Evolutionary Bioinformatics

ORIGINAL RESEARCH

AP2/ERF Transcription Factor in Rice: Genome-Wide Canvas and Syntenic Relationships between Monocots and Eudicots

Muhammad Rashid, He Guangyuan, Yang Guangxiao, Javeed Hussain and Yan Xu

China-UK HUST-RRes Genetic Engineering and Genomics Joint Laboratory, International Science and Technology Cooperation Base (Genetic Engineering) of Chinese Ministry of Science and Technology, The key laboratory of Molecular Biophysics of Chinese Ministry of Education, College of Life Science and Technology, Huazhong University of Science and Technology (HUST), Luoyu Road 1037, Wuhan 430074, China.
Corresponding author email: mrashid_niab@yahoo.com

Abstract: The transcription factor family intimately regulates gene expression in response to hormones, biotic and abiotic factors, symbiotic interactions, cell differentiation, and stress signalling pathways in plants. In this study, 170 AP2/ERF family genes are identified by phylogenetic analysis of the rice genome (*Oryza sativa* l. japonica) and they are divided into a total of 11 groups, including four major groups (AP2, ERF, DREB, and RAV), 10 subgroups, and two soloists. Gene structure analysis revealed that, at position-6, the amino acid threonine (Thr-6) is conserved in the double domain AP2 proteins compared to the amino acid arginine (Arg-6), which is preserved in the single domain of ERF proteins. In addition, the histidine (His) amino acid is found in both domains of the double domain AP2 protein, which is missing in single domain ERF proteins. Motif analysis indicates that most of the conserved motifs, apart from the AP2/ERF domain, are exclusively distributed among the specific clades in the phylogenetic tree and regulate plausible functions. Expression analysis reveals a widespread distribution of the rice AP2/ERF family genes within plant tissues. In the vegetative organs, the transcripts of these genes are found most abundant in the roots followed by the leaf and stem; whereas, in reproductive tissues, the gene expression of this family is observed high in the embryo and lemma. From chromosomal localization, it appears that repetition and tandem-duplication may contribute to the evolution of new genes in the rice genome. In this study, interspecies comparisons between rice and wheat reveal 34 rice loci and unveil the extent of collinearity between the two genomes. It was subsequently ascertained that chromosome-9 has more orthologous loci for CRT/DRE genes whereas chromosome-2 exhibits orthologs for ERF subfamily members. Maximum conserved synteny is found in chromosome-3 for AP2 double domain subfamily genes. Macrosynteny between rice and Arabidopsis, a distant, related genome, uncovered 11 homologs/orthologs loci in both genomes. The distribution of AP2/ERF family gene paralogs in Arabidopsis was most frequent in chromosome-1 followed by chromosome-5. In Arabidopsis, ERF subfamily gene orthologs are found on chromosome-1, chromosome-3, and chromosome-5, whereas DRE subfamily genes are found on chromosome-2 and chromosome-5. Orthologs for RAV and AP2 with double domains in Arabidopsis are located on chromosome-1 and chromosome-3, respectively. In conclusion, the data generated in this survey will be useful for conducting genomic research to determine the precise role of the AP2/ERF gene during stress responses with the ultimate goal of improving crops.

Keywords: AP2/ERF, CBF/DREB, RAV, phylogenetic analysis, rice, wheat, Arabidopsis

Evolutionary Bioinformatics 2012:8 321–355

doi: 10.4137/EBO.S9369

This article is available from http://www.la-press.com.

Introduction

Plants encounter various natural affronts, from pests, diseases, and ecological changes, including those affecting temperature and water. These factors greatly impinge on a plant's endurance, growth, and production. In order to stay alive and grow in diverse environments, plants have developed a complex signalling network at the molecular, cellular, and system levels. Gene co-expression at the transcription level is the foremost regulatory mechanism in biological processes. Transcription factors (TFs) control the majority of multiple stress response genes in a synchronized manner and are characterized as attractive targets for application in molecular plant biology. Among different TFs, the ethylene responsive transcription factors (ERF) family plays a vital role in plant growth and enables plants capable of fighting ambiance changes.[1–4] Therefore, it is important to understand the function of these genes in order to ameliorate crop yield and allow them to grow in sundry environmental conditions.

In the plant Kingdom, APETALA2/ethylene-responsive element-binding protein (AP2/EREBP) is a large family of TFs that contain AP2, RAV, and ERF family genes. This super family is characterized by the conserved AP2/ERF DNA binding domain, comprising of 60–70 amino acid residues.[5,6] This domain, first reported in the homeotic gene AP2 of Arabidopsis, is involved in flower development.[7] The ERF domain was first described as a conserved motif in four DNA-binding proteins (EREBP1, 2, 3, and 4; presently renamed ERF1, 2, 3, and 4, respectively) in tobacco plants. This domain binds to a GCC box, which is a DNA sequence involved in the ethylene-responsive transcription of genes.[8] In tomatoes, three proteins (Pti-4, Pti-5, and Pti-6) are reported to interact with disease resistance proteins, determined using yeast two-hybrid assays,[9] and each protein has a domain homologous to the AP2/ERF domain. Based on the number of AP2/ERF domains and their gene function, the AP2/ERF gene family is divided into four subfamilies: AP2, RAV, dehydration-responsive element-binding protein (DREB), and ERF.[10]

The AP2 family genes possess two repeats of the AP2/ERF domain and the ERF family proteins contain a single AP2/ERF domain whereas RAV (related to VP1/ABI3) family proteins have an additional B3 DNA binding domain. The AP2 subfamily is further subdivided into two monophyletic groups: AP2 and AINTEGUMENTA (ANT).[11] Based on ERF domain binding to DNA sequences, the ERF family is further split up into two subfamilies: ERF and CBF/DREB.[6] Proteins encoded by ERF subfamily genes bind to the core motif AGCCGCC;[8,12] whereas, CBF/DREB subfamily genes containing C-repeats recognize the *cis*-acting element, A/GCCGAC.[13]

ERF subfamily genes have conserved alanine (Ala) at position-14 and aspartic acid (Asp) at position-19, while CBF/DREB subfamily genes have valanine (Val) at position-14 and glutamine (Glu) at position-19.[14] The three-dimensional structure of the AP2/ERF domain from AtERF1 is portrayed using hetero-nuclear multidimensional Nuclear Magnetic Resonance and is comprised of three anti-parallel β-sheets and an α-helix.[15] Moreover, tryptophan (Trp) and arginine (Arg) residues present in the β-sheets come into contact with DNA during transcription.[14]

The AP2/ERF family is identified in the Arabidopsis,[6,10] grapevine,[16] poplar,[17] and rice genomes with 145, 132, 200, and 163 genes, respectively.[18] Using genetic and molecular approaches, AP2/ERF proteins were determined to play an imperative role in the transcriptional regulation of diverse biological processes. These processes are related to plant growth, development, and response to environmental stimuli.[2,19] The genes in this family participate in different pathways in response to hormones and biotic and abiotic stresses, such a jasmonate,[20,21] abscisic acid (ABA),[22,23] drought,[24,25] salinity,[26,27] cold,[28,29] disease,[21,30,31] and flooding stress.[32] The AP2 family contributes to the regulation of developmental processes, such as flower development,[7,33] early floral meristem identity,[34] crown and lateral roots,[35,36] and somatic embryogenesis.[37] In addition, the involution of RAV family members is reported in the AP2/ERF family in response to ethylene,[38] brassinosteroid,[39,40] and biotic and abiotic stresses.[30,41,42]

Determination of the phylogenetic relationship is a crucial step in elucidating the evolution of crop species. In the past, AP2/ERF genes were considered plant specific but,[5] recently, this domain was reported in non-plants for the first time in the ciliate *Tetrahymena thermophila*,[43] cyanobacterium *Trichodesmium erythraeum*, and viruses Enterobacteria phage Rb49 and Bacteriophage Felix 01.[44] The non-plant AP2-domain containing proteins have HNH domains edging the

AP2/ERF domain. These are known as HNH endonuclease encoded by the His-Ans-His sequence motif.[45] It is hypothesized that AP2/ERF TFs may have originated by horizontal transfer of the (HNH)-AP2 endonuclease from bacteria/viruses into plants via transpositions and homing processes.[44]

The above cited work was done on different plants whereas the present work is reported on rice. Belonging to the gramene family, rice is the world's most important staple crop that ensures food security. In addition to its agricultural utility and small genome size (394Mb), rice provides an excellent model plant system to study plant-pathogen interactions, hormonal pathways, and resistance to environmental stresses. Some AP2/ERF family genes have been cloned from rice, but the function of the majority of them remains unclear. The completion of high-quality sequencing of the rice genome has furnished an excellent opportunity for elaborate gene analysis.

In this study, 170 OsAP2/ERF rice genes are obtained from different database searches and are classified into their respective clades according to their homology with known genes. The functional analysis of each transcriptional factor belonging to the AP2/ERF family is carried out keeping in mind their functional redundancy. In addition, phylogenetic analysis of AP2/ERF genes, complete alignment of each subfamily gene, and distribution of the conserved motifs outside the AP2/ERF domain exhibiting the specificity for gene function is performed. Moreover, to generate a clear picture of AP2/ERF genes, the position of each gene on the chromosomes is determined and gene expression profile canvasing is executed to determine tissue specificity. Comparative mapping between a monocot (wheat) and a eudicot (Arabidopsis) is performed to identify homologs/orthologs among the genomes. The data brought forth in this canvas will help in the selection of appropriate genes for further functional characterization and understanding of precise regulatory checkpoints operating during development and stress responses in crop plants.

Materials and Methods

Identification of AP2/ERF genes in rice genome

First, rice AP2/ERF genes were identified in the genome of *Oryza sativa* l. japonica cultivar Nipponbare using ESTs and cDNA sequences. The data were downloaded from various public repositories, including The National Centre for Biotechnology Information (NCBI),[46] The Database of Rice Transcription Factors (DRTF),[47] The MSU Rice Genome Annotation Project Database,[48] Knowledge-based Oryza Molecular Biological Encyclopedia (KOME),[49] and Plant Genome Database (PlantGDB).[50] Next, all retrieved sequences were subjected to the BLAT online tool available on the RAP-DB website to find homologous sequences in the rice genome.[51] The sequences showing more than 80% coverage areas were expanded approximately 2000 bp on both sides of the hit to find the open reading frame (ORF) using the GENSCAN online tool.[52] Data assembly was performed using a DNA Assembly Sequence Programme CAP3.[53] In addition, Simple Modular Architecture Research Tool (SMART) is used to confirm the presence of the AP2/ERF domain in the resulting sequences.[54]

Phylogenetic and MEME motif analysis

The AP2/ERF domain comprising protein sequences obtained from various sources are aligned using ClustalX 2.0[54] and redundant entries are removed.[55] A combined un-rooted neighbor-joining (NJ) tree was generated in MEGA 4.0[56] with the following default parameters:[56] poisson correction, pairwise deletion, and bootstrap (5000 replicates). Conserved motifs in rice AP2/ERF protein sequences were identified using a motif based sequence analysis tool, MEME Suite version 4.7.0,[57] with following parameters: optimum width 6–200 amino acids, any number of repetitions of a motif, and maximum number of motifs set at 25. The BLAST search for the resulting motifs in NCBI and MS-Homology databases was carried out to determine their significance.

Intron/exon size distribution of AP2/ERF family genes

Intron positions in genes are ascertained through the identification of gaps in alignment of full-length cDNA transcripts with genomic sequences using the online tool Gene Structure Display Server (GSDS).[58] Concisely, for a single full-length cDNA aligned against a conterminous stretch of genomic sequence, exons are proximal blocks of homologous sequence between full-length cDNA and genomic sequences, whereas introns are gaps between exons consisting entirely of genomic sequence.

The general distribution of introns within each coding DNA sequence (CDS) is analyzed by the distribution of exon sizes. The mean exon size for a full length cDNA containing n_i introns (regardless of pattern of distribution) is calculated as [Length of coding DNA sequence/(n_i + 1)].[59] The total length of a gene is calculated by adding the length of all of the exons of each gene. Moreover, exon size and total gene length of each gene are plotted graphically to clarify the range and size of AP2/ERF family genes.

Gene expression analysis

To further investigate and confirm AP2/ERF gene expression, the rice expression profile database (RiceXPro), which is a repository of gene expression data derived from microarray experiments encompassing the entire life cycle of the rice plant from germination, seedling, tillering, stem elongation, panicle initiation, booting, heading, flowering, and ripening stage,[60] was used. This tool generates a heat map of normalized signal intensity values for each plant tissue for each gene and provides a quantitative measure of the transcript of particular genes.

In Silico identification of AP2/ERF gene family orthologs/homologs in rice, wheat and Arabidopsis genomes

Comparative genetic mapping among closely related (wheat) and distantly related (Arabidopsis) genomes divulges a high level of synteny in gene content and gene function. The Arabidopsis AP2/ERF genes were retrieved from the supplementary file provided by a genome wide survey of ERF family genes in Arabidopsis and rice,[10] and were confirmed using a database of genetic and molecular biology for the model plant *Arabidopsis thaliana* (TAIR).[61] Regarding wheat (*Triticum aestivum*), AP2/ERF sequences were retrieved from NCBI,[46] Plant Genome Database (PlantGDB),[50] and Chinese-Spring draft genome assembly or raw sequence reads using BLAST available online at CerealsDB, University of Bristol. Wheat and Arabidopsis AP2/ERF consensus gene sequences were used to BLAST against rice annotated genes and rice full length (FL) cDNAs with BLASTx and BLASTn, respectively, to identify orthologs/homologs between the genomes.

Results and Discussions

To ascertain the AP2/ERF family genes in the rice genome, sequences obtained from various resources as mentioned in the materials and methods were assembled using the CAP3 online tool. After confirmation of the AP2/ERF domain using the SMART online tool and removing the redundant sequences by alignment using ClustalX 2.0, it was determined that 170 sequences encode the AP2/ERF domain (Supplementary File 2). It was also discovered that 143 genes encode a single AP2/ERF domain. Twenty-three genes are anticipated to encode proteins comprising double AP2/ERF domains whereas four genes are prognosticated to encode a single AP2/ERF domain together with one B3 domain. Based on these results, the former and latter genes are appointed to the AP2 with double domain and the RAV families, respectively.

These results are an improvement on the AP2/ERF super family described by Nakano and his co-workers as they do not use these two subfamilies in their investigation under genome-wide analysis of the ERF gene family in Arabidopsis and rice (Supplementary File 2).[10] Regarding the importance of the AP2 subfamily genes with double domains, the ANT gene plays a critical role in the development of gynoecium marginal tissues (eg, stigma, style, and septa), and in the fusion of carpels and medial ridges that lead to ovule primordial.[62] BABY BOOM (BBM) promotes cell proliferation, differentiation, and morphogenesis, especially during embryogenesis.[63] WRINKLED (WRl) genes are involved in the regulation of gene expression by stress factors and by components of stress signal transduction pathways. These are involved in the activation of a subset of sugar-responsive genes and the control of carbon flow from sucrose import to oil accumulation in developing seeds.[64] The PLETHORA (PLT) gene acts as a transcriptional activator and functions as a master regulator of basal/root fate; it is essential for root quiescent centre (QC) and columella specification, stem cell activity, as well as for the establishment of the stem cell niche during embryogenesis. PLT also modulates root polar auxin transport by regulating the distribution of PIN genes and plays an essential role in specifying pattern and polarity in damaged roots.[65] Besides, genes with floral homeotic proteins promote

early floral meristem identity and are subsequently required for the transition of an inflorescence meristem into a floral meristem; they play a central role in the specification of floral identity, particularly for the normal development of sepals and petals in the wild-type flower.[66]

Other subfamily genes with one B3 domain function as negative regulators of plant growth and development. These genes are expressed in all tissues: roots, rosette leaves, cauline leaves, inflorescence stems, flowers, and siliques. They have highest expression in roots and rosette leaves and low expression in flowers.[39]

Phylogenetic relationships between AP2/ERF family transcription factors in rice

To dissect the phylogenetic relationships between rice AP2/ERF family proteins, multiple alignment analyses were executed using the amino acid sequences of the AP2/ERF domains. The alignment predicts that Gly-4, Val-5, Gly-11, Ile-17, Arg-26, Trp-28, Leu-29, Gly-30, Ala-38, Ala-39, Asp-43, Gly-51, Asn-58, and Phe-59 are completely conserved in all AP2/ERF single domain family members in rice (Fig. 2A). In addition, more than 95% of ERF members have conserved Arg-3, Arg-6, Arg-8, Glu-16, Arg-18, Ala-35, Ala-45, Ala-46, and Ala-55 amino acid residues. The AP2/ERF proteins with a single domain in the AP2/ERF subfamily contain conserved Alanine (Ala) at position-14 and Aspartic acid (Asp) at position-19, whereas the CBF/DREB subfamily genes have Valine (Val) at position-14 and Glutamine acid (Glu) at position-19 (Fig. 1).[14] Instead, few gene domains are without Glu-19 and exhibit significant similarity with the rest of the domain although genes in the RAV family with a B3 domain have Gly-14 instead of (Val/Ala-14) (Fig. 2A).

