the FEBS Journal

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

Keywords
abiotic stress; homeobox genes; microarray analysis; reproductive development; rice (*Oryza sativa*)

Correspondence
J. P. Khurana, Department of Plant Molecular Biology, University of Delhi South Campus, Benito Juarez Road, New Delhi 110021, India
Fax: +91 011 24115270
Tel: +91 011 24115126
E-mail: khuranaj@genomeindia.org

(Received 6 November 2007, revised 3 March 2008, accepted 31 March 2008)

doi:10.1111/j.1742-4658.2008.06424.x

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 α-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 *KNOTTED1*-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 1 *KNOX* genes, *SHOOT MERISTEMLESS* in *Arabidopsis* and *KNOTTED1* in maize, results in the development of embryos lacking SAM [15,17,18]. The *WUSCHEL*-like

Abbreviations
BLH, BEL1-like homeobox; DAP, days after pollination; EST, expressed sequence tag; FL-cDNA, full-length cDNA; HD, homeodomain; HD-ZIP, HD-leucine zipper; KNOX, KNOTTED1-like homeobox; MPSS, massively parallel signature sequencing; PHD, plant HD; SAM, shoot apical meristem; UDT1, UNDEVELOPED TAPETUM 1; WOX, WUSCHEL-like homeobox; WUS, WUSCHEL; ZF-HD, zinc finger-HD.

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

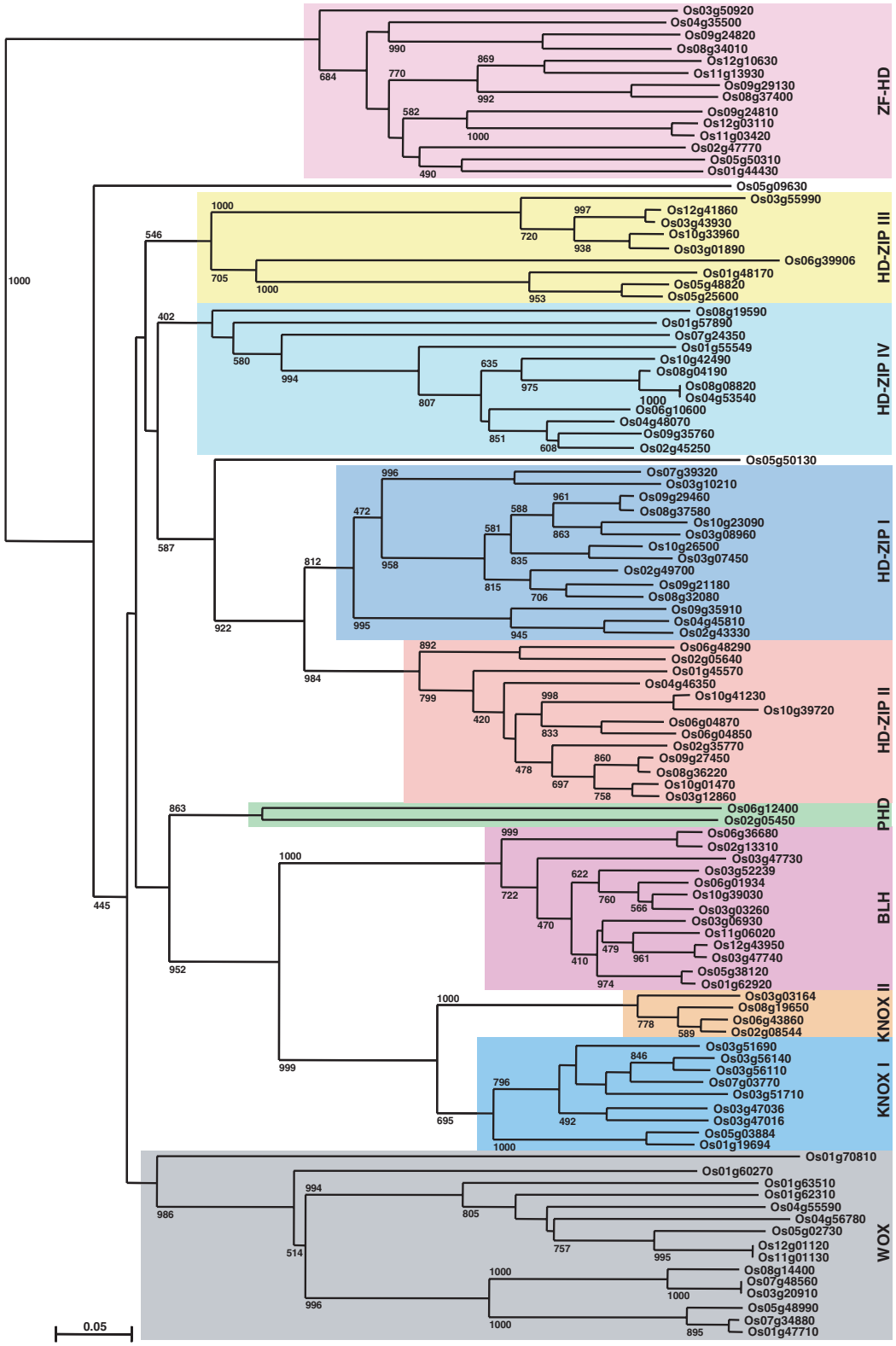


Fig. 1. Phylogenetic relationship among the rice homeobox proteins. The unrooted tree was generated from multiple sequence alignments of HD sequences. The sequences were aligned using CLUSTALX by the Neighbour-joining method. Ten distinct clades of homeobox proteins identified are represented by shading. The bootstrap values (> 40%) from 1000 replicates are indicated at each node.

