Genome-wide identification, classification, evolutionary expansion and expression analyses of homeobox genes in rice

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, India

Homeobox genes play a critical role in regulating various aspects of plant growth and development. In the present study, we identified a total of 107 homeobox genes in the rice genome and grouped them into ten distinct subfamilies based upon their domain composition and phylogenetic analysis. A significantly large number of homeobox genes are located in the duplicated segments of the rice genome, which suggests that the expansion of homeobox gene family, in large part, might have occurred due to segmental duplications in rice. Furthermore, microarray analysis was performed to elucidate the expression profiles of these genes in different tissues and during various stages of vegetative and reproductive development. Several genes with predominant expression during various stages of panicle and seed development were identified. At least 37 homeobox genes were found to be differentially expressed significantly (more than two-fold; \( P < 0.05 \)) under various abiotic stress conditions. The results of the study suggest a critical role of homeobox genes in reproductive development and abiotic stress signaling in rice, and will facilitate the selection of candidate genes of agronomic importance for functional validation.

The homeobox genes are key regulators of cell fate and body plan specification at the early stages of embryogenesis in higher organisms. Homeobox genes contain a conserved 180 bp long DNA sequence termed a homeobox, which encodes a 60 amino acid long DNA-binding domain termed a homeodomain (HD). The HD consists of three \( \alpha \)-helices that form a helix-turn-helix DNA-binding motif [1,2]. This motif recognizes and binds to specific DNA sequences to regulate the expression of target genes. HD-containing proteins have been identified in diverse organisms, such as humans, *Drosophila*, nematode and plants [3–5].

Several homeobox genes have been identified in plants and catalogued into different groups, based on the amino acid sequence of HD and the presence of other conserved motifs [4]. The plant homeobox genes have been implicated in various developmental processes and hormone response pathways [4,6,7]. Convincing evidence is available demonstrating that homeobox genes are involved in abiotic and biotic stress responses as well [6,8–12]. The members of class I *KNODDEDI*-like homeobox (*KNOX*) genes are typically expressed in shoot apical meristem (SAM) and control the balance between meristematic and determinate growth during plant development [13–16]. Loss-of-function mutation in one of the class I *KNOX* genes, *SHOOT MERISTEMLESS* in *Arabidopsis* and *KNOTTED1* in maize, results in the development of embryos lacking SAM [15,17,18]. The *WUSCHEL*-like
homeobox (WOX) class of genes appears to be essential for embryonic patterning [19,20]. The homeobox gene WUSCHEL (WUS) was originally identified as a central regulator of shoot and floral meristems in Arabidopsis [21]. It has been shown that a WUS-like gene is involved in root apical meristem formation in rice [22]. BELL-type HD proteins are involved in pattern formation and development of flowers, fruits and tubers [23–27]. The class III HD-leucine zipper (HD-ZIP) homeobox genes appear to have a role in shoot meristem formation, vascular development and establishing and/or maintaining abaxial/adaxial polarity in leaves and embryo [28–33]. Zinc finger-HD (ZF-HD) gene family members play overlapping regulatory roles in floral development in Arabidopsis [34].

Although several homeobox genes have been isolated from rice and a few of them also have been functionally characterized, a genome-wide analysis has not been performed so far. In the present study, we identified 107 homeobox genes in rice, which were classified into ten subfamilies on the basis of their phylogenetic relationship and domain composition. Comprehensive expression analysis shows overlapping and/or specific expression patterns of these genes in the various rice tissues/organs and/or developmental stages analyzed. Furthermore, we show that the expression of a large number of rice homeobox genes is regulated by various abiotic stresses.

Results and Discussion

Identification and classification of homeobox genes in rice

In plants, homeobox genes are represented by a large multigene family [4,35,36]. In Arabidopsis, approximately 100 homeobox genes were identified and classified into several groups based on their domain composition and phylogenetic relationship [4]. In the present study, the Hidden Markov Model profile and keyword searches followed by domain analysis using SMART and Pfam databases resulted in the identification of 107 nonredundant potential homeobox proteins in the rice (Oryza sativa subsp. japonica cv Nipponbare) genome. To examine the evolutionary relationship among the predicted rice homeobox proteins, a phylogenetic tree was generated from alignments of their HD sequences by the Neighbour-joining method. This analysis resulted in the formation of ten distinct clades (Fig. 1). Based on the domain composition and phylogenetic relationship of their encoded proteins, the rice homeobox genes were divided into ten subfamilies (Table 1). The subfamily, tigr locus, domain composition, ORF length, protein length and chromosomal location of each predicted homeobox gene are given in the supplementary Table S1. A BLASTp search of these 107 homeobox proteins in the annotated proteins of indica rice (cv 93–11) genome revealed that most (at least 100) of these proteins are conserved in both subspecies (data not shown). At least 72 homeobox proteins predicted from japonica rice showed ≥ 90% identity over the entire length with the annotated proteins in indica rice. This number may increase once a more exhaustive and refined annotation of the indica genome becomes available.

