Features in OsbZIP Transcription Factors

Few bZIP transcription factors, like lotus LjBzf1 and soybean STF1 (Cheong et al., 1998; Nishimura et al., 2002), contain domains other than the characteristic bZIP domain. In case of AtbZIP transcription factor family, some of the members have been shown to have additional conserved motifs in addition to the bZIP domain and have been classified into ten groups on the basis of the sequence similarities of their basic region and the presence of additional conserved motifs (Jakoby et al., 2002). The search for the presence of additional domains in the OsbZIP transcription factors using SMART database did not predict any other domain of known function. The members of OsbZIP transcription factor family were also searched for the presence of any conserved motifs by using Multiple Em (Expectation Maximization) for Motif Elicitation (MEME) analysis tool. A total of 25 additional conserved motifs, outside the bZIP domain, could be identified in 63 OsbZIP transcription factors (Fig. 3). The multilevel consensus sequence and the amino acid length of these conserved motifs are given in Supplemental Table S4. However, the biological significance of only a few of these motifs is known.

Five OsbZIP proteins have motif 6, characterized by the presence of small stretches of proline and aromatic amino acids Y, F and W, and is a part of proline-rich domain (PRD). Similar motifs have also been identified in the N-terminal region of group G bZIP transcription factors in Arabidopsis, which have been shown to have transcriptional activation potential (Schindler et al., 1992b; Jakoby et al., 2002). Motif 7 identified in five OsbZIP transcription factors, corresponds to GCB (GBF-conserved box; consensus, NLNIGMDxW), which is a part of the trans-activation domain, conserved among plant HBP-1a/GBF-type bZIP factors (Meier and Gruissem, 1994; Okanami et al., 1996). Interestingly, all the seven OsbZIP factors possessing motif 6 and/or motif 7 are also predicted to be G-box binding factors as described later.

A part of motifs 10, 12 and 13 represents potential casein kinase II (CKII) phosphorylation sites (S/TxxD/E, where x represents any amino acid), indicated by the presence of conserved TVDE and TLED/E in motifs 12 and 13, respectively. Motif 14 also contains a phosphorylation site for Ca\textsuperscript{2+}-dependent protein kinase (R/KxxS/T) which is found two residues C-terminal to a conserved leucine within the motif. Such motifs have been identified in members of group A bZIP transcription factors in Arabidopsis,
which include AREBs/ABFs (Bensmihen et al., 2002; Jakoby et al., 2002). Eight, nine, eleven, and ten OsbZIP proteins containing motifs 10, 12, 13 and 14, respectively, have such consensus phosphorylation sites for various protein kinases. Amongst these, OsbZIP66 corresponds to already identified rice ABA-inducible bZIP transcription factor TRAB1 (Hobo et al., 1999a).

The motif 24, identified in three OsbZIP proteins, corresponds to the COP1-interaction motif through which the Arabidopsis bZIP proteins like HY5 and HYH interact with the WD40 domain of COP1 (Holm et al., 2001; Holm et al., 2002). Certain other motifs of unknown function are common to AtbZIP and OsbZIP transcription factors (Jakoby et al., 2002). For example, OsbZIP07 and 44 share motif 17 with group F; OsbZIP39 and 60 have a part of motif 21 in common with group B; 14 OsbZIP proteins share a part of motif 20 with group D AtbZIP proteins. Although the function of these motifs is not yet known, the presence of these conserved motifs certainly reflects functional similarities amongst the OsbZIP proteins sharing these common motifs with the AtbZIP proteins.

**Predicted DNA-binding Site Specificity of OsbZIP Transcription Factors**

The bZIP transcription factors owe their DNA-binding ability to the basic region of their bZIP domain. Their DNA-binding specificity, in turn, is determined by the presence of certain key amino acid residues present in the basic and hinge regions of the bZIP domain, that contact DNA bases at cis-acting elements (Suckow et al., 1993a; Niu et al., 1999). Majority of the plant bZIP proteins isolated till date recognize elements with an ACGT core (Foster et al., 1994). On the basis of their binding preferences, the plant bZIP proteins have been classified into three groups (Izawa et al., 1993). In order to predict the DNA-binding site specificity of the OsbZIP proteins, sequence alignment of the amino acids present in the basic and hinge regions of the 89 OsbZIP proteins was done. Based on this alignment and the type of amino acid residues present at four different sub-regions as defined by Niu et al. (1999) (nomenclature based on Suckow et al., (1993b): A (-31 to -27), B (-25 and -24) and C (-21 to -11) within the basic region and D (-6 and -1) in the hinge region, the binding pattern of these OsbZIP proteins could be
classified into 11 groups (I to XI) (Supplemental Fig. S2). The characteristic features of the OsbZIP proteins classified into different groups are described in Table 1.

It may be reasoned that promoters of different genes may have variations in their target sites and that sometimes binding affinity could be low in certain cases in order to have a desirably low expression level of a particular gene, hence, these predictions are made to facilitate further studies on DNA-binding patterns of OsbZIP transcription factors. Such a prediction of DNA-binding ability, followed by experimental verification, would definitely be of great utility in identifying which genes are selectively activated by different OsbZIP transcription factors.

**Predicted Dimerization Properties of the OsbZIP Transcription Factors**

The bZIP transcription factors are known to bind to DNA predominantly as homo- or hetero-dimers mediated by the leucine zipper region of the bZIP domain (Landschulz et al., 1988). The amino acid sequence of the leucine zipper determines the homo- and/or hetero-dimerization of bZIP proteins (O'Shea et al., 1992). The leucine zipper region responsible for dimerization is composed of heptad repeats of amino acids, identified as a, b, c, d, e, f and g depending upon their position in the heptad. The leucine zipper dimerization stability and specificity is determined by the amino acids present at a, d, e and g positions as they are near the leucine zipper interface. The dimerization specificity of *Arabidopsis*, human and *Drosophila* bZIP transcription factors have been predicted on the basis of presence of attractive or repulsive interhelical $g\leftrightarrow e'$ electrostatic interactions and presence of polar or charged amino acids in the a and d positions of the hydrophobic interface of leucine zipper region (Fassler et al., 2002; Vinson et al., 2002; Deppmann et al., 2004).