Table 1. Classification of homeobox genes based on their domain composition and phylogenetic relationship.

Subfamily	Number of genes
HD-ZIP I	14
HD-ZIP II	13
HD-ZIP III	9
HD-ZIP IV	12
BLH	13
KNOX I	9
KNOX II	4
WOX	15
ZF-HD	14
PHD	2
Unclassified	2

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 Cys_4-His-Cys_3 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 subfamily proteins, could not be classified into any of the subfamilies.

Chromosomal localization and gene structure

The homeobox genes were localized on rice chromosome pseudomolecules based on their 5′ and 3′ coordinates available in the TIGR database, as represented diagrammatically in Fig. 2. The exact coordinates and orientation of each homeobox gene on the rice chromosome pseudomolecules is given in the supplementary Table S1. No substantial clustering of homeobox genes on rice chromosomes was observed. Although the 107 homeobox genes are scattered on all 12 rice chromosomes, their distribution is not uniform. The highest number (21; 19.6%) of genes is present on chromosome 3 followed by 12 genes on chromosome 1. Ten genes are present on chromosome 8; nine each on chromosomes 2, 5 and 6; eight each on chromosomes 9 and 10; seven on chromosome 4; five each on chromosomes 7 and 12, and only four genes on chromosome 11 (Fig. 2).

To study the gene structure of homeobox genes, their exon–intron organization was determined. The alignment of cDNA and corresponding genomic sequences revealed that the coding sequences of all

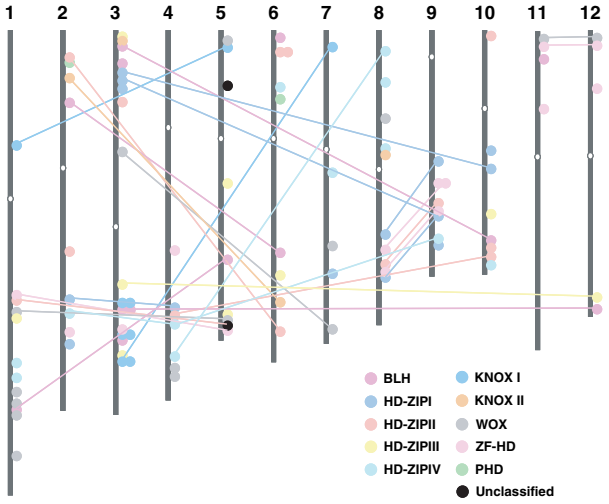


Fig. 2. Genomic distribution of homeobox genes on rice chromosomes. Homeobox genes classified in different subfamilies are shown in different colors. One circle represents one homeobox gene. White ovals on the chromosomes (vertical bar) indicate the position of centromeres. Chromosome numbers are indicated at the top of each bar. The homeobox genes present on duplicated chromosomal segments are connected by colored lines according to their subfamilies. The exact position (bp) and orientation of each homeobox gene on TIGR rice chromosome pseudomolecules (release 5) is given in the supplementary Table S1.

but 12 homeobox genes are indeed interrupted by one to 17 introns (supplementary Table S1). All the members of BLH subfamily except for *Os06g36680* 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