The multiple alignments of the amino acid sequences of AP2/ERF double domain proteins were determined. The amino acid residues were as follows: Gly-4, Val-5, Gly-12, Glu-15, His-17, Asp-20, Gly-38, Ala-46, Ala-47, Arg-48, Ala-49, Asp-51, Ala-53, Ala-54, Lys-56, Gly-59, Asn-67, and Phe-68. These were found to be conserved in the first domain of AP2/ERF double domain proteins. It is of note that at position-6, Thr-6 is conserved in the double domain AP2 genes compared to the Arg-6 conserved in single domain ERF proteins. Moreover, the His-amino acid is found in both the domains of double domain AP2/ERF proteins that are missing in the single domain AP2/ERF proteins. In the second domain of AP2/ERF double domain proteins, the alignment indicated that the amino acid residues Tyr-2, Arg-3, Gly-4, Val-5, Gly-12, Arg-13, Trp-14, Ala-16, Arg-17, Gly-19, Ala-44, Tyr-45, Asp-46, Ala-48, Ala-49, Ile-50, Gly-54, Thr-59, Asn-60, and Phe-61 are highly conserved (Fig. 2B).

Based on these observations, a phylogenetic tree of 170 AP2/ERF genes was constructed using bootstrap analysis (5000 replicates) based on the multiple sequence alignments of their protein domains (Fig. 3). The bootstrap values for the nodes in this phylogenetic tree are not high in every case, similar to the results of the phylogenetic analysis done on Arabidopsis ERF proteins.[10] The phylogram is alienated into a total of 11 groups with four main

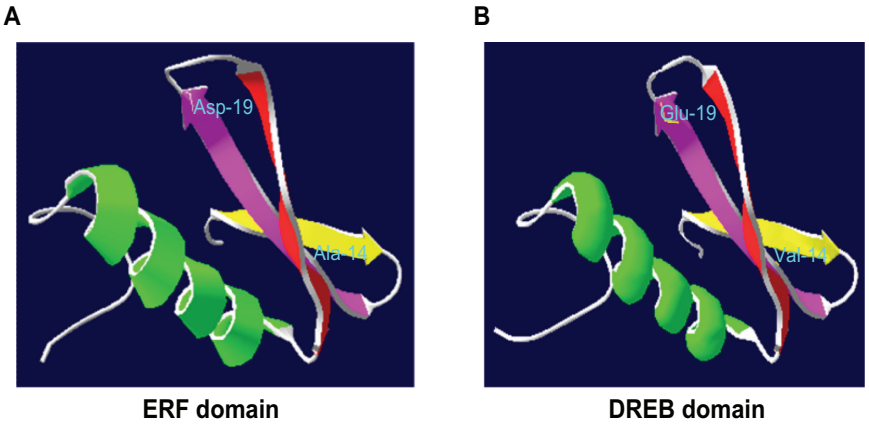


Figure 1. The conserved amino acid residues found in ERF (**A**) and CBF/DREB (**B**).

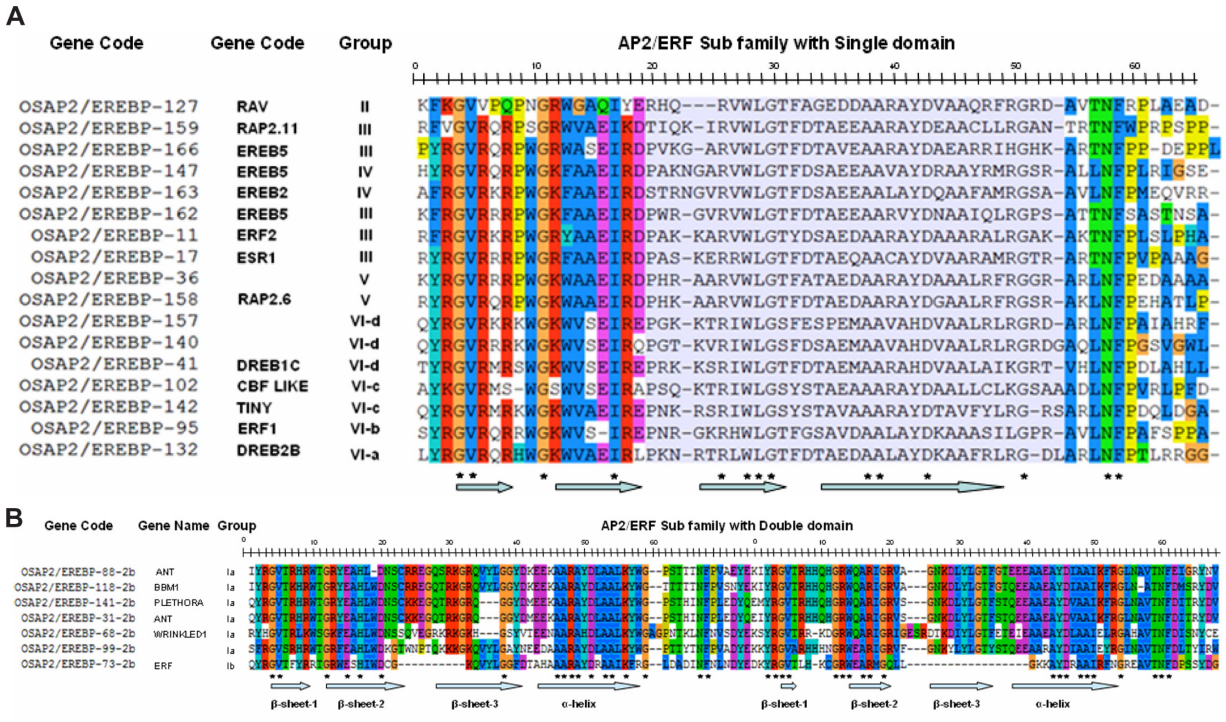


Figure 2. Alignment of the AP2/ERF domains from Rice (*O. sativa*) cv. japonica ERF proteins. (**A** and **B**) displayed the alignment of single AP2/ERF and double AP2/ERF double domain proteins respectively.
Notes: The arrows represent the predicted helix and sheets. Asterisks exhibited the conserved amino acid residues among the AP2/ERF domain proteins.

groups, namely group-I to group-IV. The members of the ERF gene family are classified into 10 groups in Arabidopsis.[10] Group-I is distinguished into three sub groups: I-a, I-b, and I-c. The proteins categorized in group-I are characterized with a double domain. These proteins play important roles in Arabidopsis, rice, *Brassica napus* (Rape), and *Medicago truncatula* for vegetative to embryonic growth, somatic embryogenesis, seed development, lateral roots formation, petal cell identity, floral meristem identity

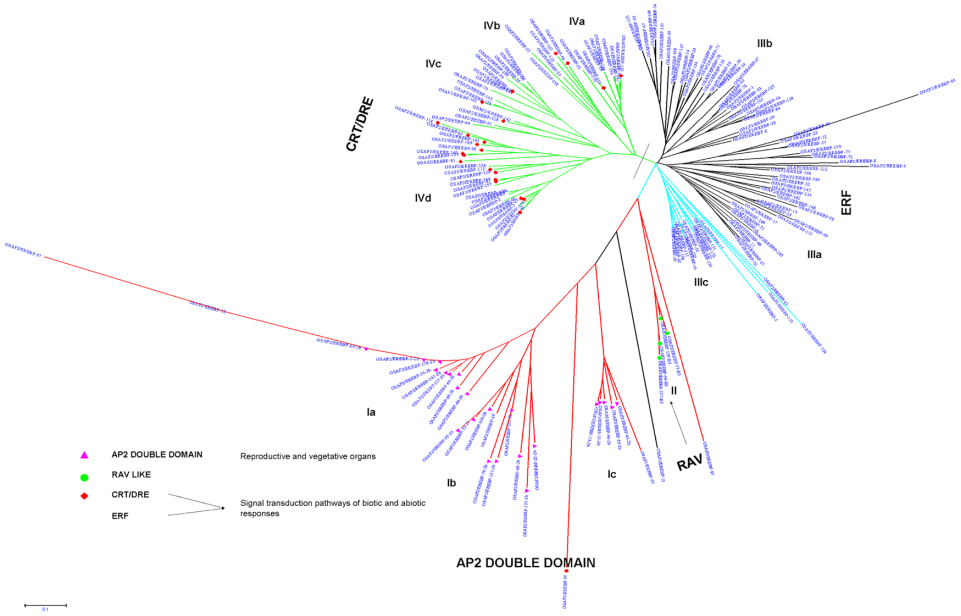


Figure 3. An unrooted phylogenetic tree of Rice AP2/ERF proteins is constructed using NJ method.
Notes: The line marked distinction between ERF and CRT/DREB protein sub families. Scale bar represents 0.01 amino acid substitution per site.

and transition to flowering time.[33,36,37,63,64,67–69] The genes in this group are expressed more in the flower, leaves, and stem as opposed to the roots.[70]

AP2/ERF genes in Group-II consist of one additional B3 domain and are distinguished as RAV proteins. The function of these proteins in Arabidopsis are as negative regulators of plant growth and development and in the tomato play a pivotal modulator role in defense mediated pathways.[39,30] The remaining genes fall in into group-III and group-IV and have a single AP2/ERF domain. The genes in group-III are part of the ERF subfamily that are used in crop plants to defend against biotic resistance, such as pathogens and diseases. The role of these genes have been studied extensively in Arabidopsis, rice, tomato, and tobacco plants with respect to hormone response (jasmonic acid, salicylic acid, absicic acid, and ethylene),[10] pathogen/disease's resistance,[4,12,31,71–77] and insect damage (Supplementary Table S2).[78] Some family members play crucial roles in transcription inhibition due to the presence of an EAR motif.[10,79] It is reported that ERF is over expressed in shoots due to low temperature, but expressed little in cultured cells and roots.[3,80,81]

Finally, the genes in Group-IV are characterized into CBF, DREB, and TINY proteins, which are used for abiotic factors, such as cold, salt, and drought resistance. These genes are characterized in Arabidopsis, rice, maize, rape seed, sunflower, tomato, and tobacco plants for their response to abiotic factors (Supplementary Table S2).[6,25,82–93] The inclusion of these genes in distinct groups based on their protein similarities need to be characterized according to different stress regimes, which may help in correlating function within their phylogenetic placement.

Distribution of conserved motifs

Complete AP2/ERF protein sequences of rice were analyzed for the presence of conserved motifs using the MEME Suite version 4.7.0. Overall, 19 conserved motifs were identified and named 1–19. The consensus sequences of these motifs are provided in Supplementary File 1. Motifs 1, 2, 3, 4, 5, 6, and 9, correspond to the AP2/ERF domain region. The remaining 18 motifs were found to characterize specific clades in the phylogentic tree. The distribution of these conserved motifs in proteins of relevant clades in the phylograms is laid out in Figure 4.

Conserved motifs outside the AP2/ERF domain

In general, transcription factors comprise functionally important conserved motifs outside the DNA binding domain, which are related to transcriptional activity, nuclear localization, and protein-protein interactions.[10] Such functional amino acid sequences motifs are often conserved among members of a subgroup in large families of transcription factors in plants; proteins with these motifs in their subgroups are likely to have similar functions.[94]

An investigation of the conserved motifs in the proteins of each clade in the AP2/ERF family of rice was accomplished using multiple alignment analysis with ClustalX 2.0.[55] The conserved motifs found in the OsAP2/ERF family are summarized in Supplementary Table S1. Most of the motifs are selectively distributed among the specific clades. For instance, in Arabidopsis, the conserved motifs CM-16 and CM-17 found in the C-terminal region of group-II with consensus sequences KGVLLNFED[A/G][A/E/D]GKVW[R/K]FRYSYWNSSQSYV and [AW]AR[ED][HP]LF[DE]K[AT]VTPSDVGKLNRLV[IV]PKQ[HQ]AE[KR]HFP[LF], respectively, function as a transcriptional repressor for plant growth and development (Fig. 5 and Supplementary Table S1).[39,49] CM-18, MCGGAI[LI]AD[LF]IP, and CM-19, D[DE]D[FW]EA[AD]F[ER]EF[EDL][DSV][DR][DS][DGH]D[DS][DE]D[ED], are found in the N-terminal region of members of group-III. They have two small blocks of conserved amino acid's sequences, eg, MCCGAI and DFEA, which play a role in ethylene transcriptional activation (Fig. 6, Supplementary Table S1).[95,96] CM-11 in the members of group-IVc and IVd have, at the C-terminal, the consensus sequence [LR]PR[PA]A[ST]A[SA]PKD[VI][QR]AAAA[LA]AAA[AM]ARPPP. The alignment of the members predicted four small blocks of conserved amino acid residues: LPR[P/A], D[I/V]QAA, D[I/V]R[A/L/R]AA, and [L/R]AAA. These are the essential signatures in Arabidopsis for CBL-interacting serine/threonine-protein kinase-12, Ethylene-responsive transcription factor ERF037, dehydration responsive element binding proteins-1C, and proteins-1G and Auxin response factor-19, respectively (Fig. 7, Supplementary Table S1).[97–100] The [L/R]AAA motif is a homologue of auxin response factors (ARFs). It binds specifically to the DNA sequence 5′-TGTCTC-3′ and is