The largest number (48) of homeobox proteins were classified into the HD-ZIP family, which was further subdivided into four distinct subfamilies, termed HD-ZIP I (14 members), HD-ZIP II (13 members), HD-ZIP III (nine members) and HD-ZIP IV (12 members), based on their domain composition and phylogenetic relationship (Table 1 and Fig. 1). All but one member of HD-ZIP I and HD-ZIP II subfamilies harbor a plant-specific leucine zipper domain, termed the HALZ (homeobox associated leucine zipper) domain, which is associated with the homeobox. The presence of leucine zipper motif in plant homeobox proteins is markedly different from the case in animal systems in which none of the homeobox genes examined contain a leucine zipper [37]. Leucine zippers may allow the HD-ZIP proteins to interact with each other and other leucine zipper proteins, which may be important for their function. The members of HD-ZIP III and HD-ZIP IV subfamilies encode an additional domain, termed the START (steroidogenic acute regulatory protein-related lipid-transfer) domain, which is a putative lipid-binding domain [38,39]. Recently, a novel domain, MEKHLA, with significant similarity to the PAS domain, was identified at the C-terminus of HD-ZIP III proteins [40]. Four rice homeobox proteins grouped into the HD-ZIP III subfamily also harbor the MEKHLA domain, which was proposed to function as a sensory domain [40].

The BEL1-like homeobox (BLH) subfamily comprises 13 homeobox genes, and shows homology to Arabidopsis BEL1 protein. These proteins harbor a domain of unknown function towards the N-terminal region of the HD, termed the POX domain, which is found exclusively in plant proteins associated with HD [41]. KNOX genes belong to a superfamily TALE (three amino acid loop extension) because of presence of an atypical HD, which contains three extra amino acid stretches between the first and second helices [42]. Thirteen rice homeobox proteins belong to the KNOX family, which were further divided into two subfamilies, KNOX I (nine members) and KNOX II (four
Fig. 1. Phylogenetic relationship among the rice homeobox proteins. The unrooted tree was generated from multiple sequence alignments of HD sequences. The sequences were aligned using CLUSTALX by the Neighbour-joining method. Ten distinct clades of homeobox proteins identified are represented by shading. The bootstrap values (> 40%) from 1000 replicates are indicated at each node.
members), based on the amino acid sequence of HD and phylogenetic analysis. These proteins contain ELK, KNOX1 and KNOX2 domains other than HD, which are required for nuclear localization, suppressing target gene expression and homo-dimerization, respectively [43,44].

The WOX subfamily consists of 15 members in rice, and shows strong homology to Arabidopsis WUS protein. A similar number of WOX genes has been identified in Arabidopsis [19]. Several members of the WOX subfamily have been identified in rice and maize based on the phylogenetic identification of WUS orthologs [45]. The WOX proteins do not contain any known domain other than HD. The expression dynamics showed that WOX genes mark cell fate decisions during early embryo patterning in Arabidopsis and maize [19,20]. Fourteen rice homeobox proteins have been grouped into the ZF-HD subfamily. These proteins contain a conserved region upstream of HD, termed the ZF-HD domain. This region is involved in protein–protein interaction by mediating homo- and hetero-dimerization [46]. Recently, it was demonstrated that ZF-HD genes play overlapping regulatory roles in floral development in Arabidopsis [34]. The plant HD (PHD) subfamily of homeobox genes is represented by only two members (Os02g05450 and Os06g12400) in rice. The proteins encoded by this subfamily contain a Cys4-His-Cys3 type zinc-finger domain, termed the PHD finger domain N-terminal to the HD [47,48]. Although the function of this domain is not yet known, its analogy with the LIM domain suggests that it could be involved in protein–protein interaction and be important for the assembly or activity of multicomponent complexes involved in transcriptional activation or repression [48].

Two homeobox genes, Os05g09630 and Os05g50130, although apparently closely related to HD-ZIP sub-
but 12 homeobox genes are indeed interrupted by one to 17 introns (supplementary Table S1). All the members of BLH subfamily except for Os06g36660 are disrupted by three introns at perfectly conserved positions with respect to their amino acid sequence. Quite a few (eight of 14) of the members of ZF-HD subfamily are intronless. The members of other subfamilies harbor highly variable numbers of introns. The largest number of introns is present in the members of HD-ZIP III subfamily. Genes with multiple exons and introns can be regulated by alternative splicing, which was proposed as a mechanism to expand the proteomic diversity within a genome [49]. Interestingly, 27% (29 of 107) of homeobox genes were predicted to be alternatively spliced by tigr, which is slightly higher than that predicted for rice genes overall [50,51]. These homeobox genes exist in two to four alternatively spliced forms, giving rise to a total of 74 transcripts (supplementary Table S2). The alternative splicing of these genes was validated manually or using the tigr program to assemble spliced alignments by alignment of rice full-length cDNA (FL-cDNA) and/or expressed sequence tag (EST) sequences [50]. Notably, different alternatively spliced transcripts of these genes have been derived from various tissues/organs of rice (data not shown), indicating their differential expression. Several alternatively spliced forms of Os02g08544/HOS58, Os06g43860/HOS59 and Os03g03164/HOS66 have been reported previously that exhibited organ-specific expression patterns [52]. A more detailed analysis of developmental and temporal differential expression of these and other alternatively spliced homeobox genes will help in the understanding of their post-transcriptional regulation.