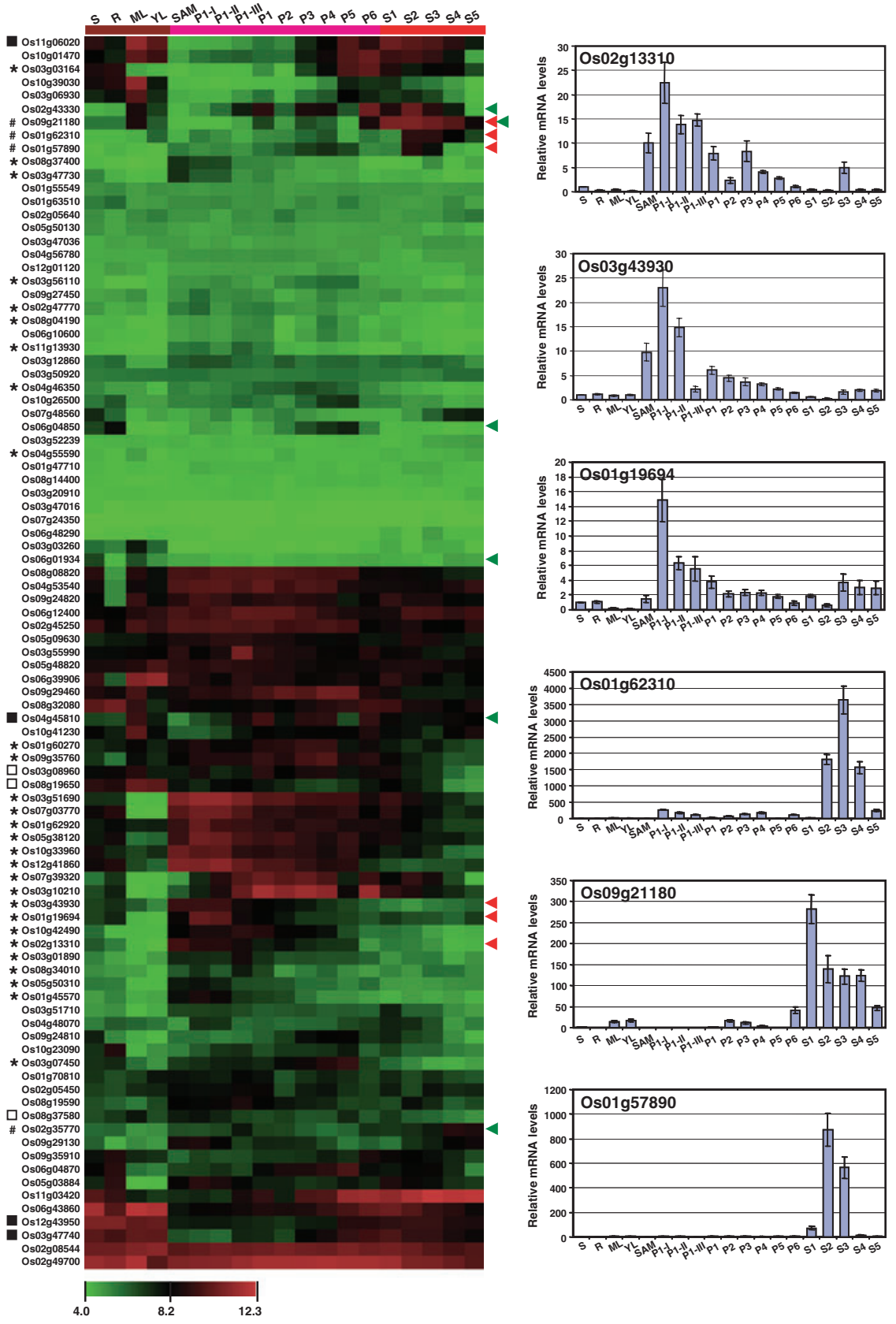
of 107 (86%) homeobox genes have at least one corresponding FL-cDNA and/or EST sequence (supplementary Table S1), indicating that most of these genes are expressed in rice. Fifty-nine homeobox genes have both FL-cDNA and EST evidence, whereas three and 30 genes have only FL-cDNA and EST evidence, respectively. 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

Fig. 3. Expression profiles of homeobox genes in various tissues/organs and developmental stages of rice. (A) Heatmap representing hierarchical clustering of average log signal values of all the homeobox genes in various tissues/organs and developmental stages (indicated at the top of each lane). The color scale (representing average log signal values) is shown at the bottom. The genes significantly up-regulated in at least one of the panicle and seed developmental stages are marked by asterisks and hash symbols, respectively, to the left. The genes significantly down-regulated in at least one of the panicle and seed developmental stages are marked by filled and open rectangles, respectively, to the left. The representative homeobox genes differentially expressed during various stages of development and under different abiotic stress conditions for which real-time PCR analysis was performed are indicated by red and green arrow heads, respectively, on the right. (B) Real-time PCR analysis to confirm the differential expression of representative homeobox genes during various stages of development. The mRNA 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.


S R ML YL SAM P1-I P1-II P1-III P1 P2 P3 P4 P5 P6 S1 S2 S3 S4 S5
Os11g06020
Os10g01470
Os03g03164
Os10g39030
Os03g06930
Os02g43330
Os09g21180
Os01g62310
Os01g57890
Os08g37400
Os03g47730
Os01g55549
Os01g63510
Os02g05640
Os05g50130
Os03g47036
Os04g56780
Os12g01120
Os03g56110
Os09g27450
Os02g47770
Os08g04190
Os06g10600
Os11g13930
Os03g12860
Os03g50920
Os04g46350
Os10g26500
Os07g48560
Os06g04850
Os03g52239
Os04g55590
Os01g47710
Os08g14400
Os03g20910
Os03g47016
Os07g24350
Os06g48290
Os03g03260
Os06g01934
Os08g08820
Os04g53540
Os09g24820
Os06g12400
Os02g45250
Os05g09630
Os03g55990
Os05g48820
Os06g39906
Os09g29460
Os08g32080
Os04g45810
Os10g41230
Os01g60270
Os09g35760
Os03g08960
Os08g19650
Os03g51690
Os07g03770
Os01g62920
Os05g38120
Os10g33960
Os12g41860
Os07g39320
Os03g10210
Os03g43930
Os01g19694
Os10g42490
Os02g13310
Os03g01890
Os08g34010
Os05g50310
Os01g45570
Os03g51710
Os04g48070
Os09g24810
Os10g23090
Os03g07450
Os01g70810
Os02g05450
Os08g19590
Os08g37580
Os02g35770
Os09g29130
Os09g35910
Os06g04870
Os05g03884
Os11g03420
Os06g43860
Os12g43950
Os03g47740
Os02g08544
Os02g49700
4.0
8.2
12.3
Os02g13310
Os03g43930
Os01g19694
Os01g62310
Os09g21180
Os01g57890
Relative mRNA levels

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/organs 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 then 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 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.