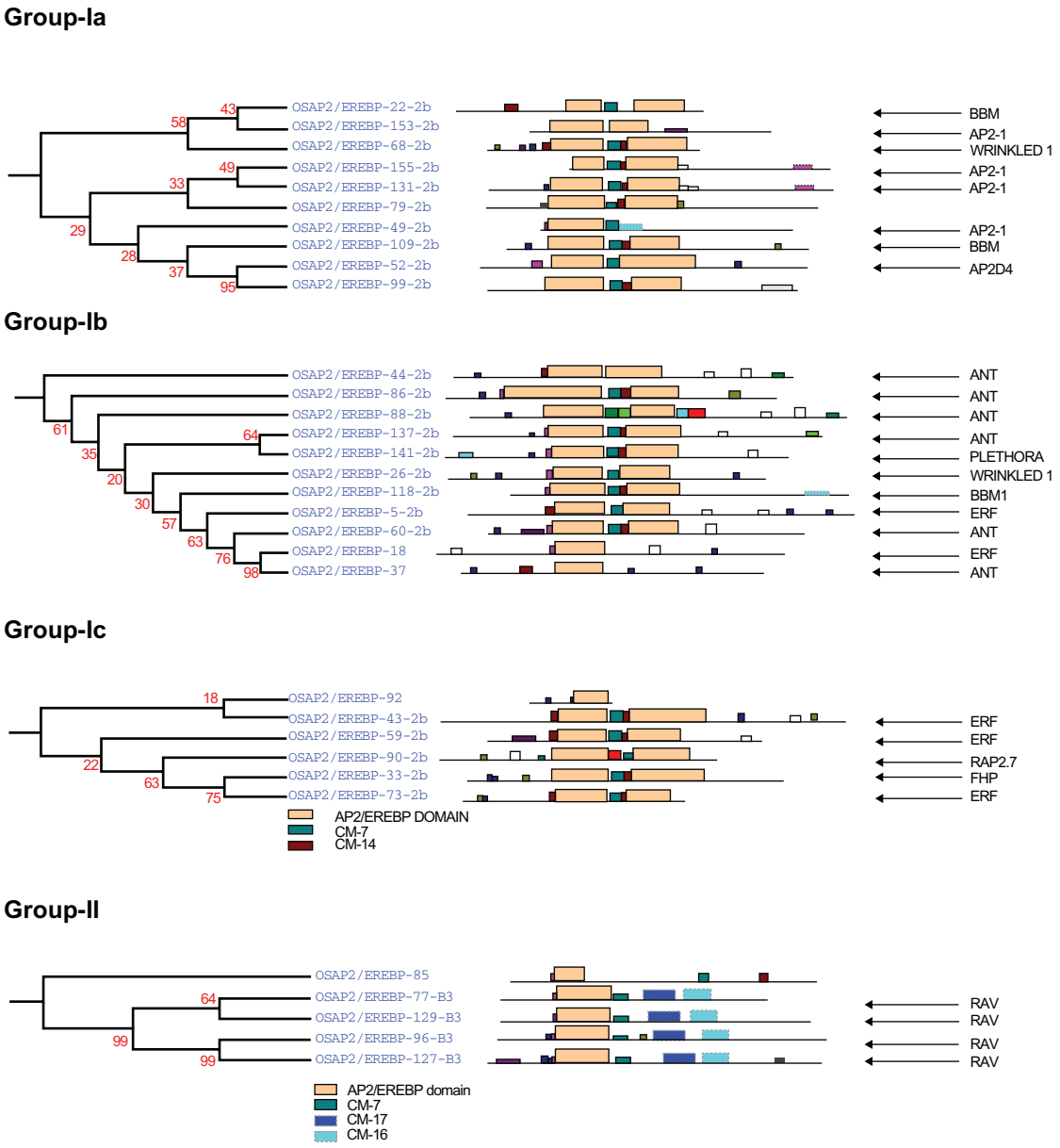


Figure 4A. The phylogenetic relationship in the AP2/ERF family genes. Group-Ia, Ib and group-Ic represent the AP2/ERF double domain proteins with their conserved motifs. Group-II exhibited the RAV family genes with one B3 domain. Their conserved motifs determined by the MEME online tools are given the Supplementary File 1.

discovered in the auxin-responsive promoter elements (AuxREs). This motif acts as a transcriptional activator and is involved in ethylene responses and it regulates lateral root formation through direct regulation of LBD16 and/or LBD29.[101] The conserved amino acid sequence LPR[P/A] detected in the CBL proteins (serine-threonine protein kinases) binds to the regulatory NAF domain of the CIPK protein and activates the kinase in a calcium-dependent pathway.[97] The [H/V/A/D/E/R/Q]LNFP motif is found to be involved in disease resistance.[102]

Four small preserved amino acid residues were found after alignment in the CM-10 motif with the consensus sequence PEMEKLDFTEAPWDESETF-HLRKYPSWEIDWDSILS (Fig. 8, Supplementary Table S1). Among them, a unique small conserved sequence of amino acids, LDF[S/T]E at the C-terminal region, plays a key role in disease resistance.[103] The APWDE motif was found to be involved in transcriptional regulation and acts as a histone acetyltransferase, which mediates the acetylation of histone H3 and H4 of target loci. It is also implied

in an auxin-independent regulation of shoot, branching, and flowering time. In addition, it is expressed in leaves, buds, flowers, stems, and over-methylated genomic DNA.[104] The KYPS motif is involved in the DNA methylation process, which plays an important role in genome management and in regulating gene expression during development.[105] A distinctive small motif, EIDWD, was also found to be involved in Arabidopsis response to ethylene.[106] It is clear that most of the conserved motifs identified in this study have plausible features in their amino acid compositions as shown in Supplementary Table S1.

Characteristics of each group in the rice AP2/ERF gene family

The characteristics of each group of the rice AP2/ERF family are described below. For reference, current information regarding the functions of the genes in the AP2/ERF family is summarized in Supplementary Table S2.

Group-I

Group-I is divided into three subgroups: Ia, Ib, and Ic (Fig. 4, Group-Ia, Ib and Ic). All genes in this group have double AP2/ERF domains except OsAP2/EREBP-049, OsAP2/EREBP-018, OsAP2/EREBP-037, and OsAP2/EREBP-092. These genes fall in group-I due to their domain similarities. At this time, the functions of the OsAP2/EREBP-079-2b, OsAP2/EREBP-092-2b, and OsAP2/EREBP-099-2b genes are unknown. All genes have the conserved motifs CM-7 and CM-14 that separates the two domains. The loci with these two conserved motifs are provided in Supplementary File 2 and are reported as ANT, BBM, and AP2 like (AIL) proteins of Arabidopsis.[10,37,67] Genes grouped in group-Ia, Ib, and group-Ic are involved in seed (embryonic growth, seed development, seed germination) and flower traits (petal cell identity, flowering time, and floral meristem identity).[33,63,64,67,69,]

Group-II

Group-II proteins share three conserved motifs: CM-7, CM-16 and CM-17 in the C-terminal region (Fig. 4, Group-II), which are contiguous with the AP2/ERF domain. In addition to a single AP2/ERF domain, these genes have one B3 domain. The RAV family proteins act as negative regulators of plant growth and development in Arabidopsis and take part in defence mediated responses in tomatoes against different stresses.[39,40,30]

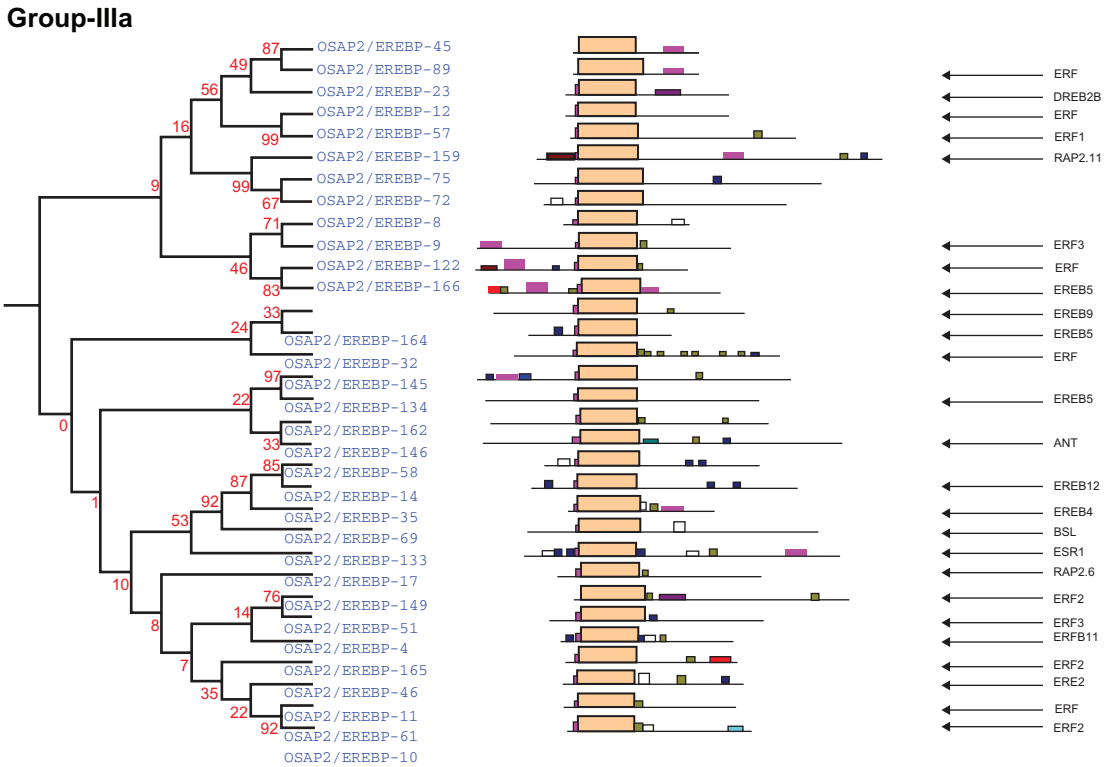


Figure 4B. (*Continued*)

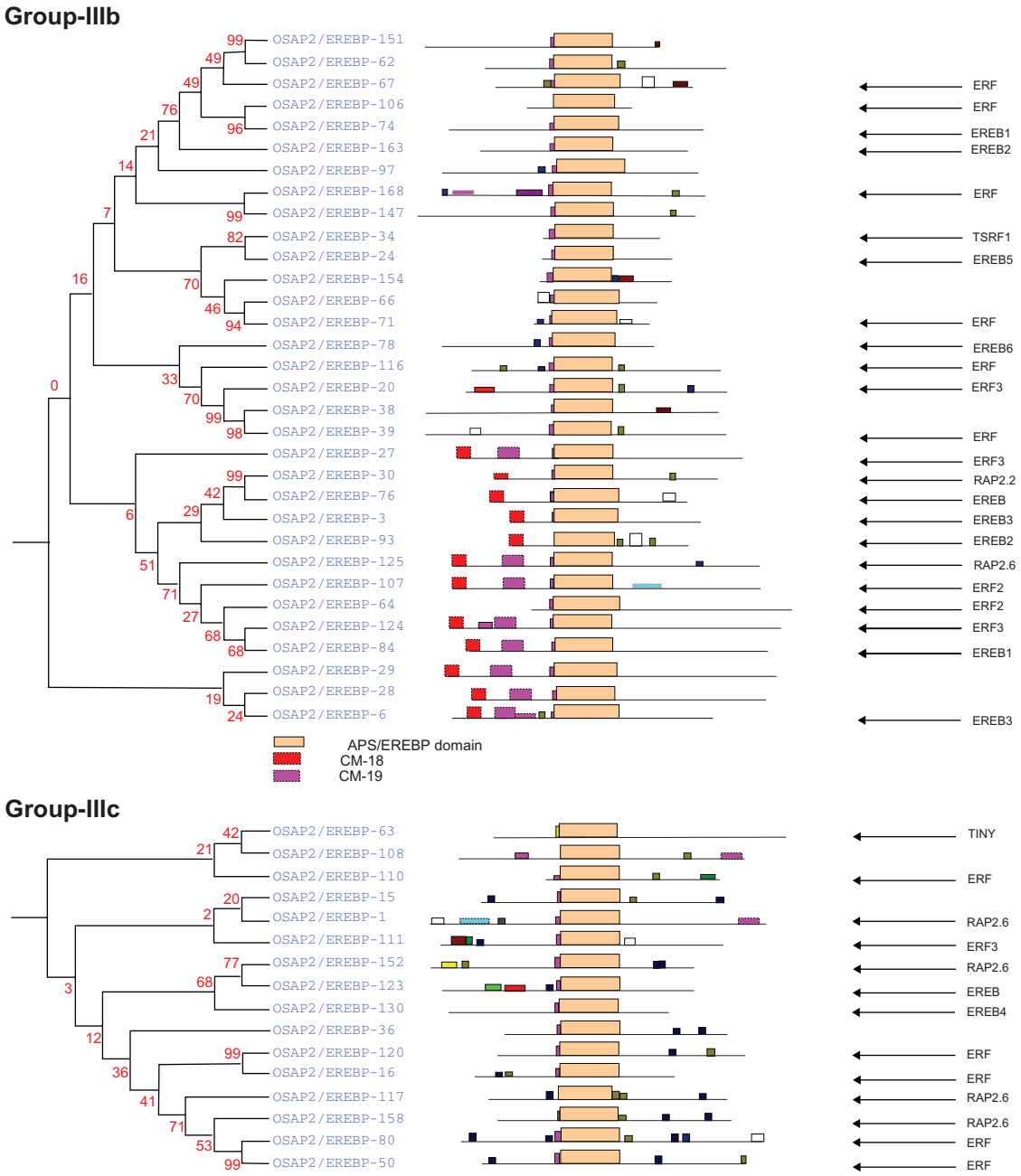


Figure 4B. The phylogenetic relationship among the AP2/ERF family genes in rice. Group-IIIa, IIIb and IIIc corresponds to the AP2/ERF single domain proteins. Group-IIIa, IIIb and IIIc falls in the ERF. Their conserved motifs determined by the MEME online tools are given the Supplementary File 1.

Group-IIIa

Group-IIIa consists of 32 genes (Fig. 4, group-IIIa). The AP2/ERF genes with the generic names OsAP2/EREBP-010, OsAP2/EREBP-011, OsAP2/EREBP-046, OsAP2/EREBP-051, and OsAP2/EREBP-057 interact as a transcriptional activator and are involved in disease resistance pathways in Arabidopsis, rice, tobacco, and tomato plants.[4,31,71,72] Their domains bind to the GCC-box pathogenesis-related promoter element. They are involved in the regulation of gene expression induced by stress factors mediated by ethylene that seem dependent on a protein kinase cascade. The genes, coded as OsAP2/EREBP-009 and OsAP2/EREBP-012, are transcriptional inhibitors and may regulate the expression of other genes in a co-expression manner. The N-terminal regions of the genes OsAP2/EREBP-009, OsAP2/EREBP-122, OsAP2/EREBP-134, and OsAP2/EREBP-166 consist of the conserved motif CM-19. Furthermore, defence-related phytohormones, such as ethylene,

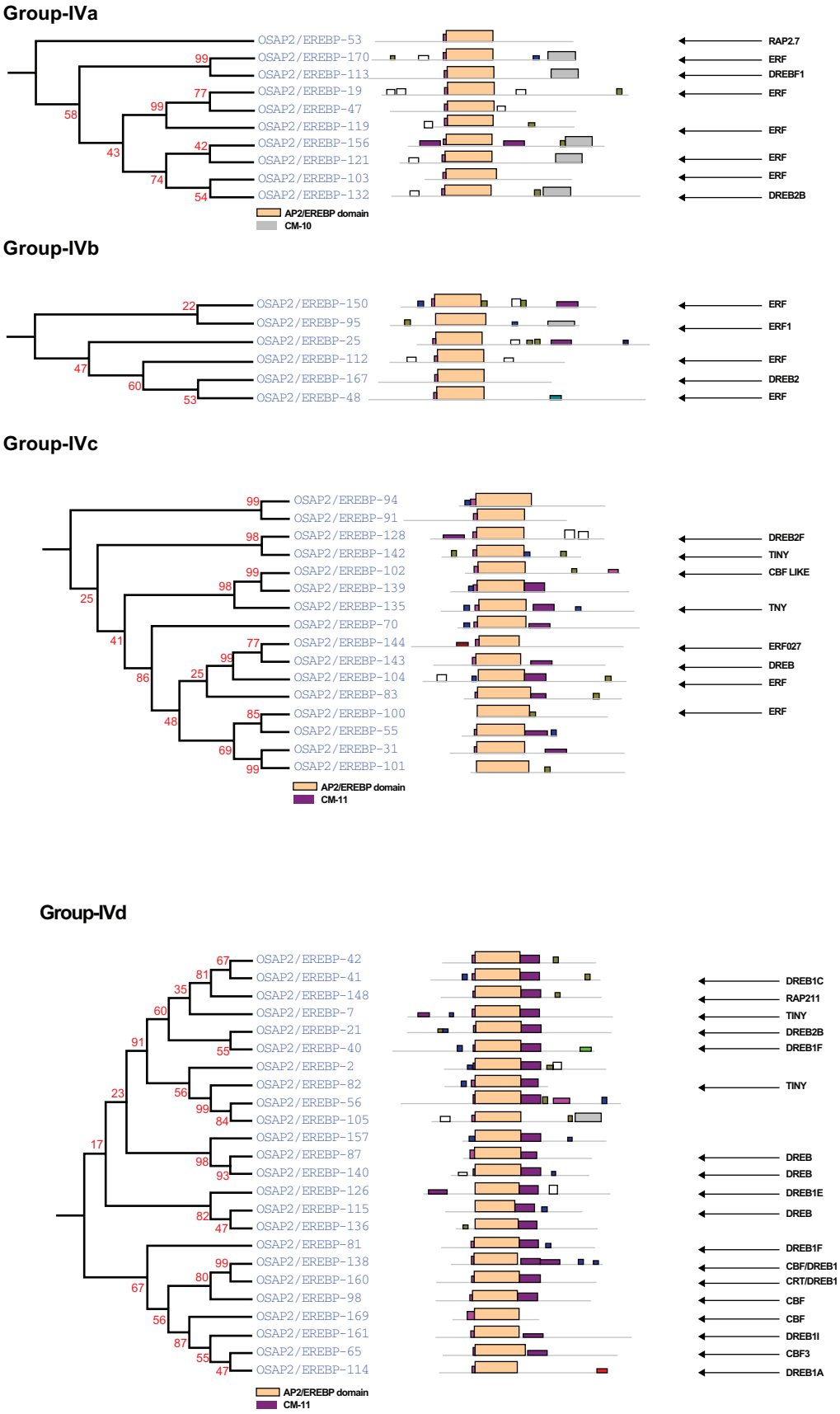


Figure 4C. The phylogenetic relationship among the AP2/ERF family genes in rice. Group-IV symbolises the AP2/ERF single domain proteins. Group-IVa, IVb, IVc and IVd are predicted to relate CBF/DREB genes. Their conserved motifs determined by the MEME online tools are given the Supplementary File 1.