**Evolutionary expansion**

The plant genomes have evolved essentially through polyploidization and segmental and tandem duplications, which have great impact on the amplification of members of a gene family in the genome. The *Arabidopsis* and rice genome sequences revealed that a majority of plant genes belong to large gene families. Both tandem gene duplication and chromosomal segmental duplication followed by dispersal and diversification are ascribed for the expansion and evolution of gene families [53–55]. The large size of the homeobox gene family indicates that it has evolved through a large number of duplication events in rice. Therefore, we studied the contribution of segmental and tandem duplications in homeobox gene family expansion. Interestingly, a very large number (54; 50.5%) of homeobox genes were present on duplicated chromosomal segments of rice (supplementary Table S3). All the homeobox gene-containing chromosomal segments have a homeobox gene in its duplicate block. This suggests that all the homeobox genes have been retained in rice after segmental duplications. All but two of the homeobox genes located on duplicated segments belong to same subfamily. However, two homeobox genes present on duplicated chromosomal segments between chromosomes 1 and 5 (Os01g45570 and Os05g50130) belong to different subfamilies. Additionally, 12 genes were found to be tandemly duplicated (representing six individual duplication events), which were separated by a maximum of three intervening genes. In all six cases, only two genes were present in tandem. Eight of 12 tandemly duplicated genes were present on chromosome 3; whereas two genes were present each on chromosomes 6 and 9. At all positions, the homeobox genes present in tandem belonged to same subfamilies. Because the number of homeobox genes present on duplicated chromosomal segments is much higher than those present in tandem, the segmental duplications appear to have played a major role in expansion of this gene family.

To analyze the evolutionary relationship among rice and *Arabidopsis* homeobox proteins, an unrooted phylogenetic tree was generated from the alignments of their HD sequences. Based on their sequence homology, and with few exceptions, all the homeobox proteins clustered into distinct clades representing different subfamilies, similar to rice proteins (supplementary Fig. S1). Because all the clades contain representatives from both rice (monocotyledonous) and *Arabidopsis* (dicotyledonous), a common ancestor of each subfamily must have existed before the divergence of monocot and dicot lineages. However, within a clade, species-specific clustering of homeobox proteins was observed, which indicates that the expansion of homeobox subfamilies has occurred independently in rice and *Arabidopsis* after their divergence by duplication (segmental or tandem) of common ancestral genes. Similar examples of gene family expansion have been reported previously [53–56].

**Gene expression during reproductive development**

Several approaches were employed for expression analysis of rice homeobox genes. In the first approach, evidence for the expression of homeobox genes was provided by the availability of any corresponding FL-cDNA and/or ESTs in the databases. Ninety-two
of 107 (86%) homeobox genes have at least one corresponding FL-cDNA and/or EST sequence (supplementary Table S1), indicating that most of these genes are expressed in rice. Fifty-nine homeobox genes have both FL-cDNA and EST evidence, whereas three and 30 genes have only FL-cDNA and EST evidence, respectively. However, the frequency of matched FL-cDNA and/or ESTs varies greatly and they were derived from various rice tissues, indicative of the differential expression of homeobox genes.

DNA microarrays provide a high-throughput means to analyze gene expression at the whole genome level or a subset of genes of interest. In the second approach, the gene expression profiling of homeobox genes was achieved by microarray analysis performed using Affymetrix rice whole genome arrays, as described previously [55]. The various vegetative and reproductive developmental stages of rice used for microarray analysis include seedling, seedling root, mature leaf, Y leaf (fully expanded youngest leaf subtending the shoot apical meristem), SAM and various stages of panicle (P1-I to P1-III and P1 to P6) and seed (S1 to S5) development. Different developmental stages of panicle and seed development have been categorized according to panicle length and days after pollination (DAP), respectively, based on the landmark developmental event(s) (supplementary Table S4) [57]. We have used the RNA samples from indica rice for hybridization on the Affymetrix rice genome arrays, which have been designed primarily based on the gene sequences from japonica rice. Several other studies also report the successful use of these arrays for studying gene expression in different varieties of both japonica and indica rice [55,58–60].