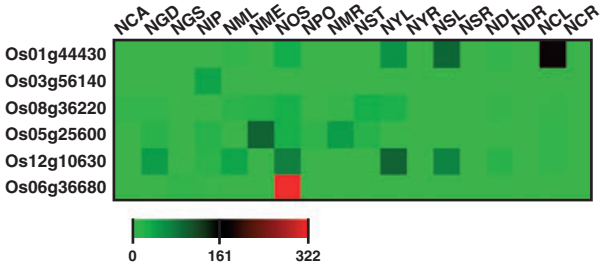


Fig. 4. Heatmap showing the expression profiles of six homeobox genes based on mRNA MPSS data. The number of significant signatures are shown as tpm. The color scale (representing tpm) is shown at the bottom. NCA, 35 days callus; NGD, 10 days germinating seedlings grown in dark; NGS, 3 days germinating seed; NIP, 90 days immature panicle; NML, 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.

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.

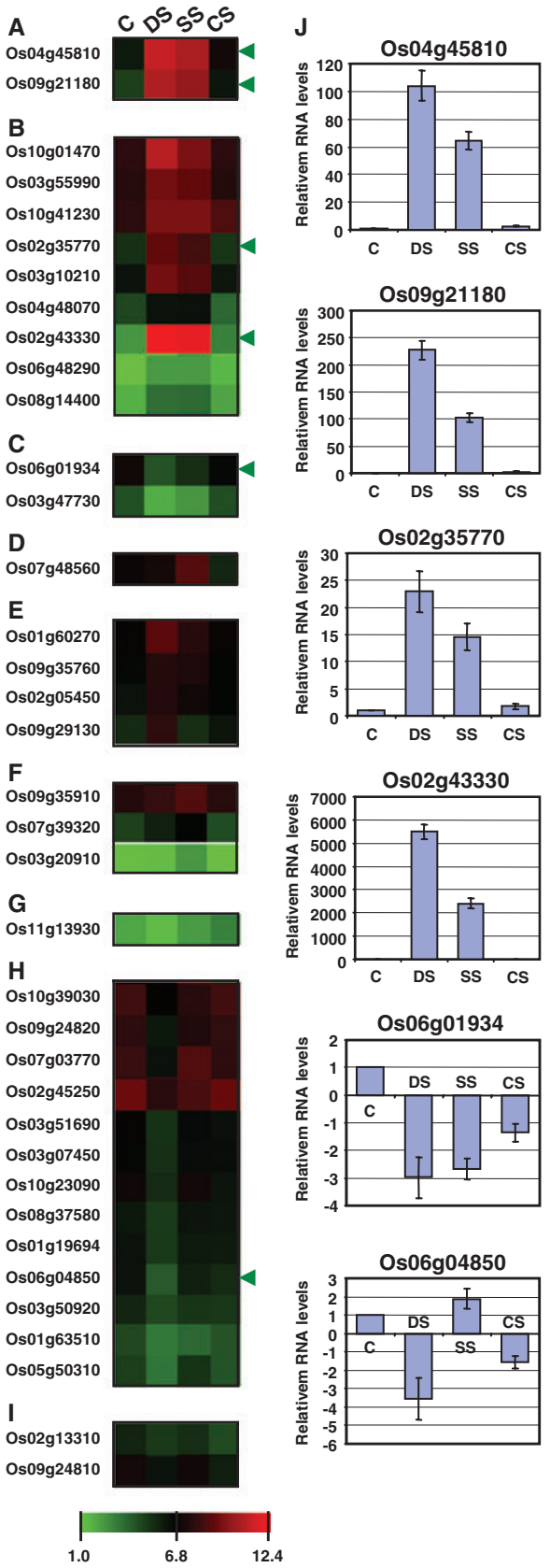
Expression profiles of duplicated homeobox genes

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



all the three stress conditions. Nine genes (i.e. *Os02g43330*, *Os03g10210*, *Os02g35770*, *Os06g48290*, *Os10g01470*, *Os10g41230*, *Os03g55990*, *Os04g48070* and *Os08g14400*) were up-regulated under desiccation and salt stress. However, four (*Os09g35760*, *Os05g50130*, *Os01g60270* and *Os09g29130*), three (*Os07g39320*, *Os09g35910* and *Os03g20910*), and one (*Os11g13930*) gene(s) were up-regulated specifically under desiccation, salt and cold stress, respectively. Two genes (*Os03g47730* and *Os06g01934*) were down-regulated under both desiccation and salt stress conditions. However, 13 (*Os03g07450*, *Os08g37580*, *Os10g23090*, *Os06g04850*, *Os02g45250*, *Os01g63510*, *Os10g39030*, *Os01g19694*, *Os03g51690*, *Os07g03770*, *Os03g50920*, *Os05g50310* and *Os09g24820*) and two (*Os02g13310* and *Os09g24810*) genes were down-regulated under desiccation and salt stress, respectively. Only one (*Os07g48560*) gene up-regulated by salt stress, was down-regulated by cold stress. The microarray data were validated by real-time PCR analysis of at least six genes exhibiting differential expression under various stress conditions (Fig. 5J).