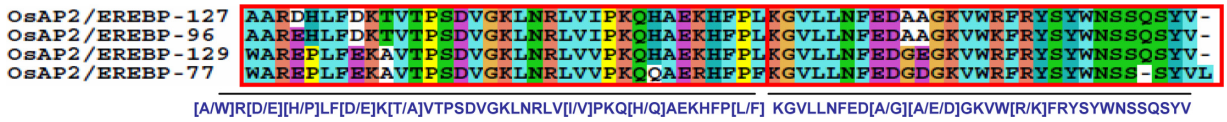


Figure 5. RAV like motif sequences conserved in the C-terminal region of group-II in rice.
Notes: The conserved sequences are under lined calculated by MEME programme. Consensus amino acid sequences are given below the line.

jasmonate, and salicylic acid differentially induce the expression of this group of genes.[10]

Group-IIIb

This subgroup of AP2/ERF genes is also related to biotic stresses, such as disease and pathogen resistance and responds to jasmonite and ethylene signal transduction pathways. Out of the 32 genes, the function of nine of these genes has not been studied. This group shares two motifs, CM-18 and CM-19, in the N-terminal region (Fig. 4, group-IIIb). The functional characterization of the two small conserved proteins blocks, MCGGAI and DFEA, are studied by comparing with the Arabidopsis ERF genes.[95,96] The OsAP2/EREBP-074 and OsAP2/EREBP-084 (EREBP1) genes in wheat and cotton repress GCC box-mediated transcription by improving pathogen and abiotic stress tolerance in transgenic plants.[73,74] Recently, genes OsAP2/EREBP-020, OsAP2/EREBP-027, and OsAP2/EREBP-124 (ERF3), are up regulated by the feeding of the rice striped stem borer on rice.[78] In addition, OsAP2/EREBP-003, OsAP2/EREBP-006, OsAP2/EREBP-024, OsAP2/EREBP-078, OsAP2/EREBP-093, and OsAP2/EREBP-163 display important roles in cultured cells and roots.[80,81]

Group-IIIc

Group-IIIc consists of 16 genes (Fig. 4. group-IIIc). The function of three genes, OsAP2/EREBP-015, OsAP2/EREBP-036, and OsAP2/EREBP-108 has not previously been studied. Except for the gene OsAP2/EREBP-063 (TINY), which plays an important part in abiotic stresses, the other genes have a crucial function in biotic stresses (insect resistance, responsive to stress hormones like jasmonic acid, salicylic acid, abscisic acid, and ethylene).[10]

Group-IVa

Except for the genes OsAP2/EREBP-019, OsAP2/EREBP-053, OsAP2/EREBP-103, OsAP2/EREBP-047, and OsAP2/EREBP-119, genes in this group contain the CM-10 motif in the C-terminal region of the AP2/ERF domain (Fig. 4 group-IVa). Close inspection of this conserved motif divides it into four blocks with the following protein sequences: APWDE, LDF[S/T]E, EIDWD, and KYPS. In Arabidopsis, these four conserved protein blocks play key roles in disease resistance, auxin responsiveness, and regulation of genes during expression and transcription activation (Supplementary Table S1).[103–106] The overexpression of the genes OsAP2/EREBP-132 and OsAP2/EREBP-113 in Arabidopsis are characterized for high salt resistance.[6]

Group-IVb

This small group consists of only six genes. Among them, the gene OsAP2/EREBP-167 is involved in water deficit tolerance in rice and sunflowers.[25,82] Moreover, the OsAP2/EREBP-095 gene is a disease resistance gene in Arabidopsis.[31] The gene OsAP2/EREBP-025 requires characterization (Fig. 4 group-IVb).

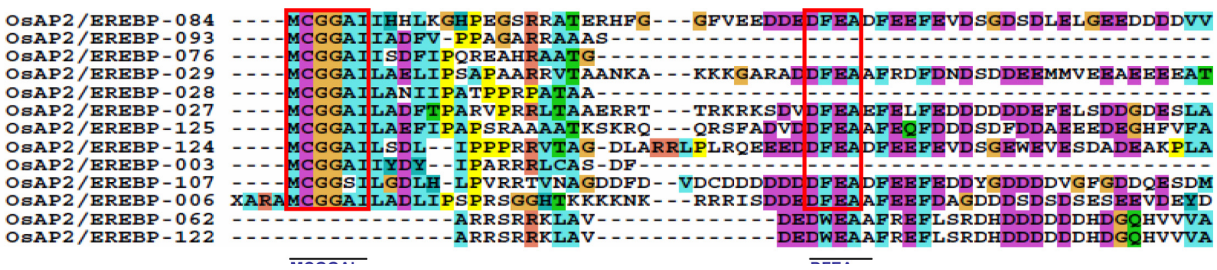


Figure 6. The conserved motif amino acid sequences in group-III are identified in the N-terminal region.
Note: The consensus sequences are given under the line.

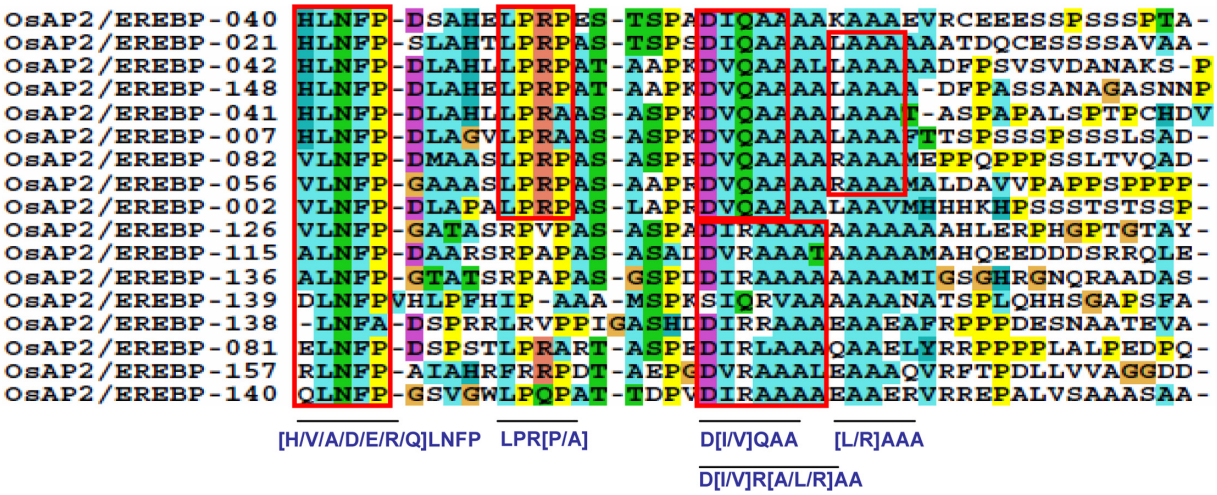


Figure 7. The conserved motif amino acid sequences in group-IVc and group-IVd respectively are identified in the C-terminal region.
Note: The consensus sequences are given under the line.

Group-IVc

The roles of the genes in subgroup-IVc have been extensively studied. These genes play essential functions in response to high salinity and cold-stress.[6] Although the functions of the OsAP2/EREBP-031, OsAP2/EREBP-055, OsAP2/EREBP-070, OsAP2/EREBP-083,OsAP2/EREBP-91,OsAP2/EREBP-94, OsAP2/EREBP-101, and OsAP2/EREBP-139 genes are unknown, these genes may play a role as transcriptional activators in gene expression in response to abiotic stresses due to their placement in the phylogram (Fig. 4 group-IVc). In Arabidopsis and corn, TINY (OsAP2/EREBP-142 and OsAP2/EREBP-135) is a homolog of the group-IVc and responds to cold stress.[92,93]

Group-IVd

The proteins in group-IVd possess the CM-11 motif that is homologous to proteins of group-IVc. Possession of these motifs and their phylogenetic relationships indicate a strong similarity between groups-IVd and IVc (Fig. 4 group-IVd). The function of genes with the generic names, OsAP2/EREBP-002, OsAP2/EREBP-042, OsAP2/EREBP-056, OsAP2/EREBP-105, OsAP2/EREBP-136, and OsAP2/EREBP-157, need to be investigated. The CM-11 motif has a conserved [H/V/A/D/E/R/Q]LNFP amino acid residue sequence whose function is studied in rice and other homologous plant species.[102] Another conserved protein sequence, LPR[P/A] in motif CM-11, also plays an important role in serine-threonine protein kinase mechanisms to activate the kinase in a calcium-dependent manner.[97] The functional characterization of these genes for low temperature, salt, dehydration, drought resistance, and osmotic tolerance has been investigated in tomato, rice, Arabidopsis, and corn plants.

Gene length with respect to intron/exon size

Usually, the intron/exon position in the CDS furnishes clues on evolutionary relationships of genomes.[10] To explore the rice genome to find more AP2/ERF family genes, the full cDNA information is computed

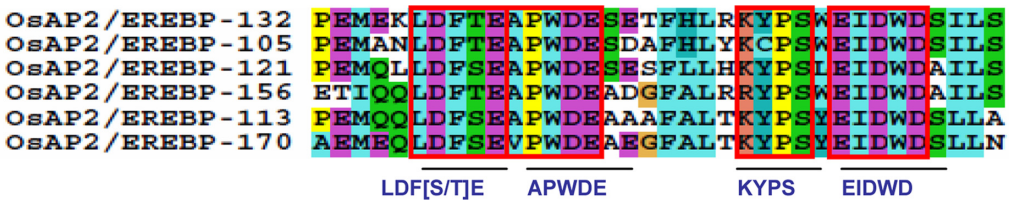


Figure 8. The conserved motif amino acid sequences in group-IVa are identified in the C-terminal region.
Note: The consensus sequences are given under the line.

using the coverage of 2000 bp region on both sides of the hit to avoid any possible gene losses. It was found that gene length ranged from 781.1 to 1679.7 bp with an average gene length of 1230.4 ± 449.3 bp. These findings indicate that the average gene length is within the 2000 bp limit used in our survey to find these genes. Therefore, it is evident that every exon covered the ESTs and this makes it possible to map the rice genome and this 2000 bp limit ensures that there is no loss of genes (Fig. 9A). For furthering probing, we analyzed the size distribution of exons in our cDNA region by comparing the CDS with genomic sequences in each gene. It was found that the length of exons ranges between 138.4 to 274.79 bp with the average size of exons being 205.3 bp with a standard deviation of 68.96, which is also within the limit and indicates that all the genes are in the expanding range of 2000 bp on both side of the hit (Fig. 9B, Supplementary Table S3).

Expression profile analysis of the AP2/ERF gene family

Rice, being a staple food crop, suffers yield losses as a result of environmental cues and pathogens during plant development. It is evident that AP2/ERF family genes play important roles in plant development in response to biotic and abiotic stresses and thus, they present as ideal candidates to investigate the molecular regulation of these processes. To determine the expression pattern of putative AP2/ERF family genes, the RiceXPro database was probed for 12 different kinds of rice tissues. The expression of these genes was detected in the root, stem, leaf, sheath, inflorescence, ovary, anther, embryo, pistil, lemma, palea, and endosperm. Most of the AP2/ERF genes indicate some degree of tissue specificity.[107] Among the vegetative organs, the expression of these genes was found most abundant in the root (14%), followed by the leaf (8%) and stem (6%). With regards to reproductive organs, they were most common in the embryo (12%) and lemma (11%) (Fig. 10). The expression analysis of each gene in different tissues is determined by signal intensity values derived from the RiceXPro database. Most of the genes have a similar degree of tissue specificity except for OsAP2/ERF#001, OsAP2/ERF#128, and OsAP2/ERF#048, which exhibit more expression in 12, 10, and 9 rice tissues, respectively (Fig. 11). It should be noted that the genes with higher expression levels belong to the subgroups IIIc, IVc, IVb, and Ib. The genes in group-IIIb and group-IVd have the same level of gene expression. It is apparent that the members of the subgroup Ia and group-II have the lowest gene expression in the rice tissues used in this canvas (Fig. 11).

Furthermore, the function of genes in each group divulges that the members of group-IIIb were highly present in the root. Gene expression during inflorescence, pistil, and anther development was high in members of group-IIIa. During embryogenesis, expression was high in group-Ia and group-IIIb genes. During the heading stage in rice, these AP2/

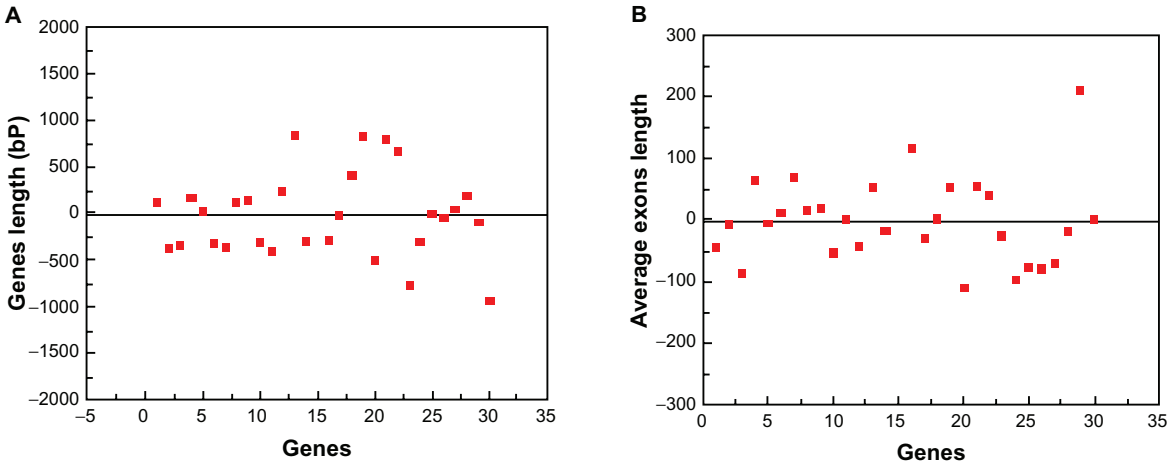


Figure 9. Distribution of AP2/ERF family genes with full cDNA information. (**A**) Distribution of gene length in 2000 bp mapping hits. (**B**) Exon size distribution in gene length.