Following whole-chip data processing, the log signal values for 93 homeobox genes, represented on the array in all the developmental stages analyzed, were extracted. A hierarchical cluster display generated from the average log signal values indicates the differential expression profiles of homeobox genes (Fig. 3A; the average log signal values are provided in supplementary Table S5). Moreover, the signal values indicate that most of homeobox genes are expressed in at least one of the rice vegetative and/or reproductive stage(s) of development analyzed. Subsequently, differential expression analysis was performed in two steps to identify the homeobox genes exhibiting stage-specific expression. In the first step, the homeobox genes differentially expressed during any of the reproductive development stages (SAM, P1-I to P1-III, P1 to P6 and S1 to S5) were identified compared to all the four vegetative stages (seedling, root, mature leaf and Y leaf). This analysis revealed that a large number (55) of homeobox genes were differentially expressed in at least one of the stages of reproductive development compared to vegetative stages. Further analysis revealed that, out of these 55 genes, 33 and four genes were differentially expressed during panicle and seed development, respectively. However, 18 genes were differentially expressed in both panicle and seed development stages. In the second step, the homeobox genes differentially expressed at any stage(s) of the panicle development were identified compared to seed development stages. Similarly, the differentially expressed homeobox genes at any stage(s) of the seed development compared to panicle development stages were identified. This analysis revealed that 32 (28 up- and four down-regulated) and seven (four up- and three down-regulated) homeobox genes were differentially expressed in at least one of the stages of panicle and seed development, respectively (Fig. 3A). To validate the differential expression of homeobox genes as revealed from microarray analysis, real-time PCR analysis was performed for some representative genes in all the tissues/organs and developmental stages analyzed from indica rice. The real-time PCR results were found to be in very good agreement with the microarray data (Fig. 3B). The expression profiles of various genes obtained from our microarray data also have been validated by real-time PCR analysis in other studies [61–63].

Massively parallel signature sequencing (MPSS) generates hundreds of thousands of molecules per reaction and provides a quantitative assessment of transcript levels for each candidate gene in different tissue samples were calculated relative to its expression in seedlings. S, seedling; R, root; ML, mature leaf; YL, Y leaf; P1-I to P1-III and P1 to P6, stages of panicle development; S1 to S5, stages of seed development.

![Fig. 3. Expression profiles of homeobox genes in various tissues/organs and developmental stages of rice.](image-url)
abundance [64]. In the third approach, we investigated the expression of 14 homeobox genes for which microarray data were unavailable in the mRNA MPSS database of rice (http://mpss.udel.edu/rice/) [65]. MPSS data from 22 libraries (17 base signature) representing 18 different tissues/organisms of rice was used for this analysis. MPSS signatures were available for all the 14 genes in at least one of the libraries, indicating their expression. However, the significant signatures (i.e. that uniquely identify individual gene) were found only for six homeobox genes, which showed low to moderate expression levels (Fig. 4). Among these, Os01g44430 was highly expressed in salt and cold stressed 14-day-old young leaves, whereas Os05g25600 transcript was abundant in the meristematic tissue. Os12g10630 showed significant expression in 14-day-old young leaves. Interestingly, Os06g36680 appears to be specifically expressed only in ovary and mature stigma [322 tags per million (tpm)] with marginal expression in germinating seed and immature panicle (7 and 3 tpm, respectively).

In most plant species, the fundamental body plan is established during embryogenesis, and each organ formation occurs successively after germination from shoot and root apical meristems that are established at early stages of embryogenesis. Genetic and morphological characterization of some of the mutants defective in various steps of embryogenesis has indicated the existence of several major developmental processes occurring during embryogenesis in rice [57,66–68]. By analogy to animals, homeobox genes in plants are also thought to mediate important processes during embryogenesis, and there is much evidence available to support this notion. Early evidence for the involvement of plant homeobox genes in embryogenesis came from the analysis of an Arabidopsis embryogenesis defective mutant, stm [17]. SHOOT MERISTEMLESS is expressed in the rudimentary SAM of embryo and its expression is necessary to maintain the integrity of SAM [15,17]. WUS is also expressed very early in embryogenesis before the morphological appearance of embryonic SAM [21]. Some members of the KNOX gene family Oryza sativa homebox (OSH) genes, of rice are expressed in a restricted region of embryo that defines the position at which the SAM would eventually develop, prior to visible organ formation, and their expression continues after seed germination until development of inflorescence meristem and differentiation of floral organs [69]. Due to their specific expression in the SAM, they are considered to be reliable markers for studying plant development [57,70–73]. Our results also show that several members of KNOX I and KNOX II subfamilies are differentially expressed during various stages of panicle and seed development (Fig. 3A). The analysis of expression patterns and loss-of-function alleles of the HD-ZIP III family members in Arabidopsis revealed their roles in meristem initiation, meristem regulation, organ polarity and embryo patterning [28–33]. The members of HD-ZIP III subfamily in rice, Os03g01890, Os03g43930, Os10g33960 and Os12g41860, which represent the orthologs of Arabidopsis PHABULOSA (PHB), PHAVOLUTA (PHV) and/or REVOLUTA (REV) genes (supplementary Fig. S1), are also preferentially expressed during various stages of panicle and seed development (Fig. 3A), suggesting similar roles of these genes in rice. Altogether, we have identified a large number of homeobox genes that are preferentially expressed during various stages of development, including floral transition, floral organ differentiation and development, maturation of male and female reproductive organs, and embryogenesis. It is thus conceivable that these homeobox genes may perform specific roles during different stages of reproductive development in rice. Although our analyses provide evidence for tissue-/organ- and developmental stage-specific expression of homeobox genes, definitive clues for their cell type-specific expression and function will come from in situ hybridization and functional validation in transgenics.
Gene expression during anther and stigma development