In order to predict the dimerization specificity of the 89 OsbZIP proteins, the amino acid sequence of their bZIP domains was manually arranged in the form of heptads starting from four amino acids (corresponding to the g position in the first heptad) before the occurrence of first leucine (N-x$_7$-R/K-x$_9$-L) in the bZIP domain (Supplemental Fig. S3). The N-terminal and C-terminal boundaries of the OsbZIP leucine zippers have been demarcated following the criteria used for *Arabidopsis* bZIP proteins (Deppmann et al., 2004). It was observed that leucine zipper length was variable ranging from 2-9 heptads
in case of OsbZIP proteins. A detailed analysis of the type of amino acids present at \(a, d, e\) and \(g\) positions in OsbZIP proteins was carried out and a cross comparison was made with that of *Arabidopsis* and human (HsbZIPs). A comparison of charged amino acids present in \(g\) and \(e\) positions of different heptads showed that OsbZIPs, like AtbZIPs, have a lesser frequency of charged amino acids in these positions as compared to HsbZIPs (Fig. 4A). However, the frequency of charged residues was more in OsbZIPs than AtbZIPs. This indicates that interactions between \(g\leftrightarrow e'\) pairs may be more prominent in regulating dimerization specificity in OsbZIPs as compared to AtbZIPs. About 23% of amino acids present at \(a\) position were asparagines (Fig. 4A), which is almost equal to the frequency observed in AtbZIPs (Deppmann et al., 2004). This indicates that in case of OsbZIPs there will be greater number of homo-dimerizing leucine zippers as N-N interhelical interaction is preferred over interaction with other amino acids at \(a\) position similar to AtbZIPs (Acharya et al., 2002). Interestingly, the frequency of leucine, responsible for dimer stability, in the \(d\) position in OsbZIPs was found to be 71%, which is significantly greater than in AtbZIPs (56%) (Fig. 4A). However, the abundance of other aliphatic amino acids at \(d\) position was comparable in OsbZIPs (17%) and AtbZIPs (19%). The presence of leucine might be responsible for dimer stability of longer zippers as indicated by its higher frequency in the 7-9 heptads of the leucine zipper. The frequency of asparagine residues in the \(a\) positions is the highest in heptad 2 followed by heptad 5 (Fig. 4B), as observed earlier for AtbZIP proteins (Deppmann et al., 2004). This further supports the assumption that a larger number of OsbZIPs may prefer to homo-dimerize. Polar amino acids like serine and threonine and charged amino acids were also found in the \(a\) positions of some heptads in few OsbZIP proteins (Supplemental Fig. S3).

To analyze the contribution of charged residues in governing dimerization properties of OsbZIP proteins, the frequency of attractive and repulsive \(g\leftrightarrow e'\) pairs in each heptad of OsbZIP leucine zippers was calculated and the corresponding histogram is represented in the figure 5. This analysis was based on the frequency of attractive basic-acidic pairs (\(R\leftrightarrow E\) and \(K\leftrightarrow E\)), attractive acidic-basic pairs (\(E\leftrightarrow R, E\leftrightarrow K, D\leftrightarrow R,\) and \(D\leftrightarrow K\)), repulsive acidic pairs (\(E\leftrightarrow E, E\leftrightarrow D, E\leftrightarrow Q,\) and \(Q\leftrightarrow E\)) and repulsive basic pairs (\(K\leftrightarrow K, R\leftrightarrow K, Q\leftrightarrow K, R\leftrightarrow Q,\) and \(K\leftrightarrow Q\)) in the heptads. It was observed that the frequency of interactive \(g\leftrightarrow e'\) pairs was the maximum in the first heptads, with a sharp
decrease in the next three heptads. The frequency of attractive $g\leftrightarrow e'$ pairs increases in the fifth heptad. Moreover, only attractive $g\leftrightarrow e'$ pairs were found in the ninth heptad. However, repulsive $g\leftrightarrow e'$ pairs are predominant in sixth and seventh heptads, thereby suggesting the chances of hetero-dimerization. It should be noted that few OsbZIP proteins had multiple repulsive $g\leftrightarrow e'$ pairs, which were found to be completely absent in Arabidopsis leucine zippers.

Based on the above analyses, the OsbZIP proteins were classified into 29 sub-families (BZ1-BZ29) on the basis of similar predicted dimerization specificities. The various properties of these sub-families have been described in detail in the Supplemental Table S5). Known interaction patterns between different bZIP transcription factors in rice obey the predictions made in this study (Supplemental Table S5). It is evident from this analysis that OsbZIP proteins are expected to display a complex and varied dimerization patterns, with the potential to homo-dimerize with themselves or with the same sub-family members as well as hetero-dimerize with the members of other sub-families. Since many of the OsbZIP transcription factors show the potential to hetero-dimerize, it should be pointed out that most of the OsbZIP proteins which do not show differential expression profiles in our microarray analysis (described later) and are rather ubiquitously expressed, might assume specific functions when they hetero-dimerize in planta.

**Phylogenetic Analysis of OsbZIP Genes**

To analyze the evolutionary history of the OsbZIP genes, an unrooted phylogenetic tree was generated using the sequence alignments of the OsbZIP proteins (Fig. 6). Excluding one, i.e. OsbZIP80, all the other OsbZIP transcription factors could be subdivided into 10 clades, designated A to J, with well supported bootstrap values. It was observed that majority of the members, predicted to have similar DNA-binding properties, clustered together. However, certain members of groups III, IV and VI were exceptions as they clustered apart into different clades. To find the relatedness at amino acid level, three more phylograms based on (a) bZIP domain (b) basic and hinge regions and (c) leucine zipper region of OsbZIP proteins were generated (Supplemental Fig. S4). These three phylograms are similar yet have some interesting differences (marked in
Supplemental Fig. S4), highlighting the fact that more than one OsbZIP proteins might recognize same DNA sequence for binding but may have different dimerization properties, thereby capable of controlling a wide range of transcriptional responses. Most of the members belonging to one clade also shared one or more conserved motifs outside the bZIP domain. Further, 62 OsbZIP proteins formed 31 sister pairs, out of which 17 pairs were located on the duplicated segmental regions of rice chromosomes mapped by TIGR. Since the remaining 14 sister pairs have a high degree of similarity in their protein sequences, some of these might be present on the unidentified duplicated chromosomal segments.

To examine the phylogenetic relationships of rice and other known bZIP transcription factors, another unrooted phylogenetic tree was constructed from the sequence alignments of their proteins (Supplemental Fig. S5). All the bZIP transcription factors could be divided into distinct clades. Interspecies clustering was observed, indicating that homologous bZIP transcription factors exist in rice and other plants. For example, it was found that OsbZIP proteins predicted to be G-box (Group I) and C-box (Group VII) binding in our analysis also clustered together with other known G-box (e.g. EmBP-1, ROM2 and SGBF-1) and C-box (e.g. TGA1a, MBF2 and HBP-1b) binding bZIP proteins from other plants, respectively (Supplemental Fig. S5). Most of the clades had both OsbZIP and AtbZIP proteins. This indicates that bZIP transcription factor genes have appeared before divergence between monocots and dicots. It is also worth mentioning that most of the OsbZIP proteins which clustered together with AtbZIP proteins in the same clade shared one or more additional conserved motifs outside the bZIP domain. Therefore, majority of the OsbZIP proteins are expected to be orthologs of the AtbZIP proteins. Furthermore, majority of the clades contain members from different plants suggesting that the structure and function of most of the bZIP genes has probably remained conserved during angiosperm evolution.