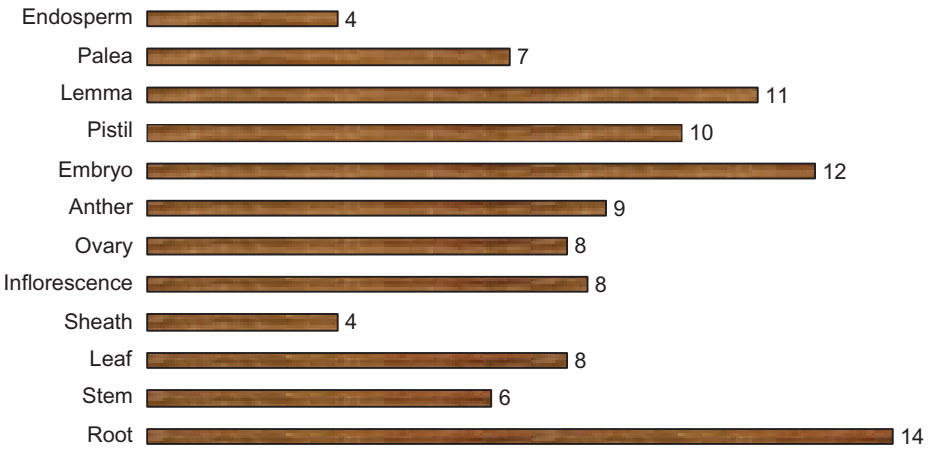


Figure 10. Distribution of rice AP2/ERF family genes in vegetative (root, stem, leaf and sheath) and reproductive tissues (inflorescence, ovary, anther, embryo, pistil, lemma, palea and endosperm).
Note: The expression profile is determined by using the RiceXPro database.

ERF genes were expressed in the lemma and palea in members of group-IVd. Moderate gene expression in the leaf was observed in group-IIIa, whereas the lowest expression of AP2/ERF genes in the stem, sheath, and endosperm was observed in all groups (Fig. 12).

Recently, transcriptome analysis of rice in mature root tissue by massive parallel sequencing unveiled previously unrecognized tissue-specific expression profiles and render an interesting platform to study the differential regulation of transcribed regions of root tissues.[108] Moreover, this technique foregoes canvas

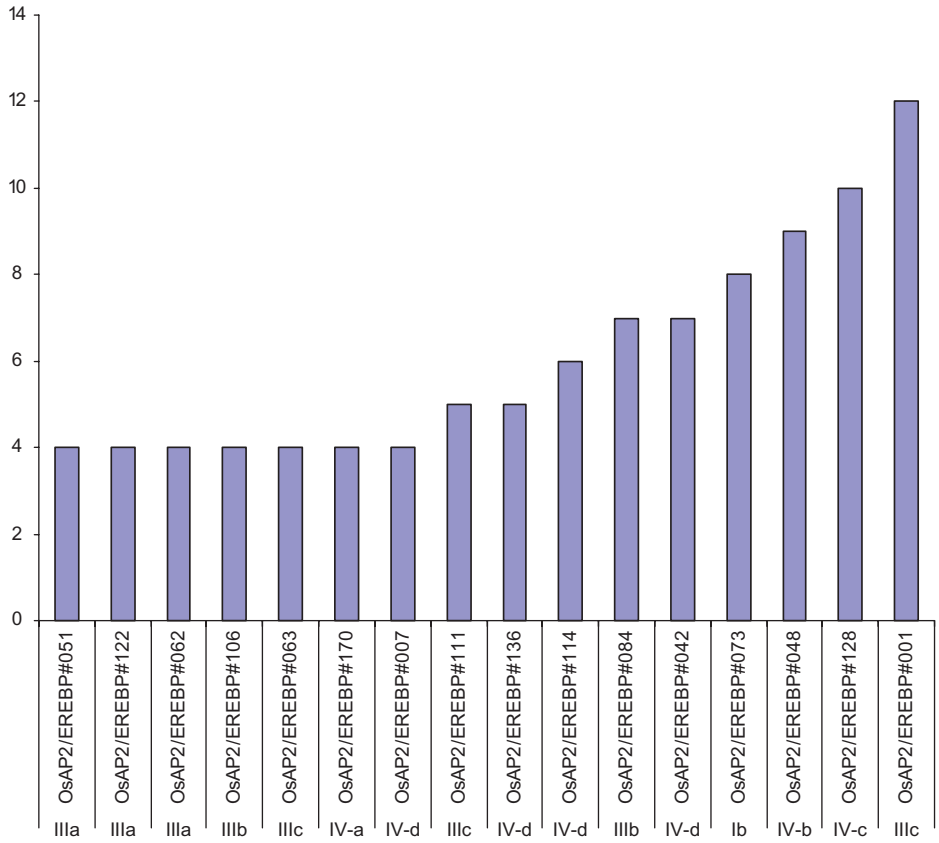


Figure 11. Gene's classification according to their expression level in each tissue.

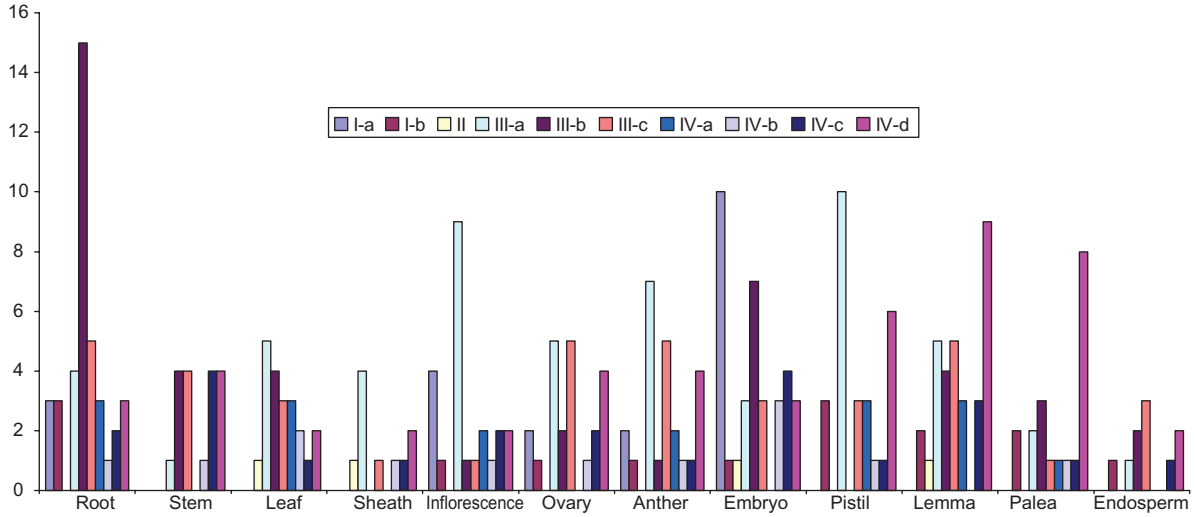


Figure 12. Tissue specific expression of the AP2/ERF family genes in each group.

spotlighting in which OsDREB1B (a DREB subfamily gene) was expressed significantly in roots compared to the leaves, shoots, growing points, mature seeds, and other tissues of unstressed rice plants.[109] The subfamily gene, OsAP211, is extremely prominent in shoot tips in mature rice and immature seeds at the booting stage compared to other rice tissues.[110] The ERF subfamily gene, OsAP25, is detected in the leaves, shoots, roots, growing points, flower, immature seeds, and mature seeds of rice.[111] In wheat, 9 types of tissues are used for expression profile analysis and the transcripts of these genes are most abundant in leaves followed by roots, seeds, and stem.[107]

Chromosome position of the identified rice AP2/ERF genes

To examine the genomic distribution of AP2/EREBP genes on rice chromosomes (chr), we identified their positions using a Rice TOGO Browser database search.[112] A total of 170 rice AP2/EREBP genes were localized on the 12 chromosomes with an uneven distribution. A similar situation also occurred with respect to the OsERF and AtERF family genes, thereby indicating that ERF genes are distributed widely among monocot and eudicot genomes.[10] The OsAP2/EREBP genes are present in all regions on a single chromosome (eg, at the telomeric ends, near the centromere, and in between) and are distributed individually or in clusters (Fig. 13). Chr2 has the maximum number (26) of OsAP2/EREBP genes, and chr4 and chr6 have 22 and 20 genes, respectively. The high number of AP2/ERF sequences in these chromosomes is due to the repetition of adjacent genes. Interestingly, the same tendency was found for the chromosomal location of group-II (RAV binding domain) genes. Only four OsAP2/EREBP genes were identified on chr11, all of which are found on the short arm. Two OsAP2/EREBP genes (Os07g0410300 and Os07g0410700) arranged as tandem duplications were found around the centromere on chr7. Less than 10 chromosome locations are found on each of chr10, chr11, and chr12.

Comparative mapping between rice, wheat, and Arabidopsis genomes

To explore orthologs/homologs, wheat and Arabidopsis sequences were used to BLAST against the rice genome using BLASTx and BLASTn, respectively. The aligned sequences with more than 60% similarity to rice genes elsewhere in rice genome produced high-scoring pairs (HSPs). These HSPs were selected based on similarity percentage criteria (Supplementary Tables S4 and 5), which facilitate the differentiation of orthologous regions in the genome with high sureness and, subsequently, describes shared duplications between these genomes. These syntenic surveys facilitate finding different orthologous/homologous gene annotations in the comparative study of rice, wheat, and Arabidopsis genomes.[113–117]

In wheat (*Triticum aestivum* L.), AP2/ERF family genes were compared to rice orthologous loci in the rice genome on chr1 to chr9, and not chr10, chr11, or chr12.[118]

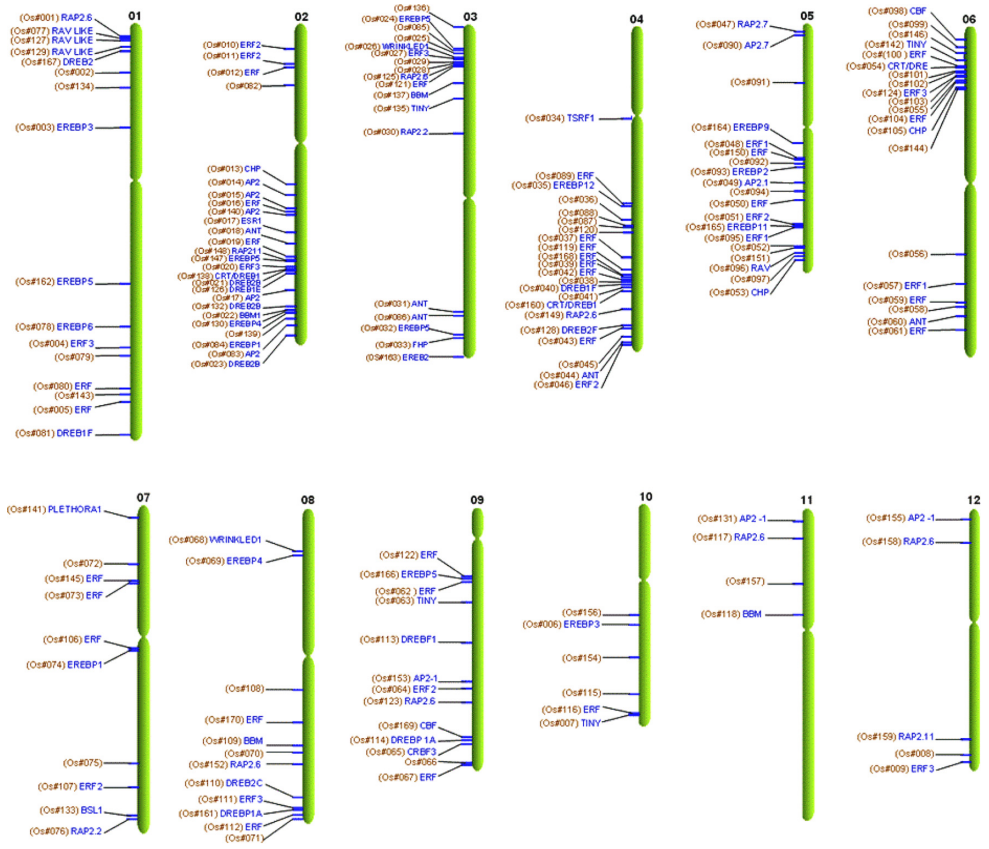


Figure 13. The physical location of the AP2/ERF genes on rice chromosomes.
Notes: The chromosome number is indicated at the top of each chromosome. The genes are indicated on the locus by their generic names and function.

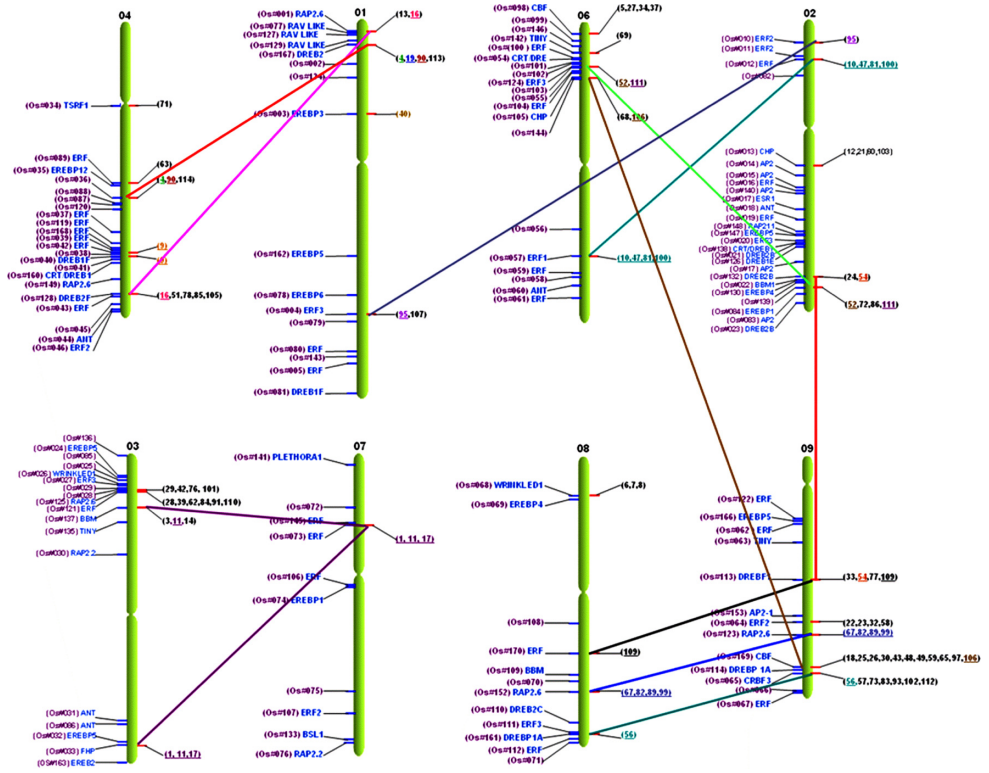


Figure 14. Comparative mapping of wheat AP2/ERF genes on rice chromosomes.