Several genes involved in anther development have been identified in *Arabidopsis*. Despite its importance in crop yield and hybrid seed production, very few studies on anther development in cereal plants have been performed. Recently, a basic helix-loop-helix transcription factor, UNDEVELOPED TAPETUM 1 (UDT1), was identified as a major regulator of tapetum (i.e. the innermost sporophytic layer in the anther wall, which is thought to play a crucial role in the development and maturation of microspores) development and pollen mother cell meiosis in rice [74]. Several downstream target genes of UDT1 were identified by microarray analysis [74]. Our further analysis of these microarray data revealed that at least four (*Os08g32080, Os10g01470, Os05g48990 and Os08g37400*) homeobox genes were down-regulated and one (*Os04g45810*) gene up-regulated significantly (more than two-fold) in *udt1-1* anthers compared to wild-type anthers. These genes were differentially expressed at different stages of anther development, including meiosis, young microspore, vacuolated pollen and pollen mitosis, compared to palea/lemma. These genes may interact with UDT1 to regulate anther development in rice.

The stigma, a female reproductive organ, provides nutrients and guidance cues for pollen grain germination and pollen tube growth. Recently, a genome-wide gene expression profiling identified several genes specifically or preferentially expressed in stigma that may regulate these processes in rice [75]. A survey of these genes during the present study identified at least seven (*Os03g08960, Os06g04850, Os06g04870, Os06g39906, Os09g35760, Os01g60270 and Os09g24810*) homeobox genes that express preferentially in rice stigma, indicating their role in stigma development and other processes involved therein.

Expression profiles of duplicated homeobox genes

Gene duplications serve as a mechanism to increase diversity at the molecular level. The mutation in a duplicated locus might not have a morphological or physiological effect, or may be the primary contributor to innovative developmental programs. After duplication, the coding regions may be brought into a new regulatory context via acquisition/deletion of regulatory sequences (i.e. tissue-specific enhancers and repressors), which can cause spatial and/or temporal change in gene expression contributing to diversification of gene function by sub- (one of the gene acquires a novel function) or neo-functionalization (the duplicated genes perform different aspects of the original gene’s function) [76,77]. To study the role of gene duplication in the diversification of gene function, the expression profiles of homeobox genes localized on duplicated chromosomal segments and tandemly duplicated genes were investigated. Out of 27 gene pairs localized on duplicated segments, microarray data for 21 gene pairs were available. Fourteen of the 21 gene pairs exhibited similar gene expression patterns (supplementary Fig. S2), indicating their overlapping functions. However, the expression of one of the duplicated genes was very low in some cases. This may be due to the duplicated genes with low expression losing their function over the course of evolution. Other seven gene pairs exhibited significantly divergent expression patterns for two duplicated genes (supplementary Fig. S2), indicating sub- or neo-functionalization. For example, *Os03g20910* is preferentially expressed in young tissues (seedling and root) and late stages of seed development, whereas *Os07g48560* is expressed at very low level in all the tissues/stages examined. Similarly, the expression of *Os10g39030* was significantly higher during late panicle and early seed developmental stages compared to *Os03g03260*.

Out of six gene pairs representing the tandem gene duplication events, microarray data were available for five gene pairs. Three gene pairs exhibited highly similar expression patterns (supplementary Fig. S3). However, the expression patterns of two gene pairs, *Os03g47730/Os03g47740* and *Os09g24810/Os09g24820* were highly divergent (supplementary Fig. S3). Taken together, these results suggest that homeobox genes appear to have evolved through gene duplication events followed by conservation and sub- or neo-functionalization of the duplicated gene.