**FL-cDNA, EST and Microarray Based Expression Profiling of OsbZIP Genes in Various Tissues/Organs**

In order to extract expression information for the rice bZIP transcription factors, a survey for the availability of any full-length cDNA (FL-cDNA), and/or expressed
sequence tag (EST) in databases was done. This analysis was based on the information given on the gene expression evidence search page available at TIGR. It was found that one or more FL-cDNA, and/or EST(s) were available for 75 of 89 OsbZIP transcription factor genes (Supplemental Table S6). This depicts that a very large percentage (84.27%) of these transcription factor genes are expressed. The total number of mapped EST sequences for OsbZIP transcription factors were highly variable ranging from a minimum of one for few OsbZIP genes, e.g. OsbZIP07, to a maximum of 277 for OsbZIP38 (Supplemental Table S6). This expression profiling was based on EST sequences obtained from various rice tissue/organ libraries, which in turn suggests that OsbZIP genes show organ/tissue specific differential expression. Most of the fourteen OsbZIP proteins for which we did not find any FL-cDNA and/or EST evidence in TIGR, showed significant similarity with bZIP proteins from other plant species, as revealed by BLASTP and TBLASTN search in various databases (Supplemental Table S6).

Another approach was used for expression profiling, which centred around the analysis of organ-specific expression of OsbZIP genes. This analysis was based on the use of microarray data from an earlier study (Ma et al., 2005), which described the expression pattern in several representative rice organs/tissues, including cultured cell, seedling, shoot, root, heading-stage panicle, and filling-stage panicle at whole genome level. Expression information for a total of 54 OsbZIP genes could be extracted. Further analysis demonstrated that 50 OsbZIP genes were expressed in at least one of the above mentioned rice organs/tissues (Supplemental Table S7). Only 22 of these were expressed in all six organs/cell types and a very few of them exhibited tissue/organ specific expression (Supplemental Fig. S6). The overall expression data analysis signifies that bZIP transcription factors may play regulatory roles at multiple developmental stages in rice.

Microarray Analysis of Expression Patterns of OsbZIP Genes during Floral Transition, Panicle, and Seed Development

Microarray analysis was performed using Affymetrix rice whole genome arrays to look at the gene expression profiles of OsbZIP transcription factor genes. Different rice tissues/organs and developmental stages (panicle and seed) were selected for microarray
analysis, which include seedling, seedling root, mature leaf, Y leaf (leaf subtending the shoot apical meristem, SAM), SAM, various stages of panicle (P1-P6) and seed (S1-S5) development, as described earlier (Jain et al., 2007b). However, in addition to the six (P1-P6) different stages of panicle development, three more stages of early panicle development (P1-I to P1-III) were included in this analysis which have been characterized as follows: 0.5-2 mm (P1-I), branch meristem, spikelet meristem, formation of secondary and tertiary branches and differentiation of rudimentary glumes; 2-5 mm (P1-II), floral meristem and floral organ initiation; 5-10 mm (P1-III), early floral organ differentiation. These stage specifications are also approximations based on information from Itoh et al. (2005).

The whole-chip data were processed as described in materials and methods. In order to analyze the expression profiles of the OsbZIP genes, the log signal values of 85 OsbZIP genes represented on the array were extracted. Average log signal values for all 85 OsbZIP genes from three biological replicates of each sample are given in Supplemental Table S8, and the corresponding hierarchial cluster display of these genes is depicted in Supplemental Figure S7. On the basis of the signal values, it was evident that majority of the OsbZIP genes are expressed in at least one of the rice vegetative organs and/or stages of development analyzed. This also includes at least nine of the 14 OsbZIP genes for which no EST or FL-cDNA could be detected from the data available at TIGR as described earlier. Of the remaining five OsbZIP genes, the Affy IDs were not available for three genes and the other two had very low expression (Supplemental Fig. S7). With the aim to reveal OsbZIP gene expression profiles, both during panicle and seed development stage(s), differential expression analysis was performed. A gene was defined as differentially expressed at a given stage only if the expression level of the gene at that stage was significantly higher (more than 2-fold) with p value less than 0.05 in comparison to the levels at all the other stages. To facilitate the analysis, two separate data subsets were created having genes differentially expressed during panicle and seed development in comparison to seedling, root, mature leaf, and Y leaf as controls. We found that a total of 28 and 25 OsbZIP genes were showing differential expression patterns in at least one of the stages of panicle and seed development, respectively, as compared to vegetative organs. Out of these, there was an overlap of 13 OsbZIP genes
between panicle and seed development. After excluding these 13 OsbZIP genes, the remaining were analyzed for preferential differential expression in any stage of panicle development as compared to seed development stages and vice-versa. This resulted in the identification of 11 and 10 OsbZIP genes, which displayed a preferential differential expression profile in at least one of the stages of early/late panicle and seed development, respectively (Fig. 7). Further confirmation of some of the representative genes in panicle and seed development stages was done by real time PCR analysis (Supplemental Figs. S8A and S8B). This analysis highlights the potential role of different OsbZIP transcription factor genes during different stages of early and/or late panicle and seed development. These data have been discussed in some detail below, in light of the earlier observations on the role of bZIP proteins in regulating reproductive development in plants.