Several HD-ZIP proteins (mainly members of HD-ZIP I subfamily) have been suggested to be dependent on abscisic acid signaling to act as transcriptional activator in various plant species [9,78–82]. The overexpression of rice *BIHD1/Os03g47740* gene in transgenic tobacco resulted in an elevated level of defence-related gene expression and enhanced sensitivity to salt and oxidative stress [12]. Our study

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, salt and cold stress (A), up-regulated by desiccation and salt (B), down-regulated by desiccation and salt (C), up-regulated by salt and down-regulated by cold (D), up-regulated by desiccation (E), up-regulated by salt (F), up-regulated by cold (G), down-regulated by desiccation (H) and down-regulated by cold (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 homeobox 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 *Os11g13930*) 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 homeobox proteins in regulatory networks

Homeobox 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 stages of panicle (up to 0.5 mm, SAM; 0.5–2 mm, P1-I; 2–5 mm, P1-II; 5–10 mm, P1-III; 0–3 cm, P1; 3–5 cm, P2; 5–10 cm, P3; 10–15 cm, P4; 15–22 cm, P5; 22–30 cm, P6) and seed (0–2 DAP, S1; 3–4 DAP, S2; 5–10 DAP, S3; 11–20 DAP, S4; 21–29 DAP, S5) development. Roots were harvested from 7-day-old seedlings grown in water. The desiccation (between folds of tissue paper), salt (200 mM NaCl solution) and cold (4 ± 1 °C) stress treatments were given to 7-day-old rice seedlings each for 3 h as described previously [55]. The control seedlings were kept in water for 3 h, at 28 ± 1 °C.

Database search

The Hidden Markov Model profile (build 2.3.2) of homeobox 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 by TIGR (release 5; http://www.tigr.org/tdb/e2k1/osa1) 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 homeobox proteins classified as zinc-finger homeobox 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 homeobox proteins for convenience. Domains in homeobox 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 homeobox 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 homeobox 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 search 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 abiotic 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 $\Delta\Delta C_T$ 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/organs, 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; NML, 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

1 Desplan C, Theis J & O'Farrell PH (1988) The sequence specificity of homeodomain-DNA interaction. *Cell* **54**, 1081–1090.
2 Otting G, Qian YQ, Billeter M, Muller M, Affolter M, Gehring WJ & Wuthrich K (1990) Protein–DNA contacts in the structure of a homeodomain–DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J* **9**, 3085–3092.
3 McGinnis W, Garber RL, Wirz J, Kuroiwa A & Gehring WJ (1984) A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. *Cell* **37**, 403–408.
4 Chan RL, Gago GM, Palena CM & Gonzalez DH (1998) Homeoboxes in plant development. *Biochim Biophys Acta* **1442**, 1–19.
5 Veraksa A, Del Campo M & McGinnis W (2000) Developmental patterning genes and their conserved functions: from model organisms to humans. *Mol Genet Metab* **69**, 85–100.
6 Himmelbach A, Hoffmann T, Leube M, Hohener B & Grill E (2002) Homeodomain protein ATHB6 is a target of the protein phosphatase ABI1 and regulates hormone responses in *Arabidopsis*. *EMBO J* **21**, 3029–3038.
7 Sawa S, Ohgishi M, Goda H, Higuchi K, Shimada Y, Yoshida S & Koshiba T (2002) The *HAT2* gene, a member of the HD-Zip gene family, isolated as an auxin inducible gene by DNA microarray screening, affects auxin response in *Arabidopsis*. *Plant J* **32**, 1011–1022.
8 Korfhage U, Trezzini GF, Meier I, Hahlbrock K & Somssich IE (1994) Plant homeodomain protein involved in transcriptional regulation of a pathogen defense-related gene. *Plant Cell* **6**, 695–708.
9 Lee YH & Chun JY (1998) A new homeodomain-leucine zipper gene from *Arabidopsis thaliana* induced by water stress and abscisic acid treatment. *Plant Mol Biol* **37**, 377–384.
10 Lee H, Xiong L, Gong Z, Ishitani M, Stevenson B & Zhu JK (2001) The *Arabidopsis HOS1* gene negatively regulates cold signal transduction and encodes a RING finger protein that displays cold-regulated nucleo–cytoplasmic partitioning. *Genes Dev* **15**, 912–924.
11 Zhu J, Shi H, Lee BH, Damsz B, Cheng S, Stirm V, Zhu JK, Hasegawa PM & Bressan RA (2004) An *Arabidopsis* homeodomain transcription factor gene, *HOS9*, mediates cold tolerance through a CBF-independent pathway. *Proc Natl Acad Sci USA* **101**, 9873–9878.
12 Luo H, Song F & Zheng Z (2005) Overexpression in transgenic tobacco reveals different roles for the rice homeodomain gene *OsBIHD1* in biotic and abiotic stress responses. *J Exp Bot* **56**, 2673–2682.
13 Lincoln C, Long J, Yamaguchi J, Serikawa K & Hake S (1994) A *knotted1*-like homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. *Plant Cell* **6**, 1859–1876.
14 Jackson D, Veit B & Hake S (1994) Expression of maize *KNOTTED1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. *Development* **120**, 405–413.
15 Long JA, Moan EI, Medford JI & Barton MK (1996) A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. *Nature* **379**, 66–69.
16 Hake S, Smith HM, Holtan H, Magnani E, Mele G & Ramirez J (2004) The role of *knox* genes in plant development. *Annu Rev Cell Dev Biol* **20**, 125–151.
17 Barton MK & Poethig RS (1993) Formation of the shoot apical meristem in *Arabidopsis thaliana*: an analysis of development in the wild-type and s*hoot meristemless* mutant. *Development* **119**, 823–831.
18 Vollbrecht E, Reiser L & Hake S (2000) Shoot meristem size is dependent on inbred background and presence of the maize homeobox gene, *knotted1*. *Development* **127**, 3161–3172.
19 Haecker A, Gross-Hardt R, Geiges B, Sarkar A, Breuninger H, Herrmann M & Laux T (2004) Expression dynamics of *WOX* genes mark cell fate decisions during