Gene expression under abiotic stress conditions

In various plant species, the role of several homeobox genes has been implicated in abiotic stress responses [9,11,78–81]. However, the role of only one rice homeobox gene, *BIHD1/Os03g47740*, in abiotic stress response has been reported to date [12]. To investigate the abiotic stress response of all the rice homeobox genes, microarray analysis of 7-day-old seedlings subjected to desiccation, salt and cold stress treatments was performed. Data analysis revealed that a total of 37 homeobox genes were differentially expressed significantly (at least two-fold; *P* < 0.05) in at least one of the stress conditions examined (Fig. 5A–I and supplementary Table S6). The transcript levels of two genes (*Os04g45810 and Os09g21180*) were up-regulated under
Genomic analyses of homeobox genes in rice

M. Jain et al.

Expression profiles of rice homeobox genes differentially expressed under various abiotic stress conditions. Expression profiles are presented of homeobox genes up-regulated by desiccation and salt stress, and cold stress (A), up-regulated by desiccation and salt stress (B), down-regulated by desiccation and salt stress (C), up-regulated by salt and down-regulated by cold stress (D), up-regulated by desiccation and salt stress (E), up-regulated by salt and down-regulated by cold stress (F), up-regulated by cold stress (G), down-regulated by desiccation and cold stress (H), and down-regulated by cold stress (I) compared to the control seedlings. The average log signal values of homeobox genes under control and various stress conditions (indicated at the top of each lane) are presented by heatmaps. Only those genes that exhibited two-fold or more differential expression with a $P < 0.05$, under any of the given abiotic stress conditions, are shown. The color scale (representing average log signal values) is shown at the bottom. The representative homeobox genes differentially expressed under different abiotic stress conditions for which real-time PCR analysis was performed are indicated by green arrow heads, to the right. (J) Real-time PCR analysis to confirm the differential expression of representative homeobox genes during various abiotic stress conditions. The mRNA levels for each candidate gene in different tissue samples were calculated relative to its expression in control seedlings. C, control; DS, desiccation stress; SS, salt stress; CS, cold stress.

Fig. 5. Expression profiles of rice homeobox genes differentially expressed under various abiotic stress conditions. Expression profiles are presented of homeobox genes up-regulated by desiccation and salt stress, and cold stress (A), up-regulated by desiccation and salt stress (B), down-regulated by desiccation and salt stress (C), up-regulated by salt and down-regulated by cold stress (D), up-regulated by desiccation and salt stress (E), up-regulated by salt and down-regulated by cold stress (F), up-regulated by cold stress (G), down-regulated by desiccation and cold stress (H), and down-regulated by cold stress (I) compared to the control seedlings. The average log signal values of homeobox genes under control and various stress conditions (indicated at the top of each lane) are presented by heatmaps. Only those genes that exhibited two-fold or more differential expression with a $P < 0.05$, under any of the given abiotic stress conditions, are shown. The color scale (representing average log signal values) is shown at the bottom. The representative homeobox genes differentially expressed under different abiotic stress conditions for which real-time PCR analysis was performed are indicated by green arrow heads, to the right. (J) Real-time PCR analysis to confirm the differential expression of representative homeobox genes during various abiotic stress conditions. The mRNA levels for each candidate gene in different tissue samples were calculated relative to its expression in control seedlings. C, control; DS, desiccation stress; SS, salt stress; CS, cold stress.
demonstrates that a large number of rice homebox genes belonging to all subfamilies (except for KNOX II) are involved in abiotic stress responses. These genes may play a significant role in the abiotic stress pathway and provide a valuable resource for generating stress-tolerant transgenic crop plants.

There is much evidence demonstrating the interaction of developmental processes and stress responses [55,83–86]. The interaction between plant development and environmental conditions implies that some genes must be co-regulated by both environmental factors and developmental cues. Cooper et al. [87] reported a network of rice genes that are associated with stress response and seed development. Furthermore, it was demonstrated that a significant number of pollination/fertilization-related genes are indeed regulated by dehydration and wounding in rice [88]. Recently, an interaction network of proteins associated with abiotic stress response and development in wheat was proposed [89]. We also found that 14 (Os03g07450, Os03g10210, Os04g45810, Os07g39320, Os03g55990, Os09g35760, Os02g13310, Os03g47730, Os01g60270, Os01g19694, Os03g51690, Os07g03770, Os05g50310 and Os11g39360) and three (Os08g37580, Os09g21180 and Os02g35770) homeobox genes that were differentially expressed during at least one of the panicle and seed developmental stages, respectively, are regulated by one or more of the stress conditions. This suggests that a number of candidate homeobox genes are likely to be involved in critical developmental processes and stress responses, but their direct relationship requires experimental validation. Such genes may act as mediators of plant growth response to different abiotic stress conditions during various developmental stages.