Molecular studies in both monocots and dicots have reported the involvement of bZIP transcription factors in transition from the vegetative to the reproductive phase, viz. FD in Arabidopsis (Abe et al., 2005; Wigge et al., 2005); LIGULELESS2 (LG2) and DELAYED FLOWERING1 (DLF1) in maize (Walsh and Freeling, 1999; Muszynski et al., 2006). Our analysis also highlights few OsbZIP genes with specific expression in SAM and early panicle (P1-I, P1-II, P1-III, and P1) development stages (Figs. 7A and 7B). Such expression profiles suggest the possible involvement of OsbZIP transcription factors in regulating the expression pattern of genes involved in floral transition, determination of floral meristem, floral organ identity and early panicle development in rice. A few bZIP transcription factors have also been linked with other aspects of flower development, viz. PERIANTHIA (PAN) gene of Arabidopsis linked with floral organ number as well as their relative positions (Chuang et al., 1999); BZI proteins of tobacco involved in regulation of flower organ size (Strathmann et al., 2001); TGA2.1 implicated in defining organ identity in tobacco flowers (Thurow et al., 2005). A few OsbZIP genes were also found to be differentially expressed in one or more stages of panicle development (P2-P6) as well (Figs. 7A and 7B). These OsbZIP transcription factors, in turn, might be involved in the regulation of certain genes whose expression governs similar processes in rice.
In plants, quite a large number of the bZIP proteins characterized till date have been shown to play a role in the regulation of seed-specific genes, thereby linking them with different seed developmental processes (Izawa et al., 1994; Chern et al., 1996; Onodera et al., 2001). Interestingly, in cereals, bZIP proteins function as regulatory factor for genes encoding storage proteins in the endosperm (Schmidt et al., 1992; Vicente-Carbajosa et al., 1998; Onate et al., 1999). Our analysis also resulted in the identification of ten OsbZIP genes which depicted differential expression in at least one or more of the developmental stages of seed in rice (Figs. 7C and 7D), as described earlier. Out of these, nine OsbZIP genes displayed more than 2-fold increase in the transcript level during different stages of seed development (Fig. 7C), while a single OsbZIP gene, i.e. OsbZIP13, showed a decrease in the transcript level in late stages (S4 and S5) of seed development (Fig. 7D). The seed-specific high abundance expression profile of these nine OsbZIP genes could be correlated to their specific role in regulating the expression of genes involved in different stages of seed development, from embryo and endosperm development to seed maturation. OsbZIP58, up-regulated in stages from S2-S5 in our study, corresponds to an earlier characterized seed-specific rice bZIP transcription factor, RISBZ1 (Onodera et al., 2001). On similar lines, genes like OsbZIP34 and 76 can be proposed to have an important role in seed development. Two other rice bZIP transcription factors, one of them being RITA-1, has been shown to be highly expressed during seed development (Izawa et al., 1994) and the other, REB, isolated from maturing seed, has been shown to bind to the promoter of rice seed storage protein gene alpha-globulin (Nakase et al., 1997). Surprisingly, the OsbZIP transcription factor genes corresponding to RITA-1 and REB in our study, i.e. OsbZIP20 and 33, respectively, do not show seed-specific expression. Their expression profiles (Supplemental Fig. S7) reveal that both these are highly expressed throughout rice plant development. This probably depicts that these might either play multifunctional roles in other developmental processes in rice apart from being involved in seed development or may gain seed-specificity as a result of hetero-dimerization and hence function to regulate seed-specific genes. This assumption is further supported by the fact that REB has high degree of sequence similarity with maize bZIP factor OHP1, whose expression in not confined to seed only and is known to hetero-dimerize with another maize bZIP factor, OPAQUE2.
(O2), to regulate storage protein synthesis in the endosperm (Schmidt et al., 1992; Nakase et al., 1997). Recently, it has been demonstrated that there exists a synergistic interaction between a rice Dof transcriptional activator, RPBF and a bZIP activator, RISBZ1, which in turn regulates the expression of endosperm-specific genes in a quantitative manner (Yamamoto et al., 2006). Such interactions between other bZIP transcription factors and transcriptional regulators from other groups might exist in nature as well.

Furthermore, the relationship between expression pattern of OsbZIP genes in one or more stages of panicle and/or seed development and sequence features as well as phylogeny was analyzed (Supplemental Table S9). It was found that a number of OsbZIP genes having same conserved motif(s) and/or lying close in phylogenetic tree had similar expression pattern. Few duplicated OsbZIP genes and those having same intron patterns were also found to be co-expressed in one or more stages of panicle and/or seed development. Some OsbZIP genes predicted to have similar binding affinities were also found to exhibit similar expression profiles in various panicle and/or seed developmental stages.

**Light Regulation**

The bZIP transcription factors have also been shown to regulate promoters of light responsive genes. Amongst these are the *Arabidopsis* GBFs and parsley CPRF proteins, which have been shown to be light-regulated either at the expression level or in their intracellular localization (Jakoby et al., 2002). To test whether such a light-responsive relationship exists in case of OsbZIP genes and to identify the OsbZIP transcription factor genes regulated by light, analysis of the microarray data generated in an earlier study by Jiao et al. (2005) was carried out. This genomic study of rice gene expression in response to light revealed that different light qualities regulate a large number of rice genes. The expression values of only 47 OsbZIP genes could be retrieved under different light conditions in three different rice tissues/organs, i.e. seedling, shoot and root. Applying a 1.5-fold threshold with a \( p \) value less than 0.05, examination of expression ratios between light- and dark-grown rice tissue/organs resulted in the identification of 14 differentially expressed OsbZIP genes, with 9 up- and 5 down-regulated OsbZIP transcription factor genes in white light in comparison to dark (Fig. 8A,
Supplemental Table S10). In contrast to white light, less number (8) of OsbZIP transcription factor genes were regulated under specific monochromatic lights in rice seedlings with the maximum number of differentially expressed genes in case of blue light, followed by red and far-red light (Fig. 8B, Supplemental Table S10). Amongst the different activated and/or repressed OsbZIP genes, certain genes are differentially regulated similarly under different light treatments (Supplemental Figs. S9A and S9B). Two bZIP transcriptional activators, HY5 and HYH, have been shown to mediate transcription of light-inducible genes in Arabidopsis. HY5 has been well documented as a positive regulator of photomorphogenesis (Jakoby et al., 2002). Likewise, genetic analysis has linked HYH to blue-light regulated expression of genes (Holm et al., 2002). The OsbZIP proteins, OsbZIP01, 18 and 48, might represent closely related orthologs of Arabidopsis HY5 and/or HYH, as evident from the phylogenetic analysis as well as presence of additional conserved COP1 interaction motif specific to HY5 and HYH and might perform similar functions in rice. Incidently, the expression data for OsbZIP01 and 48 could not be retrieved from the data as these genes were not represented on the microarray chip. However, OsbZIP18 was found to be down-regulated under red light. The expression data were also correlated with pattern of introns, gene duplications, motif analysis, predicted DNA-binding specificity and phylogeny. A link could be seen in case of OsbZIP23 and 40 as well as OsbZIP28 and 41 (Supplemental Table S9).

Since a reasonable number of OsbZIP genes were displaying light-dependent differential expression, it can be assumed that other OsbZIP transcription factors, which might be involved in light signal transduction, may respond differently to light as depicted by some other bZIP transcription factors. The above analysis is, however, a first step towards understanding the light responsive nature of OsbZIP genes in terms of their expression level. Extensive studies on the light dependent localization as well as binding patterns of OsbZIP transcription factors would be important to understand their role in regulating the expression of light-responsive genes.

Expression of OsbZIP Genes under Abiotic Stress Conditions

The role of phytohormone abscisic acid (ABA) in the adaptation of plants to various abiotic environmental stress conditions like drought, high salinity and cold stress...
has been well characterized. The action of ABA is dependent on the modification of 
ABA-regulated genes which possess ABA-responsive element (ABRE) or certain ABRE-
like sequences in their promoter region (Busk and Pages, 1998). Interestingly, all the 
transcription factors reported till date, that bind to ABRE or ABRE-like sequences, come 
under the bZIP class of transcription factors (Busk and Pages, 1998). It is worth 
mentioning that a recent study reports that three bZIP transcription factor genes, 
\textit{AREB1/ABF2}, \textit{AREB2/ABF4} and \textit{ABF3/DPBF5}, depict ABA-, drought- and high-
salinity-inducible expression in vegetative tissues in \textit{Arabidopsis} (Fujita et al., 2005).