early embryonic patterning in *Arabidopsis thaliana*. *Development* **131**, 657–668.
20 Nardmann J, Zimmermann R, Durantini D, Kranz E & Werr W (2007) *WOX* gene phylogeny in Poaceae: a comparative approach addressing leaf and embryo development. *Mol Biol Evol* **24**, 2474–2484.
21 Mayer KF, Schoof H, Haecker A, Lenhard M, Jurgens G & Laux T (1998) Role of WUSCHEL in regulating stem cell fate in the *Arabidopsis* shoot meristem. *Cell* **95**, 805–815.
22 Kamiya N, Nagasaki H, Morikami A, Sato Y & Matsuoka M (2003) Isolation and characterization of a rice *WUSCHEL*-type homeobox gene that is specifically expressed in the central cells of a quiescent center in the root apical meristem. *Plant J* **35**, 429–441.
23 Reiser L, Modrusan Z, Margossian L, Samach A, Ohad N, Haughn GW & Fischer RL (1995) The *BELL1* gene encodes a homeodomain protein involved in pattern formation in the *Arabidopsis* ovule primordium. *Cell* **83**, 735–742.
24 Dong YH, Yao JL, Atkinson RG, Putterill JJ, Morris BA & Gardner RC (2000) *MDH1*: an apple homeobox gene belonging to the BEL1 family. *Plant Mol Biol* **42**, 623–633.
25 Byrne ME, Groover AT, Fontana JR & Martienssen RA (2003) Phyllotactic pattern and stem cell fate are determined by the *Arabidopsis* homeobox gene *BELL-RINGER*. *Development* **130**, 3941–3950.
26 Chen H, Rosin FM, Prat S & Hannapel DJ (2003) Interacting transcription factors from the three-amino acid loop extension superclass regulate tuber formation. *Plant Physiol* **132**, 1391–1404.
27 Smith HM & Hake S (2003) The interaction of two homeobox genes, *BREVIPEDICELLUS* and *PENNY-WISE*, regulates internode patterning in the *Arabidopsis* inflorescence. *Plant Cell* **15**, 1717–1727.
28 Talbert PB, Adler HT, Parks DW & Comai L (1995) The *REVOLUTA* gene is necessary for apical meristem development and for limiting cell divisions in the leaves and stems of *Arabidopsis thaliana*. *Development* **121**, 2723–2735.
29 McConnell JR & Barton MK (1998) Leaf polarity and meristem formation in *Arabidopsis*. *Development* **125**, 2935–2942.
30 McConnell JR, Emery J, Eshed Y, Bao N, Bowman J & Barton MK (2001) Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature* **411**, 709–713.
31 Otsuga D, DeGuzman B, Prigge MJ, Drews GN & Clark SE (2001) REVOLUTA regulates meristem initiation at lateral positions. *Plant J* **25**, 223–236.
32 Emery JF, Floyd SK, Alvarez J, Eshed Y, Hawker NP, Izhaki A, Baum SF & Bowman JL (2003) Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and *KANADI* genes. *Curr Biol* **13**, 1768–1774.
33 Prigge MJ, Otsuga D, Alonso JM, Ecker JR, Drews GN & Clark SE (2005) Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in *Arabidopsis* development. *Plant Cell* **17**, 61–76.
34 Tan QK & Irish VF (2006) The *Arabidopsis* zinc finger-homeodomain genes encode proteins with unique biochemical properties that are coordinately expressed during floral development. *Plant Physiol* **140**, 1095–1108.
35 Aso K, Kato M, Banks JA & Hasebe M (1999) Characterization of homeodomain-leucine zipper genes in the fern *Ceratopteris richardii* and the evolution of the homeodomain-leucine zipper gene family in vascular plants. *Mol Biol Evol* **16**, 544–552.
36 Sakakibara K, Nishiyama T, Kato M & Hasebe M (2001) Isolation of homeodomain-leucine zipper genes from the moss *Physcomitrella patens* and the evolution of homeodomain-leucine zipper genes in land plants. *Mol Biol Evol* **18**, 491–502.
37 Schena M & Davis RW (1994) Structure of homeobox-leucine zipper genes suggests a model for the evolution of gene families. *Proc Natl Acad Sci USA* **91**, 8393–8397.
38 Ponting CP & Aravind L (1999) START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. *Trends Biochem Sci* **24**, 130–132.
39 Schrick K, Nguyen D, Karlowski WM & Mayer KF (2004) START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. *Genome Biol* **5**, R41.
40 Mukherjee K & Burglin TR (2006) MEKHLA, a novel domain with similarity to PAS domains, is fused to plant homeodomain-leucine zipper III proteins. *Plant Physiol* **140**, 1142–1150.
41 Doerks T, Copley RR, Schultz J, Ponting CP & Bork P (2002) Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res* **12**, 47–56.
42 Burglin TR (1997) Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic Acids Res* **25**, 4173–4180.
43 Hofer J, Gourlay C, Michael A & Ellis TH (2001) Expression of a class 1 *knotted1*-like homeobox gene is down-regulated in pea compound leaf primordia. *Plant Mol Biol* **45**, 387–398.
44 Nagasaki H, Sakamoto T, Sato Y & Matsuoka M (2001) Functional analysis of the conserved domains of a rice KNOX homeodomain protein, OSH15. *Plant Cell* **13**, 2085–2098.
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-