**Defining the roles of homebox proteins in regulatory networks**

Homebox genes encode transcription factors, which are the key regulators of various aspects of plant development. Similar to other multigene families, many homeobox genes probably have overlapping functions that complicate the analysis of their mutant and/or transgenic phenotypes. Our identification of all the homeobox genes in rice genome is thus a prerequisite for the dissection of individual homeobox gene functions. On the basis of domain organization and phylogenetic relationship, we have defined ten groups of related homeobox proteins in which functional overlaps are more plausible. The duplicated genes might have partially redundant functions and the identification of duplicated genes will have important implications in the study of gene functions and the evolutionary consequences of gene duplication. To date, only a few homeobox genes have been functionally characterized in rice. The analysis presented in the present study suggests a crucial role of homeobox genes in reproductive development and abiotic stress responses, and should act as a major step towards a comprehensive functional characterization of the homeobox gene family in rice and other plant species.

Finally, it will be useful to identify those factors that function with or regulate homeobox proteins by protein–protein interaction studies. Together with the availability of complete rice genome sequence and the increasing ease of obtaining mutants and raising transgenics, our analysis should stimulate future studies on homeobox gene function.

**Experimental procedures**

**Plant materials**

Rice (O. sativa L. ssp. indica var. IR64) seeds were disinfected and grown as described previously [55]. Rice plants were grown under greenhouse or field conditions for collecting tissue samples of mature leaf, Y leaf and different seed developmental stages, respectively, are regulated by one or more of the stress conditions. This suggests that a number of candidate homeobox genes are likely to be involved in critical developmental processes and stress responses, but their direct relationship requires experimental validation. Such genes may act as mediators of plant growth response to different abiotic stress conditions during various developmental stages.

Rice plants were grown under greenhouse or field conditions for collecting tissue samples of mature leaf, Y leaf and different seed developmental stages, respectively, are regulated by one or more of the stress conditions. This suggests that a number of candidate homeobox genes are likely to be involved in critical developmental processes and stress responses, but their direct relationship requires experimental validation. Such genes may act as mediators of plant growth response to different abiotic stress conditions during various developmental stages.

**Database search**

The Hidden Markov Model profile (build 2.3.2) of homebox domain (PF00046) generated by alignments of 188 seed sequences, was downloaded from Pfam (http://www.sanger.ac.uk/Software/Pfam). This profile was utilized to identify all the homeobox proteins encoded by the rice genome by searching against the annotated proteins in whole rice genome tigr (release 5; http://www.tigr.org/tdb/e2k1/osaily) with an e-value cut off of 1.0. This search resulted in the identification of 160 nontransposable element proteins. Of the 160 proteins, 54 proteins were removed because they represented different gene models present at the same locus in rice genome. Among the remaining nonredundant set of 106 proteins, only 93 showed the presence of homeobox domain with confidence (e-value < 1.0) by SMART/Pfam (http://smart.embl-heidelberg.de), when checked individually.
Similarly, 14 more nonredundant homebox proteins classified as zinc-finger homebox proteins were identified by a Pfam profile search for ZF-HD dimerization domain (PF04770). In the present study, we removed the LOC prefix from all tigr locus IDs representing homebox proteins for convenience. Domains in homebox proteins were identified using SMART and Pfam with an e-value cut off of 1.0. For the blast search in the indica rice (cv 93–11) genome, the annotation available at the BGI Rise Rice Genome Database (http://rise.genomics.org.cn) [90] was used.

**Phylogenetic analysis**

Multiple sequence alignments of homebox domain identified by SMART from all the protein sequences were performed using clustalx, version 1.83 [91]. The unrooted phylogenetic trees were constructed by the Neighbour-joining method [92] and displayed using njplot [93].

**Localization of homebox genes on rice chromosomes**

The position of each of the homeobox genes on the rice chromosome pseudomolecules available at TIGR was determined by a BLASTN search. The presence of homeobox genes on duplicated chromosomal segments was investigated by segmental genome duplication of rice available at TIGR with the maximum length distance permitted between collinear gene pairs of 500 kb. The genes separated by a maximum of five genes were considered as tandemly duplicated genes.

**FL-cDNA and EST evidence search**

The gene expression evidence page, available at TIGR rice genome annotation database (http://www.tigr.org/tdb/e2k1/osa1/locus_expression_evidence.shtml), was used to find the availability of any FL-cDNA and/or EST sequence(s) corresponding to each of the homeobox genes.

**Microarray hybridization and data analysis**

Microarray analysis was performed as described previously [55] using Affymetrix GeneChip Rice Genome Arrays (Affymetrix Inc., Santa Clara, CA, USA) representing 49 824 transcripts (48 564 of japonica and 1260 of indica). At least three independent biological replicates of each tissue sample were used for microarray analysis. For data analysis, the cel files generated by Genechip Operating Software were imported into ARRAYASSIST (version 5.0) (Stratagene, La Jolla, CA, USA). The normalization, probe summarization and variance stabilization were performed as described previously [55]. Three biological replicates of each tissue sample with an overall correlation coefficient value of more than 0.94 were selected for final analysis. The microarray data for 22 tissue samples (66 arrays in total) were included in the final analysis. These samples included 19 samples for the rice development series and four samples for the biotic stress series (the control seedling sample was used in both development and stress series data analysis). The microarray data have been deposited in the Gene Expression Omnibus database at the National Center for Biotechnology Information under the series accession numbers GSE6893 and GSE6901.