The microcollinearity allowed a higher resolution of distribution in chr9, followed by chr3, chr2, and chr6, while the minimum predicted orthology was found in chr7. Comparison with rice chromosomes predicts 25% gene orthologs in chr9 followed by 13% on chr2 (Supplementary Table S4). CRT/DRE subfamily genes from wheat found syntenic loci on rice on chr1, chr2, chr4, chr6, chr8, and chr9. The majority of these genes were conserved on chr9 followed by chr1 and chr8. In addition, ERF subfamily gene orthlogous loci were found on chr1, chr2, chr3, chr6, and chr9 respective to their similarity, and maximum syntenic regions were mapped on chr2. Regarding the AP2 family with double domains, these genes were distributed on chr3, chr7, and chr8, and the maximum conserved loci were preserved on chr3 (Fig. 14). The comparative distribution of wheat genes to the rice genome clearly designates that rice chromosomes have passed through different phylogenies.[118]

Comparative mapping between rice and Arabidopsis displayed 11 orthologous/homologous loci, unveiling the broad extent of collinearity between these genomes.[119,120] The homologs in rice genome were found on chr1, chr2, chr4, chr5, chr6, chr9,

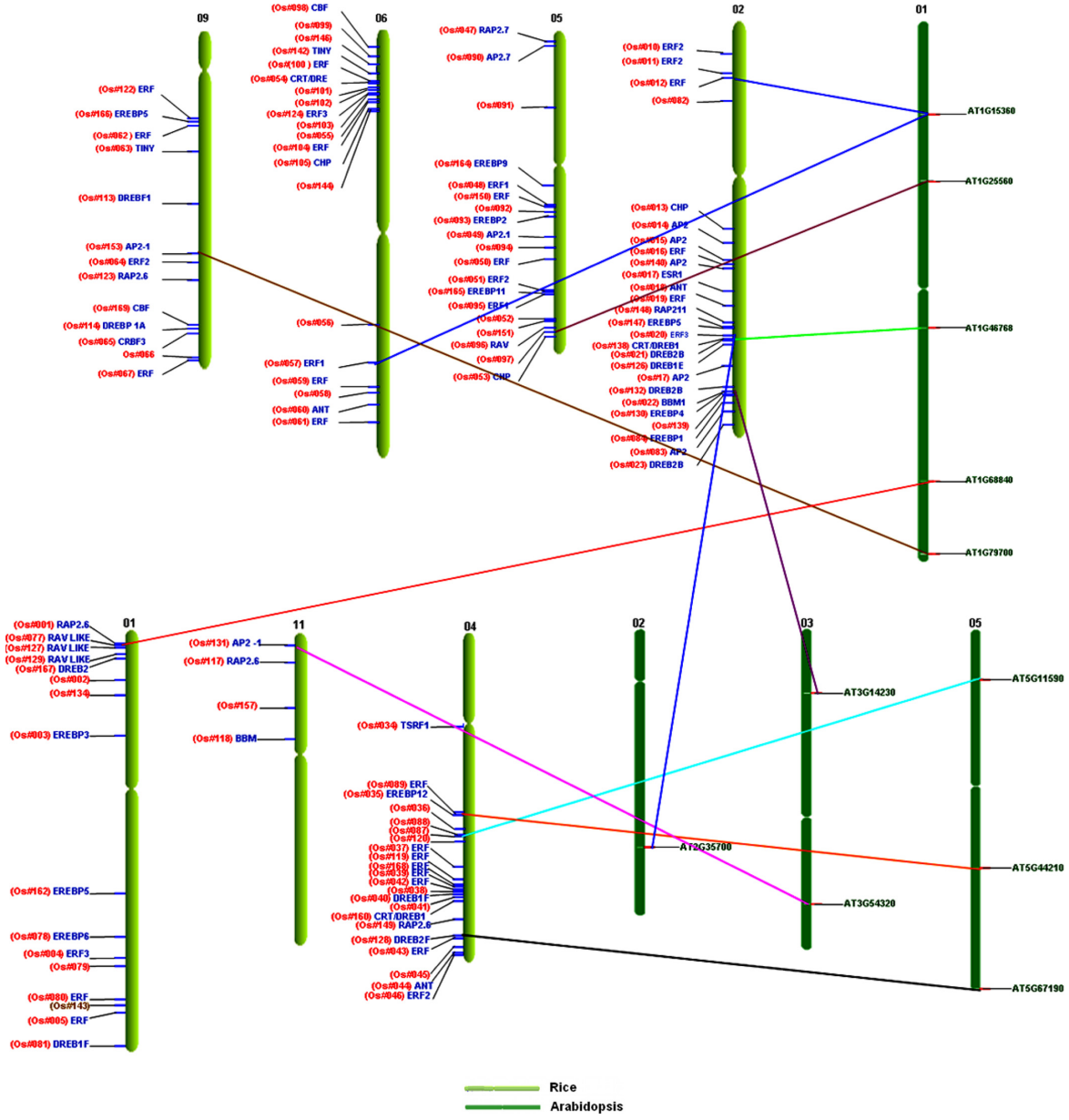


Figure 15. Comparative mapping of rice and Arabidopsis AP2/ERF family.

and chr11; whereas, in Arabidopsis, this collinearity was found preserved on chr1, chr2, chr3, and chr5. The chromosomal distribution of the AP2/ERF family genes revealed that chr1 in Arabidopsis has more homologs (45%) followed by chr5 (27%). These findings are consistent with the evolution and divergence of the ERF family genes in Arabidopsis.[10] In rice, the conserved homology is found in chr2 (36%) followed by chr4 (27%) (Supplementary Table S5). The ERF subfamily genes of Arabidopsis on chr1, chr3, and chr5 have syntenic loci in rice on chr2, chr4, chr6, and chr9. The genes related to abiotic stresses, such as DRE and TINY on Arabidopsis chr2 and chr5 have their orthologs on rice chr2 and chr4, respectively. The subfamily genes with B3 domains in Arabidopsis found on chr1 have orthologs on rice chr1 and chr5, whereas the subfamily gene with AP2 double domains in Arabidopsis on chr3 is syntenic to rice chr11 (Fig. 15). It is reported that 98% of homologs of known maize, wheat, and barley proteins are found in rice. Synteny and gene homology between monocot (rice and the other cereal) genomes are extensive, whereas synteny with dicots (Arabidopsis) is limited.[121] Moreover, scant collinearity in gene order is also observed between rice and Arabidopsis genomes using comparative genomics.[122]

Conclusions

Transcription factor proteins primarily regulate a web of biological processes and have egressed as a powerful tool for the manipulation of composite metabolic pathways. In this study, 170 AP2/ERF genes are identified from different rice databases available in the public domain. The results reveal much about the diversification of AP2/ERF family genes in the rice genome. Gene translocation and segmental duplication might have imparted towards the expansion of the AP2/ERF gene family. During the enlargement of the AP2/ERF gene family, many groups and subgroups evolved, resulting in a high level of functional divergence. The conserved motifs present in their respective clades suggest the specificity of the genes for function in this group. Comparative mapping revealed that the homologs/orthologs are present in rice, wheat, and Arabidopsis indicating that many of the genes in these species antedated the divergence of monocots and eudicots. The expression analysis of AP2/ERF family genes furnishes a new avenue for functional analyses in rice. During the domestication of rice, selection pressure may lead to selecting the genotypes that have specific conserved motifs and have related molecular functions against natural confronts. Modern bioinformatics and biotechnology tools may predict and characterized these genes in their respective clades. As a model plant, and having a great synteny with the grass family with respect to gene structure, the information generated about the AP2/ERF gene family in rice will also provide a platform for predicting the function of genes of crops whose genome sequences are in their infancy.

Author Contributions

Conceived and designed the experiments: M.R. Analysed the data: M.R. Wrote the first draft of the manuscript: M.R. Contributed to the writing of the manuscript: M.R. Agree with manuscript results and conclusions: M.R., G.H. and Y.G. Jointly developed the structure and arguments for the paper: M.R., J.H. Made critical revisions and approved final version: M.R., X.Y. All authors reviewed and approved of the final manuscript.

Funding

This work is supported by the National Major Specific Project of the People's Republic of China (2011 ZX 08002-004, 2009 ZX 08016-001A, 2009 ZX 08002-013B) and the International Science and Technology Cooperation project of China MoST (2009 DFB 20290).

Competing Interests

Author(s) disclose no potential conflicts of interest.

Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and

conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

References

1. Abogadallah GM, Nada RM, Malinowski R, Quick P. Overexpression of HARDY, an AP2/ERF gene from Arabidopsis, improves drought and salt tolerance by reducing transpiration and sodium uptake in transgenic Trifolium alexandrinum L. *Planta*. 2011;233:1265–76.
2. Agarwal P, Agarwal PK, Joshi AJ, Sopory SK, Reddy MK. Overexpression of PgDREB2A transcription factor enhances abiotic stress tolerance and activates downstream stress-responsive genes. *Molecular Biology Reports*. 2010;37(2):1125–35.
3. Gao FC JM, Xiong AS, Peng RH, Liu JG, Cai B, Yao QH. Isolation and characterization of a novel AP2/EREBP-type transcription factor OsAP211 in Oryza sativa. *Biologia Plantarum*. 2009;53(4):643–9.
4. Hu Y, Zhao L, Chong K, Wang T. Overexpression of OsERF1, a novel rice ERF gene, up-regulates ethylene-responsive genes expression besides affects growth and development in Arabidopsis. *Journal of Plant Physiology*. 2008;165(16):1717–25.
5. Riechmann JL, Meyerowitz EM. The AP2/EREBP family of plant transcription factors. *Biological Chemistry*. 1998;379:633–46.
6. Sakuma YL Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K. DNA-binding specificity of the ERF/AP2 domain of Arabidopsis DREBs, transcription factors involved in dehydration-and cold-inducible gene expression. *Biochemical and Biophysical Research Communications*. 2002;290(3):998–1009.
7. Jofuku KD, Boer BGW, Montagu MV, Okamuro JK. Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *The Plant Cell Online*. 1994;6(9):1211–25.
8. Ohme-Takagi M, Shinshi H. Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *The Plant Cell Online*. 1995;7(2):173–82.
9. Zhou J, Tang X, Martin GB. The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a cis-element of pathogenesis-related genes. *The EMBO Journal*. 1997;16(11):3207–18.
10. Nakano T, Suzuki K, Fujimura T, Shinshi H. Genome-wide analysis of the ERF gene family in Arabidopsis and rice. *Plant Physiology*. 2006;140(2):411–32.
11. Shigyo M, Ito M. Analysis of gymnosperm two-AP2-domain-containing genes. *Development Genes and Evolution*. 2004;214(3):105–14.
12. Hao D, Ohme-Takagi M, Sarai A. Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. *Journal of Biological Chemistry*. 1998;273(41):26857–61.
13. Yamaguchi-Shinozaki K, Shinozaki K. A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress. *The Plant Cell Online*. 1994;6(2):251–64.
14. Sharma MKK R, Solanke AU, Sharma R, Tyagi AK, Sharma AK. Identification, phylogeny, and transcript profiling of ERF family genes during development and abiotic stress treatments in tomato. *Molecular Genetics and Genomics*. 2010;284:455–75.
15. Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M, Suzuki M. A novel mode of DNA recognition by a β-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *The EMBO Journal*. 1998;17(18):5484–96.
16. Jaillon OAJM, Noel B, Policriti A, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*. 2007;449(7161):463–7.
17. Zhuang JC B, Peng RH, Zhu B, Jin XF, et al. Genome-wide analysis of the AP2/ERF gene family in Populus trichocarpa. *Biochemical and Biophysical Research Communications*. 2008;371(3):468–74.
18. Sharoni AMN M, Satoh K, Shimizu T, et al. Gene structures, classification and expression models of the AP2/EREBP transcription factor family in rice. *Plant and Cell Physiology*. 2011;52(2):344–60.
19. Zhu QZ J, Gao X, Tong J, Xiao L, Li W, Zhang H. The Arabidopsis AP2/ERF transcription factor RAP2. 6 participates in ABA, salt and osmotic stress responses. *Gene*. 2010;457(1–2):1–12.
20. Zarei AK AP, Younessi P, Montiel G, Champion A, Memelink J. Two GCC boxes and AP2/ERF-domain transcription factor ORA59 in jasmonate/ethylene-mediated activation of the PDF1. 2 promoter in Arabidopsis. *Plant Molecular Biology*. 2011;75:321–31.
21. Yu ZXL, JX, Yang CQ, Hu WL, Wang LJ, Chen XY. The Jasmonate-Responsive AP2/ERF Transcription Factors AaERF1 and AaERF2 Positively Regulate Artemisinin Biosynthesis in Artemisia annua L. *Molecular Plant*. 2011:1–13.
22. Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. AP2/ERF family transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*. 2011;1819(2):86–96.
23. Song CPA M, Ohta M, Guo Y, Halfter U, Wang P, Zhu JK. Role of an Arabidopsis AP2/EREBP-type transcriptional repressor in abscisic acid and drought stress responses. *The Plant Cell Online*. 2005;17(8):2384–96.
24. Liu QK M, Sakuma Y, Abe H, Miura S. Yamaguchi-Shinozaki K, Shinozaki K. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought-and low-temperature-responsive gene expression, respectively, in Arabidopsis. *The Plant Cell Online*. 1998;10(8):1391–406.
25. Bihani P, Char B, Bhargava S. Transgenic expression of sorghum DREB2 in rice improves tolerance and yield under water limitation. *The Journal of Agricultural Science*. 2011;149(01):95–101.
26. Nakashima KS ZK, Sakuma Y, Seki M, Miura S, Shinozaki K. Yamaguchi-Shinozaki K. Organization and expression of two Arabidopsis DREB2 genes encoding DRE-binding proteins involved in dehydration-and high-salinity-responsive gene expression. *Plant Molecular Biology*. 2000;42(4):657–65.
27. Golldack D, Lüking I, Yang O. Plant tolerance to drought and salinity: stress regulating transcription factors and their functional significance in the cellular transcriptional network. *Plant Cell Reports*. 2011;30:1383–91.
28. Carvallo MAP, MT, Jeknić Z, et al. A comparison of the low temperature transcriptomes and CBF regulons of three plant species that differ in freezing tolerance: Solanum commersonii, Solanum tuberosum, and Arabidopsis thaliana. *Journal of Experimental Botany*. 2011;62(11):3807–19.
29. Maki HSM, Ogawa K, Kaida R, Yamamoto N, Kaneko ST. Cloning and expression profile of an ERF gene isolated from cold-stressed poplar cells (Populus nigra). *Cytologia*. 2011;76(1):11–8.
30. Li CWS RC, Cheng CP, You SJ, Hsieh TH, Chao TC, Chan MT. Tomato RAV transcription factor is a pivotal modulator involved in the AP2/EREBP-mediated defense pathway. *Plant Physiology*. 2011;156(1):213–27.
31. Zhou YLX, MR, Zhao MF, et al. Genome-wide gene responses in a transgenic rice line carrying the maize resistance gene Rxo1 to the rice bacterial streak pathogen, Xanthomonas oryzae pv. oryzicola. *BMC Genomics*. 2010;11(1):78.
32. Rzewuski G, Sauter M. Ethylene biosynthesis and signaling in rice. *Plant Science*. 2008;175(1):32–42.
33. Aukerman MJ, Sakai H. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. *The Plant Cell Online*. 2003;15(11):2730–41.
34. Kirch T, Simon R, Grünewald M, Werr W. The DORNRÖSCHEN/ENHANCER OF SHOOT REGENERATION1 gene of Arabidopsis acts in the control of meristem cell fate and lateral organ development. *The Plant Cell Online*. 2003;15(3):694–705.
35. Kitomi YI H, Hobo T, Aya K, Kitano H, Inukai Y. The auxin responsive AP2/ERF transcription factor CROWN ROOTLESS5 is involved in crown root initiation in rice through the induction of OsRR1, a type-A response regulator of cytokinin signaling. *The Plant Journal*. 2011;67(3):472–84.