tions in the course of mono- and dicot evolution. *Mol Biol Evol* **23**, 2492–2504.
46 Windhovel A, Hein I, Dabrowa R & Stockhaus J (2001) Characterization of a novel class of plant homeodomain proteins that bind to the C4 phosphoenolpyruvate carboxylase gene of *Flaveria trinervia*. *Plant Mol Biol* **45**, 201–214.
47 Schindler U, Beckmann H & Cashmore AR (1993) HAT3.1, a novel *Arabidopsis* homeodomain protein containing a conserved cysteine-rich region. *Plant J* **4**, 137–150.
48 Aasland R, Gibson TJ & Stewart AF (1995) The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem Sci* **20**, 56–59.
49 Lareau LF, Green RE, Bhatnagar RS & Brenner SE (2004) The evolving roles of alternative splicing. *Curr Opin Struct Biol* **14**, 273–282.
50 Campbell MA, Haas BJ, Hamilton JP, Mount SM & Buell CR (2006) Comprehensive analysis of alternative splicing in rice and comparative analyses with *Arabidopsis*. *BMC Genomics* **7**, 327.
51 Wang BB & Brendel V (2006) Genome-wide comparative analysis of alternative splicing in plants. *Proc Natl Acad Sci USA* **103**, 7175–7180.
52 Ito Y, Hirochika H & Kurata N (2002) Organ-specific alternative transcripts of KNOX family class 2 homeobox genes of rice. *Gene* **288**, 41–47.
53 Jain M, Kaur N, Garg R, Thakur JK, Tyagi AK & Khurana JP (2006) Structure and expression analysis of early auxin-responsive *Aux/IAA* gene family in rice (*Oryza sativa*). *Funct Integr Genomics* **6**, 47–59.
54 Jain M, Tyagi AK & Khurana JP (2006) Genome-wide analysis, evolutionary expansion, and expression of early auxin-responsive *SAUR* gene family in rice (*Oryza sativa*). *Genomics* **88**, 360–371.
55 Jain M, Nijhawan A, Arora R, Agarwal P, Ray S, Sharma P, Kapoor S, Tyagi AK & Khurana JP (2007) F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant Physiol* **143**, 1467–1483.
56 Zhang S, Chen C, Li L, Meng L, Singh J, Jiang N, Deng XW, He ZH & Lemaux PG (2005) Evolutionary expansion, gene structure, and expression of the rice wall-associated kinase gene family. *Plant Physiol* **139**, 1107–1124.
57 Itoh J, Nonomura K, Ikeda K, Yamaki S, Inukai Y, Yamagishi H, Kitano H & Nagato Y (2005) Rice plant development: from zygote to spikelet. *Plant Cell Physiol* **46**, 23–47.
58 Walia H, Wilson C, Condamine P, Liu X, Ismail AM, Zeng L, Wanamaker SI, Mandal J, Xu J, Cui X *et al.* (2005) Comparative transcriptional profiling of two contrasting rice genotypes under salinity stress during the vegetative growth stage. *Plant Physiol* **139**, 822–835.
59 Kumar R, Qiu J, Joshi T, Valliyodan B, Xu D & Nguyen HT (2007) Single feature polymorphism discovery in rice. *PLoS ONE* **2**, e284.
60 Walia H, Wilson C, Zeng L, Ismail AM, Condamine P & Close TJ (2007) Genome-wide transcriptional analysis of salinity stressed japonica and indica rice genotypes during panicle initiation stage. *Plant Mol Biol* **63**, 609–623.
61 Agarwal P, Arora R, Ray S, Singh AK, Singh VP, Takatsuji H, Kapoor S & Tyagi AK (2007) Genome-wide identification of C2H2 zinc-finger gene family in rice and their phylogeny and expression analysis. *Plant Mol Biol* **65**, 467–485.
62 Jain M, Tyagi AK & Khurana JP (2008) Differential gene expression of rice two-component signaling elements during reproductive development and regulation by abiotic stress. *Funct Integr Genomics* **8**, 175–180.
63 Nijhawan A, Jain M, Tyagi AK & Khurana JP (2008) A genomic survey and gene expression analysis of basic leucine zipper (bZIP) transcription factor family in rice. *Plant Physiol* **146**, 333–350.
64 Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, Luo S, McCurdy S, Foy M, Ewan M *et al.* (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* **18**, 630–634.
65 Nobuta K, Venu RC, Lu C, Belo A, Vemaraju K, Kulkarni K, Wang W, Pillay M, Green PJ, Wang GL *et al.* (2007) An expression atlas of rice mRNAs and small RNAs. *Nat Biotechnol* **25**, 473–477.
66 Kitano H, Tamura Y, Satoh H & Nagato Y (1993) Hierarchical regulation of organ differentiation during embryogenesis in rice. *Plant J* **3**, 607–610.
67 Hong SK, Aoki T, Kitano H, Satoh H & Nagato Y (1995) Phenotypic diversity of 188 rice embryo mutants. *Dev Genet* **16**, 298–310.
68 Hong SK, Kitano H, Satoh H & Nagato Y (1996) How is embryo size genetically regulated in rice? *Development* **122**, 2051–2058.
69 Sentoku N, Sato Y, Kurata N, Ito Y, Kitano H & Matsuoka M (1999) Regional expression of the rice *KN1*-type homeobox gene family during embryo, shoot, and flower development. *Plant Cell* **11**, 1651–1664.
70 Sato Y, Hong SK, Tagiri A, Kitano H, Yamamoto N, Nagato Y & Matsuoka M (1996) A rice homeobox gene, *OSH1*, is expressed before organ differentiation in a specific region during early embryogenesis. *Proc Natl Acad Sci USA* **93**, 8117–8122.
71 Satoh N, Hong SK, Nishimura A, Matsuoka M, Kitano H & Nagato Y (1999) Initiation of shoot apical meristem in rice: characterization of four *SHOOTLESS* genes. *Development* **126**, 3629–3636.