To explore the possible involvement of OsbZIP transcription factors in the 
regulation of abiotic stress-related genes, the expression pattern of \textit{OsbZIP} genes was 
analyzed under abiotic stress conditions. The analysis was based on microarray analysis 
performed with total RNA isolated from rice seedlings subjected to dehydration, salinity 
and cold treatment. Examination of microarray data indicated that at least 37 \textit{OsbZIP} 
genes displayed differential expression. We found that 26 genes were up-regulated and 
11 genes were down-regulated by at least two-fold, with a \textit{p} value below 0.05, under one 
or more of the above-mentioned stress treatments (Figs. 9A and 9B). Two \textit{OsbZIP} genes 
were up-regulated under all the three stress conditions. One of these, \textit{OsbZIP45}, 
represents the closely related ortholog of \textit{Zea mays} GBF1, which has been shown to be 
induced by hypoxia (de Vetten and Ferl, 1995). Interestingly, out of the 18 \textit{OsbZIP} genes 
differentially regulated under any two stress conditions, 16 \textit{OsbZIP} genes were 
differentially expressed under dehydration and salt stresses but not under cold stress (Fig. 
9, Supplemental Table S11). As far as alteration in expression to independent stress 
treatments was concerned, 13 \textit{OsbZIP} genes showed differential expression (10 up- and 3 
down-regulated) under dehydration stress and four \textit{OsbZIP} genes showed differential 
expression (2 up- and 2 down-regulated) under cold, while no \textit{OsbZIP} gene depicted 
differential expression under salt stress alone (Figs. 9C and 9D). It should be noted that 
\textit{OsbZIP38} (up-regulated under cold treatment), corresponds to the earlier identified rice 
low-temperature induced \textit{LIP19} gene (Shimizu et al., 2005). This confirms the reliability 
of our expression data. Additionally, the differential expression of some representative 
genes under abiotic stress condition(s) (as deduced from microarray data) was confirmed 
by real-time PCR results (Supplemental Fig. S8C).
The transcript levels of ABA-inducible bZIP transcription factor genes, \textit{OSBZ8} (\textit{OsbZIP05}) and \textit{TRAB1} (\textit{OsbZIP66}) (Nakagawa et al., 1996; Hobo et al., 1999a), were up-regulated under dehydration and salt stress conditions in our study, indicating the involvement of these genes in ABA-dependent stress signal transduction pathway. This observation is consistent with the earlier reports in the literature on drought stress-inducible genes being also activated by ABA (Ingram and Bartels, 1996). Further, a characteristic link was observed between sequence features, phylogeny and expression pattern of the \textit{OsbZIP} genes whose expression was altered under stress conditions (Supplemental Table S9). It was found that a majority of \textit{OsbZIP} genes possessing additional conserved motif(s) representing potential CKII and/or Ca$^{2+}$-dependent protein kinase phosphorylation sites (motifs 10, 11, 12, 13, 14), confined to group VI of predicted DNA-binding specificity and proposed to be involved in stress and/or ABA signaling are actually up-regulated under stress conditions. Thus, it would be interesting to analyse these OsbZIP transcription factors in context of stress signaling.

Recently, it has been demonstrated that an extensive gene set is shared among pollination/fertilization and/or seed development and stress responses like dehydration and wounding in rice (Cooper et al., 2003). To find out whether OsbZIP transcription factors are also a part of such interrelated genetic programs, a cross comparison between the differentially expressed \textit{OsbZIP} genes under different stress treatments and different stages of panicle and/or seed development was made. Such a cross comparison resulted in the identification of 11 \textit{OsbZIP} genes up-regulated under stress conditions (\textit{OsbZIP05}, 10, 25, 30, 34, 45, 46, 57, 68, 78 and 80) which had overlapping expression pattern at one or more stages of panicle and/or seed development. Two other genes, \textit{OsbZIP23} (up-regulated under all the three stress conditions) and \textit{OsbZIP65} (down-regulated under dehydration and cold stresses) were found to be down-regulated in one or more stages of both panicle and seed development. Thus, it is evident that \textit{OsbZIP} genes might form a subset within the genes shared among stress and developmental responses and these overlapping OsbZIP transcription factors might be common to more than one signal transduction pathways.

To find out whether \textit{OsbZIP} genes co-expressed during abiotic stress and various stages of panicle and/or seed development can be potential hetero-dimerizing partners,
their expression pattern was integrated with predicted dimerization properties. It was observed that genes depicting co-expression profiles, whether belonging to the same or different sub-family are the good candidates for hetero-dimerization in planta (Supplemental Table S12). However, the possibility of hetero-dimerization of OsbZIP genes, which do not show co-expression cannot be ruled out completely, because the features in the leucine zipper region governing hetero-dimerization are equally important. Only experimental evidence will provide credibility to these presumptions.

CONCLUSION

The bZIP transcription factors have been characterized in different plant species and linked to various developmental and physiological processes. An extensive study of the rice 89 bZIP genes has been performed in terms of structure, phylogeny, sequence, and expression analyses. These bZIP transcription factors could be divided into 11 groups on the basis of their predicted DNA-binding specificity, many of them supported by presence of additional conserved motifs outside the bZIP domain. We also defined 29 different sub-families based on their complex predicted dimerization patterns. Their predicted DNA-binding specificities and dimerization patterns would facilitate both in vitro and in vivo studies to reveal their binding and dimerization patterns. Structural differences between different OsbZIP members have been studied and an attempt made to link them to their functional roles in rice in a meaningful manner. Phylogenetic analysis of bZIP transcription factor encoding genes in rice and other plant species also provides useful information on their conserved functional roles. Expression data further supports the role of bZIP proteins in performing diverse developmental and physiological functions during floral transition, panicle and seed development, as well as abiotic stresses and light signaling in rice. Although the function of only a few rice bZIP genes is known till date, our analyses provide a platform for a more detailed functional analysis of bZIP genes in rice.

MATERIALS AND METHODS

Database Search for Rice bZIP Transcription Factors
For identifying the possible bZIP transcription factor genes in rice, BLAST search of all the annotated proteins in rice (*Oryza sativa*) genome at TIGR (release 5) was carried out using the HMM profile (build 2.3.2) of bZIP domain as query. The HMM profile of 65 amino acid-long bZIP domain was generated by alignments of 73 known *Arabidopsis* bZIP proteins and 71 rice protein sequences which were showing the presence of a bZIP domain. This search resulted in the identification of 130 proteins (including proteins corresponding to different gene models present at the same locus also) in rice, with a positive score cut off of 0.0. Out of these 130 proteins, those corresponding to different gene models present at the same locus were removed and the remaining were checked by SMART for the presence of the bZIP domain. Ninety proteins were predicted to have a bZIP domain with confidence ($E$ value less than 1.0). One protein, i.e. LOC Os12g06010, annotated as retrotransposon was also removed from the analysis.