36. Imin N, Nizamidin M, Wu T, Rolfe BG. Factors involved in root formation in Medicago truncatula. *Journal of Experimental Botany*. 2007;58(3): 439–51.
37. El Ouakfaoui S, Schnell J, Abdeen A, et al. Control of somatic embryogenesis and embryo development by AP2 transcription factors. *Plant Molecular Biology*. 2010;74:313–26.
38. Alonso JM, Stepanova AN, Leisse TJ, et al. Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science*. 2003;301(5633):653–7.
39. Yu Xin H, Yong Hong W, Xin Fang L, Jia Yang L. Arabidopsis RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development. *Cell Research*. 2004;14(1):8–15.
40. Salvi SS G, Morgante M, Tomes D, et al. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proceedings of the National Academy of Sciences*. 2007;104(27): 11376–81.
41. Sohn KH, Lee SC, Jung HW, Hong JK, Hwang BK. Expression and functional roles of the pepper pathogen-induced transcription factor RAV1 in bacterial disease resistance, and drought and salt stress tolerance. *Plant Molecular Biology*. 2006;61(6):897–915.
42. Zhuang J, Sun CC, Zhou XR, Xiong AS, Zhang J. Isolation and characterization of an AP2/ERF-RAV transcription factor BnaRAV-1-HY15 in Brassica napus L. HuYou15. *Molecular Biology Reports*. 2011;38:3921–8.
43. Wuitschick JD, Lindstrom PR, Meyer AE, Karrer KM. Homing endonucleases encoded by germ line-limited genes in Tetrahymena thermophila have APETELA2 DNA binding domains. *Eukaryotic Cell*. 2004;3(3):685–94.
44. Magnani E, Sjölander K, Hake S. From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. *The Plant Cell Online*. 2004;16(9):2265–77.
45. Dalgaard JZ, Klar AJ, Moser MJ, Holley WR, Chatterjee A, Mian IS. Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. *Nucleic Acids Research*. 1997;25(22):4626–38.
46. Sayers EW, Barrett T, Benson DA, et al. Database resources of the national center for biotechnology information. *Nucleic Acids Research*. 2011;39(Suppl 1):D38–51.
47. Gao G, Zhong Y, Guo A, et al. DRTF: a database of rice transcription factors. *Bioinformatics*. 2006;22(10):1286–7.
48. Ouyang S, Zhu W, Hamilton J, et al. The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Research*. 2007;35(Suppl 1):D883–7.
49. Kikuchi S, Satoh K, Nagata T, et al. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science*. 2003;301(5631): 376–9.
50. Dong Q, Schlueter SD, Brendel V. PlantGDB, plant genome database and analysis tools. *Nucleic Acids Research*. 2004;32(Suppl 1):D354–9.
51. Tanaka T, Antonio BA, Kikuchi S, et al. The rice annotation project database (RAP-DB): 2008 update. *Nucleic Acids Research*. 2008;36(Suppl 1): D1028–33.
52. Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA1. *Journal of Molecular Biology*. 1997;268(1):78–94.
53. Huang X, Madan A. CAP3: A DNA sequence assembly program. *Genome Research*. 1999;9(9):868–77.
54. Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: identification of signaling domains. *Proceedings of the National Academy of Sciences*. 1998;95(11):5857–64.
55. Larkin M, Blackshields G, Brown N, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007;23(21):2947–8.
56. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*. 2007;24(8):1596–9.
57. Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*. 2009;37(Suppl 2):W202–8.
58. Guo AY, Zhu QH, Chen X, Luo JC. GSDS: a gene structure display server]. *Yi chuan = Hereditas/Zhongguo yi chuan xue hui bian ji*. 2007;29(8):1023–6.
59. Lynch M, Kewalramani A. Messenger RNA surveillance and the evolutionary proliferation of introns. *Molecular Biology and Evolution*. 2003;20(4): 563–71.
60. Sato Y, Antonio BA, Namiki N, et al. RiceXPro: a platform for monitoring gene expression in japonica rice grown under natural field conditions. *Nucleic Acids Research*. 2011;39(Suppl 1):D1141–8.
61. Rhee SY, Beavis W, Berardini TZ, et al. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Research*. 2003;31(1):224–8.
62. Mizukami Y, Fischer RL. Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis. *Proceedings of the National Academy of Sciences*. 2000;97(2):942–497.
63. Boutilier K, Offringa R, Sharma VK, et al. Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *The Plant Cell Online*. 2002;14(8):1737–49.
64. Cernac A, Benning C. WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. *The Plant Journal*. 2004;40(4):575–85.
65. Aida M, Beis D, Heidstra R, et al. The PLETHORA genes mediate patterning of the Arabidopsis root stem cell niche. *Cell*. 2004;119(1):109–20.
66. Drews GN BJ, Meyerowitz EM. Negative regulation of the Arabidopsis homeotic gene AGAMOUS by the APETALA2 product. *Cell*. 1991;14(65):991–1002.
67. Krizek BA, Prost V, Macias A. AINTEGUMENTA promotes petal identity and acts as a negative regulator of AGAMOUS. *The Plant Cell Online*. 2000;12(8):1357–66.
68. Meyer K, Damude HG, Everard JD, Ripp KG, Stecca KL. Use of a seed specific promoter to drive odp1 expression in cruciferous oilseed plants to increase oil content while maintaining normal germination: US Patent App. 20,100/257,635; 2010.
69. Klucher KM, Chow H, Reiser L, Fischer RL. The AINTEGUMENTA gene of Arabidopsis required for ovule and female gametophyte development is related to the floral homeotic gene APETALA2. *The Plant Cell Online*. 1996;8(2):137–53.
70. Okamuro JK, Caster B, Villarroel R, Van Montagu M, Jofuku KD. The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis. *Proceedings of the National Academy of Sciences*. 1997;94(13):7076–81.
71. Nishiuchi T, Suzuki K, Kitajima S, Sato F, Shinshi H. Wounding activates immediate early transcription of genes for ERFs in tobacco plants. *Plant Molecular Biology*. 2002;49(5):473–82.
72. Zhang H, Zhang D, Chen J, et al. Tomato stress-responsive factor TSRF1 interacts with ethylene responsive element GCC box and regulates pathogen resistance to Ralstonia solanacearum. *Plant Molecular Biology*. 2004;55(6):825–34.
73. Xu ZS, Xia LQ, Chen M, et al. Isolation and molecular characterization of the Triticum aestivum L. ethylene-responsive factor 1 (TaERF1) that increases multiple stress tolerance. *Plant Molecular Biology*. 2007;65(6):719–32.
74. Meng X, Li F, Liu C, Zhang C, Wu Z, Chen Y. Isolation and characterization of an ERF transcription factor gene from cotton (Gossypium barbadense L.). *Plant Molecular Biology Reporter*. 2010;28(1):176–83.
75. Fujimoto SY, Ohta M, Usui A, Shinshi H, Ohme-Takagi M. Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box–mediated gene expression. *The Plant Cell Online*. 2000;12(3):393–404.
76. Ashida Y, Yokobatake N, Kohchi C, Shimoda K, Hirata T. Cloning of cDNA encoding ethylene-responsive element binding protein-5 in the cultured cells of Nicotiana tabacum. *Mitochondrial DNA*. 2000;11(1–2):125–9.
77. Ohta M, Ohme-Takagi M, Shinshi H. Three ethylene-responsive transcription factors in tobacco with distinct transactivation functions. *The Plant Journal*. 2000;22(1):29–38.
78. Lu J, Ju H, Zhou G, et al. An EAR-motif-containing ERF transcription factor affects herbivore-induced signaling, defense and resistance in rice. *The Plant Journal*. 2011;68(4):583–96.

79. Ohta M, Matsui K, Hiratsu K, Shinshi H, Ohme-Takagi M. Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *The Plant Cell Online*. 2001;13(8):1959–68.
80. Koyama T, Okada T, Kitajima S, Ohme-Takagi M, Shinshi H, Sato F. Isolation of tobacco ubiquitin-conjugating enzyme cDNA in a yeast two-hybrid system with tobacco ERF3 as bait and its characterization of specific interaction. *Journal of Experimental Botany*. 2003;54(385):1175–81.
81. Ye R, Yao QH, Xu ZH, Xue HW. Development of an efficient method for the isolation of factors involved in gene transcription during rice embryo development. *The Plant Journal*. 2004;38(2):348–57.
82. Díaz-Martín J, Almoguera C, Prieto-Dapena P, Espinosa JM, Jordano J. Functional interaction between two transcription factors involved in the developmental regulation of a small heat stress protein gene promoter. *Plant Physiology*. 2005;139(3):1483–94.
83. Jaglo KR, Kleff S, Amundsen KL, et al. Components of the Arabidopsis C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved inbrassica napus and other plant species. *Plant Physiology*. 2001;127(3):910–7.
84. Zhang X, Fowler SG, Cheng H, et al. Freezing-sensitive tomato has a functional CBF cold response pathway, but a CBF regulon that differs from that of freezing-tolerant Arabidopsis. *The Plant Journal*. 2004;39(6):905–19.
85. Ito Y, Katsura K, Maruyama K, et al. Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice. *Plant and Cell Physiology*. 2006;47(1):141–53.
86. Alonso-Blanco C, Gomez-Mena C, Llorente F, Koornneef M, Salinas J, Martínez-Zapater JM. Genetic and molecular analyses of natural variation indicate CBF2 as a candidate gene for underlying a freezing tolerance quantitative trait locus in Arabidopsis. *Plant Physiology*. 2005;139(3):1304–12.
87. Wang L, Luo Y, Zhang L, et al. Isolation and characterization of a C-repeat binding transcription factor from maize. *Journal of Integrative Plant Biology*. 2008;50(8):965–74.
88. Dubouzet JG, Sakuma Y, Ito Y, et al. OsDREB genes in rice, Oryza sativa L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *The Plant Journal*. 2003;33(4):751–63.
89. Matsumoto T, Wu J, Kanamori H, et al. The map-based sequence of the rice genome. *Nature*. 2005;436(7052):793–800.
90. Wang Q, Guan Y, Wu Y, Chen H, Chen F, Chu C. Overexpression of a rice OsDREB1F gene increases salt, drought, and low temperature tolerance in both Arabidopsis and rice. *Plant Molecular Biology*. 2008;67(6):589–602.
91. Mallikarjuna G, Mallikarjuna K, Reddy M, Kaul T. Expression of OsDREB2 A transcription factor confers enhanced dehydration and salt stress tolerance in rice (Oryza sativa L.). *Biotechnology Letters*. 2011;33:1689–97.
92. Matsukura S, Mizoi J, Yoshida T, et al. Comprehensive analysis of rice DREB2-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. *Molecular Genetics and Genomics*. 2010;283(2):185–96.
93. Alexandrov NN, Brover VV, Freidin S, et al. Insights into corn genes derived from large-scale cDNA sequencing. *Plant Molecular Biology*. 2009;69(1):179–94.
94. Reyes JC, Muro-Pastor MI, Florencio FJ. The GATA family of transcription factors in Arabidopsis and rice. *Plant Physiology*. 2004;134(4):1718–32.
95. Büttner M, Singh KB. Arabidopsis thaliana ethylene-responsive element binding protein (AtEBP), an ethylene-inducible, GCC box DNA-binding protein interacts with an ocs element binding protein. *Proceedings of the National Academy of Sciences*. 1997;94(11):5961–6.
96. Cheong YH, Moon BC, Kim JK, et al. BWMK1, a rice mitogen-activated protein kinase, locates in the nucleus and mediates pathogenesis-related gene expression by activation of a transcription factor. *Plant Physiology*. 2003;132(4):1961–72.
97. Albrecht V, Ritz O, Linder S, Harter K, Kudla J. The NAF domain defines a novel protein–protein interaction module conserved in Ca2+-regulated kinases. *The EMBO Journal*. 2001;20(5):1051–63.
98. Qu LJ, Zhu YX. Transcription factor families in Arabidopsis: major progress and outstanding issues for future research. *Current Opinion in Plant Biology*. 2006;9(5):544–9.
99. Feng JX, Liu D, Pan Y, et al. An annotation update via cDNA sequence analysis and comprehensive profiling of developmental, hormonal or environmental responsiveness of the Arabidopsis AP2/EREBP transcription factor gene family. *Plant Molecular Biology*. 2005;59(6):853–68.
100. Hagen G, Guilfoyle T. Auxin-responsive gene expression: genes, promoters and regulatory factors. *Plant Molecular Biology*. 2002;49(3):373–85.
101. Li J, Dai X, Zhao Y. A role for auxin response factor 19 in auxin and ethylene signaling in Arabidopsis. *Plant Physiology*. 2006;140(3):899–908.
102. Manosalva PM, Davidson RM, Liu B, et al. A germin-like protein gene family functions as a complex quantitative trait locus conferring broad-spectrum disease resistance in rice. *Plant Physiology*. 2009;149(1):286–96.
103. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. *The Plant Cell Online*. 2003;15(4):809–34.
104. Yaish MWF, Peng M, Rothstein SJ. AtMBD9 modulates Arabidopsis development through the dual epigenetic pathways of DNA methylation and histone acetylation. *The Plant Journal*. 2009;59(1):123–35.
105. Swarbreck D, Wilks C, Lamesch P, et al. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Research*. 2008;36(Suppl 1):D1009–14.
106. Weber H, Hellmann H. Arabidopsis thaliana BTB/POZ-MATH proteins interact with members of the ERF/AP2 transcription factor family. *FEBS Journal*. 2009;276(22):6624–35.
107. Zhuang J, Chen JM, Yao QH, et al. Discovery and expression profile analysis of AP2/ERF family genes from Triticum aestivum. *Molecular Biology Reports*. 2011;38(2):745–53.
108. Kyndt T, Denil S, Haegeman A, et al. Transcriptome analysis of rice mature root tissue and root tips in early development by massive parallel sequencing. *Journal of Experimental Botany*. 2012.
109. Qin Q, Liu J, Zhang Z, et al. Isolation, optimization, and functional analysis of the cDNA encoding transcription factor OsDREB1B in Oryza Sativa L. *Molecular Breeding*. 2007;19(4):329–40.
110. Gao F, Chen JM, Xiong AS, et al. Isolation and characterization of a novel AP2/EREBP-type transcription factor OsAP211 in Oryza sativa. *Biologia Plantarum*. 2009;53(4):643–9.
111. Fu XY, Zhang Z, Peng RH, et al. Isolation and characterization of a novel cDNA encoding ERF/AP2-type transcription factor OsAP25 from Oryza Sativa L. *Biotechnology Letters*. 2007;29(8):1293–9.
112. Nagamura Y, Antonio BA, Sato Y, et al. Rice TOGO browser: a platform to retrieve integrated information on rice functional and applied genomics. *Plant and Cell Physiology*. 2011;52(2):230–7.
113. Bailey P, McKibbin R, Lenton J, Holdsworth M, Flintham J, Gale M. Genetic map locations for orthologous Vp1 genes in wheat and rice. *TAG Theoretical and Applied Genetics*. 1999;98(2):281–4.
114. Distelfeld A, Pearce SP, Avni R, et al. Divergent functions of orthologous NAC transcription factors in wheat and rice. *Plant Molecular Biology*. 2012:1–10.
115. Wu H, Dorse S, Bhave M. In silico identification and analysis of the protein disulphide isomerases in wheat and rice. *Biologia*. 2012;67(1):48–60.
116. Choisne N, Demange N, Orjeda G, et al. The sequence of rice chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. *BMC Biol*. 2005;3:20.
117. Deynze AE, Nelson JC, Yglesias ES, et al. Comparative mapping in grasses. Wheat relationships. *Molecular and General Genetics MGG*. 1995;248(6):744–54.
118. Salse J, Bolot S, Throude M, et al. Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *The Plant Cell Online*. 2008;20(1):11–24.
119. Salse J, Piégu B, Cooke R, Delseny M. Synteny between Arabidopsis thaliana and rice at the genome level: a tool to identify conservation in the ongoing rice genome sequencing project. *Nucleic Acids Research*. 2002;30(11):2316–28.
120. Devos KM, Beales J, Nagamura Y, Sasaki T. Arabidopsis–rice: will colinearity allow gene prediction across the eudicot–monocot divide? *Genome Research*. 1999;9(9):825–9.
121. Goff SA, Ricke D, Lan TH, et al. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). *Science*. 2002;296(5565):92–100.
122. Liu H, Sachidanandam R, Stein L. Comparative genomics between rice and Arabidopsis shows scant collinearity in gene order. *Genome Research*. 2001;11(12):2020–6.

Supplementary Data

Table S1. The functional description of conserved motifs found on both sides of the AP2/ERF domain.