72 Ito M, Sato Y & Matsuoka M (2002) Involvement of homeobox genes in early body plan of monocot. *Int Rev Cytol* **218**, 1–35.
73 Ito M, Sentoku N, Nishimura A, Hong SK, Sato Y & Matsuoka M (2002) Position dependent expression of GL2-type homeobox gene, *Roc1*: significance for protoderm differentiation and radial pattern formation in early rice embryogenesis. *Plant J* **29**, 497–507.
74 Jung KH, Han MJ, Lee YS, Kim YW, Hwang I, Kim MJ, Kim YK, Nahm BH & An G (2005) Rice Undeveloped Tapetum1 is a major regulator of early tapetum development. *Plant Cell* **17**, 2705–2722.
75 Li M, Xu W, Yang W, Kong Z & Xue Y (2007) Genome-wide gene expression profiling reveals conserved and novel molecular functions of the stigma in rice. *Plant Physiol* **144**, 1797–1812.
76 Prince VE & Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet* **3**, 827–837.
77 He X & Zhang J (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**, 1157–1164.
78 Soderman E, Mattsson J & Engstrom P (1996) The *Arabidopsis* homeobox gene *ATHB-7* is induced by water deficit and by abscisic acid. *Plant J* **10**, 375–381.
79 Soderman E, Hjellstrom M, Fahleson J & Engstrom P (1999) The HD-Zip gene *ATHB6* in *Arabidopsis* is expressed in developing leaves, roots and carpels and up-regulated by water deficit conditions. *Plant Mol Biol* **40**, 1073–1083.
80 Deng X, Phillips J, Meijer AH, Salamini F & Bartels D (2002) Characterization of five novel dehydration-responsive homeodomain leucine zipper genes from the resurrection plant *Craterostigma plantagineum*. *Plant Mol Biol* **49**, 601–610.
81 Gago GM, Almoguera C, Jordano J, Gonzalez DH & Chan RL (2002) *Hahb-4*, a homeobox-leucine zipper gene potentially involved in abscisic acid-dependent responses to water stress in sunflower. *Plant Cell Environ* **25**, 633–640.
82 Yu S-W, Zhang L-D, Zuo K-J, Tang D-Q, Sun X-F & Tang K-X (2005) *Brassica napus* L. homeodomain leucine zipper gene *BnHB6* responds to abiotic and biotic stresses. *J Integr Plant Biol* **47**, 1236–1248.
83 Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, Mauch F, Luan S, Zou G, Whitham SA *et al.* (2002) Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. *Plant Cell* **14**, 559–574.
84 Wang X (2002) Phospholipase D in hormonal and stress signaling. *Curr Opin Plant Biol* **5**, 408–414.
85 Potocky M, Elias M, Profotova B, Novotna Z, Valentova O & Zarsky V (2003) Phosphatidic acid produced by phospholipase D is required for tobacco pollen tube growth. *Planta* **217**, 122–130.
86 Zonia L & Munnik T (2004) Osmotically induced cell swelling versus cell shrinking elicits specific changes in phospholipid signals in tobacco pollen tubes. *Plant Physiol* **134**, 813–823.
87 Cooper B, Clarke JD, Budworth P, Kreps J, Hutchison D, Park S, Guimil S, Dunn M, Luginbuhl P, Ellero C *et al.* (2003) A network of rice genes associated with stress response and seed development. *Proc Natl Acad Sci USA* **100**, 4945–4950.
88 Lan L, Li M, Lai Y, Xu W, Kong Z, Ying K, Han B & Xue Y (2005) Microarray analysis reveals similarities and variations in genetic programs controlling pollination/fertilization and stress responses in rice (*Oryza sativa* L.). *Plant Mol Biol* **59**, 151–164.
89 Tardif G, Kane NA, Adam H, Labrie L, Major G, Gulick P, Sarhan F & Laliberte JF (2007) Interaction network of proteins associated with abiotic stress response and development in wheat. *Plant Mol Biol* **63**, 703–718.
90 Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C *et al.* (2005) The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol* **3**, e38.
91 Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F & Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.
92 Saitou N & Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.
93 Perriere G & Gouy M (1996) WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie* **78**, 364–369.
94 Jain M, Nijhawan A, Tyagi AK & Khurana JP (2006) Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochem Biophys Res Commun* **345**, 646–651.

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.