**Sequence Analysis**

The additional conserved motifs outside the bZIP domain were identified by MEME analysis tool ([http://meme.sdsc.edu/meme/meme.html](http://meme.sdsc.edu/meme/meme.html)) version 3.5.4, which is widely used for analyzing a group of protein sequences to identify motifs in them. At first, all 89 OsbZIP protein sequences were given together as input and a limit of 50 motifs was specified with all other parameters set to default. These motifs were analyzed manually based on e-value cut off less than e-001 and those motifs were considered significant which were shared by majority of OsbZIP proteins classified into same group, according to their DNA-binding site specificity. ClustalX (version 1.83) program was used to perform the multiple sequence alignments. The unrooted phylogenetic trees were generated by neighbor-joining method and diplayed using NJPLOT program. To obtain information on the intron/exon structure, the cDNA sequences of OsbZIP proteins were aligned with their corresponding genomic sequences using Spidey ([http://www.ncbi.nlm.nih.gov/spidey/](http://www.ncbi.nlm.nih.gov/spidey/)). The information on intron distribution pattern and intron splicing phase within the basic and hinge regions of the bZIP domains were also derived from the aligned cDNA sequences.

**Localization of OsbZIP Genes on Rice Chromosomes and Segmental Duplication**
Each of the OsbZIP genes was positioned on the rice chromosome pseudomolecules available at TIGR (http://www.tigr.org/tdb/e2k1/osa1/pseudomolecules/info.shtml) by the BLASTN search. The segmental genome duplication of rice available at TIGR (http://www.tigr.org/tdb/e2k1/osa1/semental_dup/index.shtml) was used to determine the presence of OsbZIP genes on duplicated chromosomal segments, with the maximum length distance permitted between collinear gene pairs of 500 kb.

**Plant Material and Growth Conditions**

Rice (subsp. indica var IR64) seeds obtained from Indian Agricultural Research Institute were used for this study. Plant material for different developmental stages and their growth conditions for microarray analysis are same as described previously (Jain et al., 2007b). Three additional early panicle development stages, i.e. 0.5-2.0 mm, P1-I; 2.1-5.0 mm, P1-II; 5.1-10.0 mm, P1-III; were taken. For collection of different stages of early panicle development, leaves and stem were carefully removed and growth stage of each sample was measured in mm as well as carefully examined under the microscope before sampling. For stress series, 7-d-old light-grown rice seedlings were dried for 3 h (dehydration stress), treated with 200 mM NaCl for 3 h (salt stress), kept at 4°C ±1°C for 3 h (cold stress) as described previously (Jain et al., 2007b).

**Microarray Hybridization and Data Analysis**

Isolation of total RNA and quality controls were done as described previously (Jain et al., 2006). The microarray analysis was performed using one-cycle target labeling and control reagents (Affymetrix, Santa Clara) using five micrograms of a high quality total RNA as starting material for each sample as described earlier (Jain et al., 2007b). Affymetrix GeneChip® Rice Genome Arrays [Gene Expression Omnibus (GEO) platform accession no. GPL2025] were used for microarray analysis.

For microarray data analysis, the image (cel) files were imported into ArrayAssist (version 5.0) software. Three biological replicates of each sample with an overall correlation coefficient value of more than 0.94 were selected for final analysis. The data, from 66 hybridizations was included in final analysis. Two separate projects were
created, one for developmental series which included data from 57 chips (three biological replicates for each of seedling, root, mature leaf, Y leaf, SAM, P1-I, P1-II, P1-III, P1, P2, P3, P4, P5, P6, S1, S2, S3, S4, and S5) and the other for stress series which included data from 12 chips (three biological replicates for each of control seedlings and seedlings subjected to salt, dehydration and cold stress). The data for three hybridizations corresponding to control seedlings were common to both the experiments (developmental and stress).

The normalization and probe summarization for all the rice genes present on the chip in all the tissue samples analyzed was performed by Gene Chip Robust Multiarray Analysis (GCRMA) algorithm followed by log transformation of data. Average log signal intensity values of three biological replicates for each sample were then computed and these were used for further analysis. To identify differentially expressed genes, a Student's t-test was performed with microarray data obtained for mRNA derived from seedling, mature leaf, Y leaf, and root as reference tissues/organs. After this, the log signal intensity values for rice probe set IDs corresponding to OsbZIP genes were extracted into a separate dataset. The IDs of probe sets present on the Affymetrix rice genome array representing the OsbZIP genes were extracted using the genome browser tool available at TIGR (http://www.tigr.org/tigr-scripts/osa1-web/gbrowse/rice/). The data for only one probe set for each OsbZIP gene was used for expression analysis. The probes for 85 (out of 89) OsbZIP genes could be identified that were represented on the Affymetrix rice genome array (probe set IDs corresponding to each OsbZIP gene has been listed in Supplemental Table S8). Two genes viz. OsbZIP55 and 56 showed 100% sequence identity. Hence both have the same probe set on the GeneChip. All the subsequent analyses were performed on this subset of data only. The genes that were up- or down-regulated equal to or more than two-fold with a p value less than or equal to 0.05 (after applying Benjamini and Hochberg correction) in any stage of panicle and/or seed development as compared to reference tissues/organs were considered to be differentially expressed significantly. A gene was defined as specifically enriched in a given organ only if the expression level of the gene in the organ was significantly higher (more than 2-fold) than the levels in all the other organs. Similar approach was followed in case of stress series also.
Real-Time PCR Analysis

Verification of microarray data was done by checking the differential expression profiles of few of the representative *OsbZIP* genes in various rice tissues/developmental stages and stress treatments by real-time PCR analysis using gene-specific primers as described earlier (Jain et al., 2006). The real-time PCR analysis was based on at least two biological replicates of each sample and three technical replicates of each biological replicate. Supplemental Table S13 includes the primer sequences used for real-time PCR analysis.

FL-cDNA and EST based Expression Profiling

For FL-cDNA and EST based expression profiling, the gene expression evidence search page (http://www.tigr.org/tdb/e2k1/osa1/locus_expression_evidence.shtml) available at TIGR rice genome annotation was used. The TIGR locus IDs corresponding to the *OsbZIP* genes were searched to find the availability of corresponding FL-cDNA and EST. The FL-cDNA and EST sequences showed a minimal alignment over 90% length of the transcript with 95% identity.

Supplemental Data

The following materials are available in the online version of this article.

**Supplemental Figure S1.** Position and pattern of introns within the basic and hinge regions of the bZIP domains of the OsbZIP transcription factors.

**Supplemental Figure S2.** Alignment of basic and hinge regions of OsbZIP proteins. The OsbZIP proteins are classified into 11 groups (I-XI) according to their DNA binding specificity and amino acid sequences in basic and hinge regions.

**Supplemental Figure S3.** Amino acid sequence alignment of the leucine zipper region of 89 OsbZIP proteins. The OsbZIP proteins are classified into 29 sub-families (BZ1-BZ29) with similar predicted dimerization properties.