The probe sets representing the homeobox genes on the Affymetrix rice genome array were identified by a Rice Multi-platform Microarray Search (http://www.ricearray.org/matrix.search.shtml). Probe sets with the entire set of 11 probes (eight to ten in some cases) present on the array aligned with 100% identity over the entire length with corresponding homeobox gene were considered to be significant. Data for only one probe set for each homeobox gene were used for expression analysis. The probes for 93 (out of 107) homeobox genes could be identified. The log signal intensity values for rice probe IDs corresponding to homeobox genes were extracted and used for further analyses. Hierarchical clustering was performed using Euclidean distance metric and complete linkage rule. The genes that are up- or down-regulated at least two-fold with $P < 0.05$ were considered to be differentially expressed significantly.

**Real-time PCR analysis**

To confirm the expression patterns of representative genes obtained by microarray analysis, real-time PCR analysis was performed using gene-specific primers as described previously [94]. The primer sequences used for real-time PCR analysis are given in the supplementary Table S7. At least two independent biological replicates of each sample and three technical replicates of each biological replicate were used for real-time PCR analysis. The expression of each gene in different RNA samples was normalized with the expression of the internal control gene, UBQ5. The mRNA levels for each candidate gene in different tissue samples were calculated using the $\Delta\Delta$CT method (Applied Biosystems, Foster City, CA, USA). Values are the mean of two biological replicates, each with three technical replicates. Error bars indicate the standard deviation.

**MPSS data analysis**

Expression evidence for mRNA from MPSS tags was determined from the Rice MPSS project mapped to TIGR gene models [http://mpss.udel.edu/rice/] [65]. The MPSS data for 17-base significant signatures (classes 1, 3, 5 and 7 that identify the sense strand), which uniquely identify an individual gene and show a perfect match (100% identity over 100% of the length of the tag), were retrieved. The normalized abundance (tpm) of these signatures for a given gene
in a given library represents a quantitative estimate of expression of that gene. MPSS expression data from 22 mRNA libraries, representing 18 different tissues/organisms, were used for the analysis. The description of the mRNA libraries is: NCA, 35 days callus; NGD, 10 days germinating seedlings grown in dark; NGS, 3 days germinating seed; NIP, 90 days immature panicle; NMR, 60 days mature leaves (representing an average of four replicates; A, B, C and D); NME, 60 days meristematic tissue; NOS, ovary and mature stigma; NPO, mature pollen; NMR, 60 days mature roots (representing an average of two replicates; A and B); NST, 60 days stem; NYL, 14 days young leaves; Nyr, 14 days young roots; NSL, 14 days young leaves stressed in 250 mM NaCl for 24 h; NSR, 14 days young roots stressed in 250 mM NaCl for 24 h; NDL, 14 days young leaves stressed in drought for 5 days; NDR, 14 days young roots stressed in drought for 5 days; NCL, 14 days young leaves stressed at 4 °C for 24 h; NCR, 14 days young roots stressed at 4 °C for 24 h.

Acknowledgements

Financial support from the Department of Biotechnology, Government of India, and the University Grants Commission, New Delhi is gratefully acknowledged. M.J. acknowledges financial support from the Department of Biotechnology, Government of India under the Innovative Young Biotechnologists Award scheme.

References

45 Nardmann J & Werr W (2006) The shoot stem cell niche in angiosperms: expression patterns of WUS orthologues in rice and maize imply major modifica-

46 Windhovel A, Hein I, Dabrowa R & Stockhaus J (2001) Characterization of a novel class of plant home-

47 odomain proteins that bind to the C4 phosphoenolpy-


49 containing a conserved cysteine-rich region. Plant J 4, 137–150.


Genomic analyses of homeobox genes in rice

M. Jain et al.


Supplementary material

The following supplementary material is available online:

Fig. S1. Phylogenetic relationship of rice and Arabidopsis homeobox proteins based on the multiple sequence alignments of their homeodomain sequences.

Fig. S2. Expression profiles of homeobox genes present on duplicated chromosomal segments.

Fig. S3. Expression profiles of homeobox genes present in tandem.
Table S1. Homeobox genes in rice.
Table S2. Alternatively spliced homeobox genes in rice.
Table S3. Homeobox genes present on duplicated chromosomal segments of rice.
Table S4. Developmental stages of rice used for microarray analysis.
Table S5. Average log signal values of homeobox genes in three biological replicates of each tissue sample.
Table S6. Homeobox genes differentially expressed (more than two-fold with $P < 0.05$) under various stress conditions.

Table S7. Primer sequences used for real-time PCR analysis.

This material is available as part of the online article from http://www.blackwell-synergy.com

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.