**Supplemental Figure S4.** Phylogenetic relationship among the OsbZIP proteins based on (a) bZIP domain (b) basic and hinge regions (c) leucine zipper defined from the first leucine.
**Supplemental Figure S5.** Phylogenetic relationship among the OsbZIP proteins and other known plant bZIP proteins.

**Supplemental Figure S6.** Organ-specific expression of *OsbZIP* genes based on the microarray analysis from an earlier study (Ma et al., 2005).

**Supplemental Figure S7.** Hierarchical clustering display of 85 *OsbZIP* genes represented on Affymetrix rice genome array in various rice organs and developmental stages (mentioned at the top of each lane).

**Supplemental Figure S8.** Real-time PCR analysis of representative *OsbZIP* genes differentially expressed in various tissues/developmental stages and stress treatments.

**Supplemental Figure S9.** Venn diagrams representing number of differentially regulated *OsbZIP* genes in various organs under different light conditions based on the microarray analysis from an earlier study (Jiao et al., 2005).

**Supplemental Table S1.** bZIP transcription factors identified in plants.

**Supplemental Table S2.** OsbZIP transcription factor genes. Detailed genomic information including TIGR locus IDs, gene names, ORF length, protein length, position of bZIP domain, number of introns within ORF, intron pattern within the basic and hinge regions of the bZIP domain, genomic locus (chromosomal location and BAC/PAC accession no.) and group on the basis of DNA-binding site specificity is provided for each *OsbZIP* gene.

**Supplemental Table S3.** *OsbZIP* genes present on duplicated chromosomal segments of rice.

**Supplemental Table S4.** Additional conserved motifs in OsbZIP proteins as predicted by MEME.

**Supplemental Table S5.** Classification of OsbZIP proteins into sub-families with similar predicted dimerization specificity.

**Supplemental Table S6.** Availability of Full Length-cDNA and/or EST corresponding to *OsbZIP* genes.

**Supplemental Table S7.** *OsbZIP* genes expressed in at least one of the six organs or cell types.

**Supplemental Table S8.** Average log signal values of 85 *OsbZIP* genes from three biological replicates of each sample.
Supplemental Table S9. Relationship between expression pattern of OsbZIP genes and intron pattern, conserved motifs, gene duplication, predicted DNA-binding specificity and phylogenetic analysis.

Supplemental Table S10. OsbZIP genes differentially expressed under different light conditions.

Supplemental Table S11. Average log signal values of 37 OsbZIP genes differentially expressed under various abiotic stress conditions.

Supplemental Table S12. Relationship between expression pattern of OsbZIP genes and hetero-dimerization of OsbZIP transcription factors.

Supplemental Table S13. Primer sequences used for real-time PCR analysis.

LITERATURE CITED


Hobo T, Asada M, Kowyama Y, Hattori T (1999b) ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent. Plant J 19: 679-689.


Suckow M, von Wilcken-Bergmann B, Muller-Hill B (1993a) Identification of three residues in the basic regions of the bZIP proteins GCN4, C/EBP and TAF-1 that are involved in specific DNA binding. EMBO J 12: 1193-1200.


Figure Legends

**Figure 1.** Intron distribution within the basic and hinge regions of the bZIP domains of OsbZIP proteins. Intron distribution patterns (a-g) are depicted. An example of sequence in the basic-hinge region is shown on the top with arrows indicating the position of two introns (as observed in majority of OsbZIP proteins), interrupted by black vertical lines in the sequence and the bars (representing the sequence in different intron patterns). The number of introns and the number of OsbZIP proteins having a particular pattern is also indicated. P0 and P2 indicate the splicing phases of the basic-hinge regions of the bZIP domains, where P0 and P2 refer to phase 0 and phase 2, respectively.

**Figure 2.** Genomic distribution of OsbZIP genes on rice chromosomes. OsbZIP genes are numbered 1 to 89. Yellow ovals on the chromosomes (vertical bar) indicate the position of centromeres. Chromosome numbers are indicated at the top of each bar and the number in the bracket corresponds to the number of OsbZIP genes present on that chromosome. The OsbZIP genes present on duplicated chromosomal segments are connected by coloured lines (one colour per chromosome). The red and blue triangles indicate the upward and downward direction of transcription, respectively. The position of OsbZIP genes on TIGR rice chromosome pseudomolecules (release 5) is given in the Supplemental Table S2.

**Figure 3.** Protein structure of representative OsbZIP proteins based on the position of bZIP domain and presence of additional conserved motifs outside the bZIP domain as identified by MEME. The bZIP domains are shown in blue except the unusual bZIP domain which is shown in pink. Different motifs are highlighted in different colour boxes with numbers 1 to 25, where same number refers to same motif present in the different OsbZIP proteins. OsbZIP protein numbers having similar protein structure are given in front of each structure on the right side. The details of predicted conserved motifs are given in Supplemental Table S4.

**Figure 4.** Amino acid analysis at g, e, a and d positions of the leucine zipper. (A) Histogram of frequency of amino acids in the g, e, a and d positions of the leucine zipper
for OsbZIP, AthZIP and HsbZIP proteins. (B) Histogram of the frequency of asparagine residue in a position of the leucine zippers for all OsbZIP proteins.

**Figure 5.** Histogram of the frequency of attractive or repulsive $g\leftrightarrow e'$ pairs per heptad for all OsbZIP proteins.

**Figure 6.** Phylogenetic relationship among the OsbZIP proteins. The phylogenetic tree is based on the sequence alignments of the OsbZIP proteins. The unrooted tree was generated using ClustalX program by neighbor-joining method. Bootstrap values from 1,000 replicates are indicated at each node. OsbZIP proteins are grouped into 10 distinct clades (A-J).

**Figure 7.** Expression profiles of *OsbZIP* genes differentially expressed during panicle and seed development. Expression profiles of 9 up-regulated (A) and 2 down-regulated (B) *OsbZIP* genes in SAM or any stage of panicle development (P1-I to P1-III and P1 to P6) as compared to vegetative reference tissues/organs (seedling, root, mature leaf, and Y leaf) and seed developmental stages (S1 to S5) are presented. Expression profiles of 9 up-regulated (C) and 1 down-regulated (D) *OsbZIP* genes in any stage of seed development (S1-S5) compared to vegetative reference tissues/organs (seedling, root, mature leaf, and Y leaf), SAM, and panicle developmental stages (P1-I to P6) are presented. The average log signal values of *OsbZIP* genes in various tissues/organs and developmental stages (mentioned at the top of each lane) are presented by cluster display. The color scale (representing log signal values) is shown at the bottom.

**Figure 8.** Differential expression of *OsbZIP* genes under different light conditions based on microarray analysis from an earlier study (Jiao et al. 2005). (A) Number of genes differentially expressed in various organs (mentioned below each bar) under white light. (B) Number of differentially expressed genes in seedling under different light conditions (mentioned below each bar; F: Far-red, R: Red, B: Blue).
Figure 9. Expression profiles of OsbZIP genes differentially expressed under abiotic stress conditions. Expression profiles of 26 significantly up-regulated (A) and 11 significantly down-regulated (B) OsbZIP genes as compared to the control seedlings. The average log signal values of OsbZIP genes under control and various stress conditions (mentioned at the top of each lane) are presented by cluster display (values are given in Supplemental Table S11). Only those genes that exhibited 2-fold or more differential expression with a $p$ value of less than or equal to 0.05, under any of the given abiotic stress conditions are shown. The color scale (representing log signal values) is shown at the bottom of each. (C, D) Venn diagram of the number of differentially expressed OsbZIP genes under abiotic stress conditions. Number of up-regulated (C) and down-regulated (D) OsbZIP genes in any of the abiotic stress conditions as compared to control seedlings are given.
Table 1. Characteristic features of the OsbZIP transcription factors classified into 11 groups (I-XI) according to their DNA binding specificity and amino acid sequences in basic and hinge regions.

<table>
<thead>
<tr>
<th>Group (no. of members)</th>
<th>Characteristic features</th>
<th>Putative Binding Site</th>
<th>Known binding sites in rice</th>
<th>Known binding sites in other species</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (7)</td>
<td>Conserved residues in positions -18 (N), -15 (S), -14 (A), -11 (S) and -10 (R). Have a QA(Q/E)E(A,T)E(D)/E(D,Q) hinge sequence specific to GBFs (Foster et al., 1994) and the RKQS conserved sequence in the basic region</td>
<td>G-box and/or G-box-like sequences</td>
<td>G-box and/or G-box-like sequences for OSBZ8 (OsbZIP05) and OsZIP-1a (OsbZIP86) (Nakagawa et al., 1996; Nantel and Quatrano, 1996).</td>
<td>wheat EmBP-1 (Niu et al., 1999)</td>
</tr>
<tr>
<td>II (3)</td>
<td>NRVSAQQAR sequence in their basic region</td>
<td>TGACGT-containing sequences (e.g. CRE/GA, Hex, Hexm2), some G-box-like sequences (e.g. G-box and Hexm1)</td>
<td>Not known</td>
<td>soybean STF1 (Cheong et al., 1998)</td>
</tr>
<tr>
<td>III (6)</td>
<td>Key residues in the basic region (RNRE/DSxxSR) and a distinct hinge region, i.e. YVE/K(E/D)</td>
<td>G- and C-boxes with equal affinity</td>
<td>Not known</td>
<td>tobacco TGA1b (Katagiri et al., 1989; Niu et al., 1999)</td>
</tr>
<tr>
<td>IV (22)</td>
<td>Specific hinge region sequence Qi(A)i(Q,H)H/(R,Y,A)L/(M,V)E/(T,A,N,K,)E(D)</td>
<td>Hybrid ACGT elements like G/C, G/A C/G boxes</td>
<td>G-, A- and C-box for RITA-1 (OsbZIP20) (Izawa et al., 1994; Niu et al., 1999), GCCACGT(C/A)AG (G/C and A/G hybrid box) for REB (OsbZIP33) (Nakase et al., 1997), RISRZ1 (OsbZIP58), RISRBZ4 (OsbZIP15), RISRBZ5 (OsbZIP52) recognize GCN4 motif (TGA(G/C)TCA) (Onodera et al., 2001), C/G hybrid for LIP19 (OsbZIP38)/OsOBF1 (OsbZIP87) hetero-dimer, G/C hybrid (hex) for OsOBF1 (OsbZIP87) homo-dimer (Shimizu et al., 2005).</td>
<td>ABREs for TRAB1 (OsbZIP66) (Hobo et al., 1999a)</td>
</tr>
<tr>
<td>V (3)</td>
<td>Basic region has A residue at -19 position and hinge region has a conserved QYISE sequence</td>
<td>Relaxed specificity or may bind to other unknown sequences</td>
<td>Not known</td>
<td>AebZIP34 and AebZIP61</td>
</tr>
<tr>
<td>VI (14)</td>
<td>Highly conserved motifs like M/A-I/M-K and Q/R-A-Y/R in the basic and hinge regions, respectively</td>
<td>ABA-responsive elements (ABREs) with a consensus sequence (T/G/C)ACGT(G/T)GC or others containing GCGT or AAGT instead of ACGT (Ono et al., 1996; Hobo et al., 1999b)</td>
<td>ABREs for TRAB1 (OsbZIP66) (Hobo et al., 1999a)</td>
<td>Not known</td>
</tr>
<tr>
<td>VII (16)</td>
<td>Conserved residues in positions -21 (L), -20 (A), -19 (Q), -18 (N), -15 (A), -14 (A), -12 (K), -11 (S) and -10 (R). Possess a KAY/(H)V/I/QQ/(N) hinge sequence specific to CBFs (Foster et al., 1994)</td>
<td>C-box sequence</td>
<td>Not known</td>
<td>tobacco TGA1a (Niu et al., 1999)</td>
</tr>
<tr>
<td></td>
<td>Conserved residue in position -15 (A) specific to CBFs</td>
<td>C-box elements preferentially Not known</td>
<td>Not known</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----------------------------------------------------</td>
<td>----------------------------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>VIII</td>
<td>Conserved lysine substitution at -10 position of the basic region instead of arginine</td>
<td>Sequences other than those containing a palindromic ACGT core CCA(N)\textsubscript{2}TGG for RF2a (OsbZIP75)\textsubscript{2}(Yin et al., 1997a), Box II for RF2a (OsbZIP75) and RF2b (OsbZIP30)(Yin et al., 1997b; Dai et al., 2004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IX</td>
<td>Lack the conserved asparagine in position -18</td>
<td>Might not be able to bind to DNA as homodimers</td>
<td>Not known</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>Hydrophobic isoleucine residue at position -10 instead of a conserved arginine/lysine</td>
<td>Might not be able to bind DNA or else possess a uniquely different DNA-binding specificity</td>
<td>corresponds to OsZIP-2a reported earlier (Nantel and Quatrano, 1996)</td>
<td>Not known</td>
</tr>
</tbody>
</table>
Figure 1
Protein Structure

OsbZIP proteins having similar protein structure

OsbZIP67, 73
OsbZIP15, 20
OsbZIP05, 45
OsbZIP34, 59
OsbZIP13, 26, 86
OsbZIP54, 55, 56
OsbZIP09, 10, 23, 42, 46, 72
OsbZIP62
OsbZIP29
OsbZIP66
OsbZIP12, 40
OsbZIP02, 06, 19
OsbZIP53
OsbZIP44
OsbZIP07
OsbZIP17
OsbZIP64
OsbZIP03, 08, 11, 28, 37, 41, 43, 47, 49, 63, 70, 74, 79, 83
OsbZIP39, 60
OsbZIP01, 18, 48
OsbZIP25, 30, 35, 36, 61, 68, 75, 76, 78, 81, 84

Figure 3
Figure 4
Figure 5
Figure 6
Figure 7
Figure 8
Figure 9