Motif	Species	Protein name	Reference
[A/W]R[D/E][H/P]LF[D/E]K[T/A] VTPSDVGKLNRLV[I/V]PKQ[H/Q] AEKHFP[L/F] KGVLLNFED[A/G][A/E/D] GKVW[R/K]FRYSYWNSSQSYV	Arabidopsis	AP2/ERF and B3 domain-containing protein Function as negative regulator of plant growth and development	39
MCGGAI	Arabidopsis	Ethylene-responsive transcription factor RAP2-3, RAP2-2, RAP2-12	95
DFEA	Rice	Ethylene-responsive transcription factor 1	96
LPR[P/A]	Arabidopsis	CBL-interacting serine/threonine-protein kinase-12	97
D[I/V]QAA	Arabidopsis	Ethylene-responsive transcription factor ERF037	98
D[I/V]R[A/L/R]AA	Arabidopsis	Dehydration responsive element binding proteins-1C and proteins-G in rice	99
[L/R]AAA	Arabidopsis	Auxin response factor-19	100
[H/V/A/D/E/R/Q]LNFP	Rice	Germin-like protein 8-4	102
LDF[S/T]E	Arabidopsis	Putative disease resistance protein At4g19050	103
APWDE	Arabidopsis	Methyl-CpG-binding domain-containing protein 9	104
KYPS	Arabidopsis	Probable methyltransferase PMT18	105
EIDWD	Arabidopsis	Ethylene-responsive transcription factor ERF060, 057, 056, RAP2-13	106

Table S2. AP2/ERF family genes whose biological functions are reported.

Group	Gene	Function	Plant	References
I	BBM BBM1	Vegetative to embryonic growth, somatic embryogenesis, mostly expressed in developing seeds, crown, and lateral roots formation	Arabidopsis, *Brassica napus* (Rape), *Medicago truncatula*	36,37,63
I	WRL1	Required for embryo development, seed germination	Arabidopsis	64
I	ANT	Petal cell identity and mediates down-regulation of AG in flower whorl	Arabidopsis	67
I	ERF	Adaptive responses to flooding stress in rice	Rice	32
I	RAP2.7	Regulates negatively the transition to flowering time	Arabidopsis	33
I	FHP	Floral meristem identity in synergy with LEAFY	Arabidopsis	69
I	RAP2.1	Expressed in flowers, leaves, and stems, and at very low levels in roots	Arabidopsis	70
II	RAV	Negative regulator of plant growth and development	Arabidopsis	30,39
II	RAV	Defense mediated pathway in tomato	Tomato	30
III	RAP2.6	Responsive to stress hormones like jasmonic acid, salicylic acid, abscisic acid and ethylene	Arabidopsis	10
III	ERF1	Disease resistance	Arabidopsis, rice, tobacco,	4,31,71
III	ERF2	Disease resistance	Tobacco	71
III	TSRF1	Disease resistance	Tomato	72
III	ERF3	Up-regulated by feeding of the rice striped stem borer	Rice	78
III	EREBP5	Regulation of gene expression by stress factors and components of stress signal transduction pathways	Arabidopsis, Tobacco	75,76
III	EREBP1	Improved pathogen and abiotic stress tolerance in transgenic plants	Wheat, cotton	73,74
III	EREBP6	Transcription activator. Regulation of gene expression by stress factors	Arabidopsis	12
III	EREBP4	Involved in the regulation of gene expression by stress factors like jasmonate (JA), cold, drought and salt	Arabidopsis, Tobacco	75,77
III	EREBP 9,11,12	Transcriptional inhibitor having EAR motif	Arabidopsis	79
III	EREBP3	Transcriptional repressor, Little expressed in roots and cultured cells	Tobacco, rice	80,81
III	RAP2.11	Strongly induced by low temperatures in shoot tips	Rice	3
IV	CBF like	Cold tolerance	*Brassica napus* (Rape)	83
IV	CBF3	Up regulated in response to low temperature	Tomato	84
IV	DREB1E	Transcriptional activator, play a key role in freezing tolerance and cold acclimation	Arabidopsis, rice	6,88
IV	DREB1F	Transcriptional activator, play a key role in freezing tolerance and cold acclimation	Arabidopsis	6,89
IV	TINY	Transcriptional activator, induced by cold stress	Maize	93
IV	DREB2F	Dehydration and cold	Arabidopsis	6
IV	DREB1A	Transcriptional activator, confers resistance to high cold and drought stress	Arabidopsis, Rice, Maize	85–87
IV	DREB1F	Enhance tolerance to salt, drought and low temperature	Rice, Arabidopsis	6,90
IV	DREB2	Tolerance to water deficit stress	Rice, sun flower	25,82
IV	DREB1C	Transcriptional activator, mediates high salinity- and dehydration-inducible transcription	Arabidopsis, rice	6,88
IV	DREB1I	Mediates high salinity- and dehydration-inducible transcription	Rice	89
IV	DREB2A	Tolerance to osmotic, salt and dehydration stresses	Rice, Arabidopsis	6,91
IV	DREB2B	Involved in drought and heat-shock stress tolerance	Arabidopsis, rice	6,92
IV	DREB2C	Transcriptional activator, mediates high salinity- and abscisic acid-inducible transcription	Arabidopsis, rice	6,92

Abbreviations: ANT, Antigumenta; FHP, Floral homeotic protein; BBM, Baby Boom; WRL1, Wrinkled1; TSRF1, Tomato stress responsive factor1.

Table S3. Full length cDNA ascertained with respect to intron, exon size distribution.

Generic name	MSU (LOC_OS ID)	Average gene length	Average exon length	No. of introns
OsAP2/ERF#022	LOC_Os02g51300.1	1284.0	160.5	7
OsAP2/ERF#153	LOC_Os09g25600.1	792.0	198.0	3
OsAP2/ERF#068	LOC_Os08g07440.1	828.0	118.3	6
OsAP2/ERF#052	LOC_Os05g45954.1	1350.0	270.0	4
OsAP2/ERF#099	LOC_Os06g05340.1	1200.0	200.0	5
OsAP2/ERF#109	LOC_Os08g34360.1	867.0	216.8	3
OsAP2/ERF#079	LOC_Os01g59780.1	825.0	275.0	2
OsAP2/ERF#155	LOC_Os12g03290.1	1332.0	222.0	5
OsAP2/ERF#131	LOC_Os11g03540.1	1350.0	225.0	5
OsAP2/ERF#049	LOC_Os05g32270.1	913.0	152.2	5
OsAP2/ERF#044	LOC_Os04g55970.1	828.0	207.0	3
OsAP2/ERF#086	LOC_Os03g19900.1	1473.0	163.7	8
OsAP2/ERF#088	LOC_Os03g56050.1	2086.0	260.8	7
OsAP2/ERF#141	LOC_Os07g03250.1	945.0	189.0	4
OsAP2/ERF#137	LOC_Os03g12950.1	973.0	324.3	2
OsAP2/ERF#026	LOC_Os03g07940.1	1242.0	177.4	6
OsAP2/ERF#118	LOC_Os11g19060.1	1680.0	210.0	7
OsAP2/ERF#005	LOC_Os01g67410.1	2088.0	261.0	7
OsAP2/ERF#060	LOC_Os06g44750.1	792.0	99.0	7
OsAP2/ERF#018	LOC_Os02g40070.1	2103.0	262.9	7
OsAP2/ERF#037	LOC_Os04g42570.1	1977.0	247.1	7
OsAP2/ERF#092	LOC_Os05g28800.1	543.0	181.0	2
OsAP2/ERF#059	LOC_Os06g43220.1	1005.0	111.7	8
OsAP2/ERF#073	LOC_Os07g13170.1	1311.0	131.1	9
OsAP2/ERF#033	LOC_Os03g60430.1	1293.0	129.3	9
OsAP2/ERF#043	LOC_Os04g55560.1	1383.0	138.3	9
OsAP2/ERF#090	LOC_Os05g03040.2	1539.0	192.4	7
OsAP2/ERF#085	LOC_Os03g06920.1	1260.0	420.0	2
OsAP2/ERF#045	LOC_Os04g56150.1	420.0	210.0	1
Over all length		**1230.4**	**205.3**	
SD		**449.3**	**68.985**	
Minimum length		**781.1**	**138.4**	
Maximum length		**1679.7**	**274.79**	
Base pair used for search of exons on both side of the hit		**2000**		

Table S4. Synteny between rice and wheat AP2/ERF family genes.

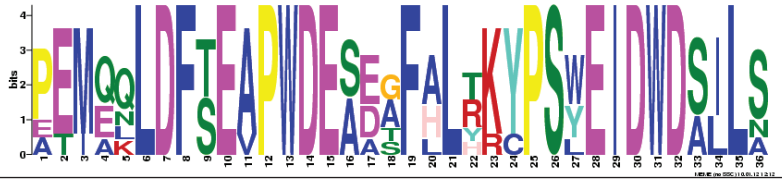
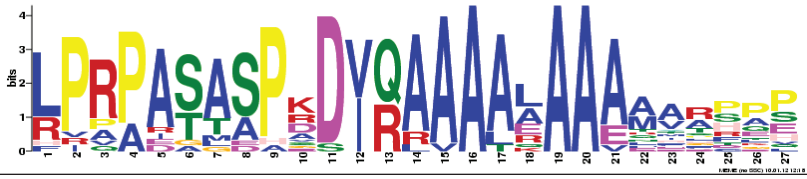
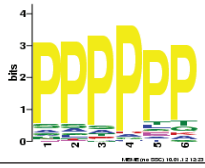
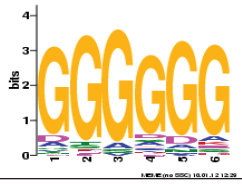
Coded wheat genes	Rice genome alignment				Rice genes (output homology)		Annotation
	Chr_Start	Chr_End	Similarity	Chr	Rice gene	Generic name	
95,107	33765994	33766955	0.606	1	LOC_Os01g58420.1	OsAP2/ERF-4	ERF-3
13,16	1745445	1752065	0.616	1	LOC_Os01g04020.1	OsAP2/ERF-1	RAP2.6
40	11783003	11783631	0.746	1	LOC_Os01g21120.1	OsAP2/ERF-3	EREB3
4,19,90,113	3355682	3357394	0.691	1	LOC_Os01g07120.2	OsAP2/ERF-167	DREB2
10,47,81,100	5688636	5689867	0.728	2	LOC_Os02g10760.1	OsAP2/ERF-12	ERF
12,21,60,103	17569037	17575069	0.768	2	LOC_Os02g29550.2	OsAP2/ERF-13	CHP
24,54	31643379	31644457	0.775	2	LOC_Os02g51670.1	OsAP2/ERF-132	DREB2B
52,72,86,111	33194322	33195272	0.814	2	LOC_Os02g54160.1	OsAP2/ERF-84	EREB1
95	3164706	3165049	0.802	2	LOC_Os02g06330.1	OsAP2/ERF-10	ERF2
1,11,17	34351602	34357234	0.717	3	LOC_Os03g60430.2	OsAP2/ERF-33	FHP
3,11,14	6993827	6995351	0.811	3	LOC_Os03g12950.1	OsAP2/ERF-137	BBM
28,39,62,84,91,110	4347196	4349534	0.633	3	LOC_Os03g08470.1	OsAP2/ERF-125	RAP2.6
29,42,76,101	4366392	4367330	0.636	3	LOC_Os03g08490.1	OsAP2/ERF-28	EREBP
4,90,114	21076935	21077459	0.794	4	LOC_Os04g34970.1	OsAP2/ERF-36	ERF
9	27332155	27332997	0.71	4	LOC_Os04g46400.1	OsAP2/ERF-40	DREB1F
9	27354771	27355159	0.792	4	LOC_Os04g46440.1	OsAP2/ERF-42	ERF
16,51,78,85,105	32845591	32846458	0.651	4	LOC_Os04g55520.1	OsAP2/ERF-128	DREB2F
63	19480884	19483547	0.75	4	LOC_Os04g32620.1	OsAP2/ERF-89	ERF
71	10307892	10308114	0.814	4	LOC_Os04g18650.1	OsAP2/ERF-34	TSRF1
55	24381783	24382150	0.698	5	LOC_Os05g41780.1	OsAP2/ERF-51	ERF-2
31,45,61,94	17187004	17187406	0.746	5	LOC_Os05g29810.1	OsAP2/ERF-93	ERFEB2
5,27,34,37	1433939	1434443	0.737	6	LOC_Os06g03670.1	OsAP2/ERF-98	CBF
10,47,81,100	23897673	23900241	0.77	6	LOC_Os06g40150.1	OsAP2/ERF-57	ERF1
19,52,111	4730450	4736020	0.721	6	LOC_Os06g09390.1	OsAP2/ERF-124	ERF3
68,106	6304249	6304712	0.825	6	LOC_Os06g11860.1	OsAP2/ERF-105	CHP
69	3336324	3336716	0.833	6	LOC_Os06g07030.1	OsAP2/ERF-142	RAP2.6
1,11,17	7544758	7548437	0.679	7	LOC_Os07g13170.1	OsAP2/ERF-73	AP2-like
6,7,8	4175116	4177421	0.811	8	LOC_Os08g07440.1	OsAP2/ERF-68	AP2D
56	27322278	27322451	0.782	8	LOC_Os08g43210.1	OsAP2/ERF-161	CRT/DREBF10
67,82,89,99	23351479	23352053	0.729	8	LOC_Os08g36920.1	OsAP2/ERF-152	CHP
109	19545283	19545935	0.739	8	LOC_Os08g31580.1	OsAP2/ERF-170	DREB
22,23,34,58	15960437	15961590	0.701	9	LOC_Os09g26420.1	OsAP2/ERF-64	ERF2
18,25,26,30,43,48,49,59,65,97,106	20394824	20395662	0.693	9	LOC_Os09g35010.1	OsAP2/ERF-169	CRT/DRE-
33,54,77,109	12212469	12217153	0.686	9	LOC_Os09g20350.1	OsAP2/ERF-113	DREBF1
56,57,73,83,93,102,112	20398977	20399625	0.659	9	LOC_Os09g35020.1	OsAP2/ERF-65	CBF3
67,82,89,99	17305377	17308135	0.713	9	LOC_Os09g28440.1	OsAP2/ERF-123	CHP

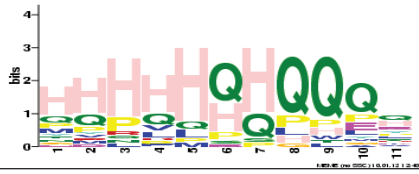
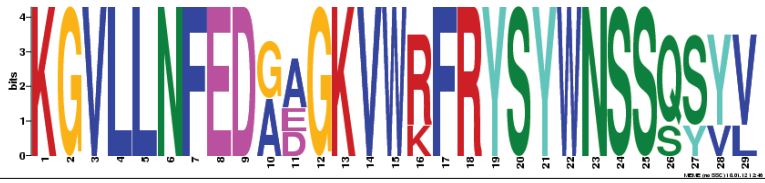
Table S5. Synteny between rice and Arabidopsis AP2/ERF family genes.

Arabidopsis genes	Generic name	Chr	Rice genome alignment				Rice genes (output homology)		Annotation
			Chr_Start	Chr_End	Similarity	Chr	Rice gene	Generic name	
AT1G15360.1	AtAP2/EREBP#001	1	24775898	24776037	0.80	6	LOC_Os06g40150.1	OsAP2/EREBP#057	ERF3
AT1G15360.1	AtAP2/EREBP#002	1	5688851	5691552	0.68	2	LOC_Os02g10760.1	OsAP2/EREBP#012	ERF1
AT2G35700.1	AtAP2/EREBP#003	2	27381026	27381177	0.71	2	LOC_Os02g43940.1	OsAP2/EREBP#021	DREB2B
AT5G11590.1	AtAP2/EREBP#004	5	22622513	22622610	0.76	4	LOC_Os04g36640.1	OsAP2/EREBP#087	TINY
AT5G67190.1	AtAP2/EREBP#005	5	33530573	33530696	0.81	4	LOC_Os04g55520.1	OsAP2/EREBP#128	DREB2F
AT3G14230.1	AtAP2/EREBP#006	3	34056887	34056986	0.87	2	LOC_Os02g54160.1	OsAP2/EREBP#084	EREBP1
AT5G44210.1	AtAP2/EREBP#007	5	20013422	20013492	0.78	4	LOC_Os04g32790.1	OsAP2/EREBP#035	EREBP12
AT1G46768.1	AtAP2/EREBP#008	1	27380978	27381083	0.66	2	LOC_Os02g43940.1	OsAP2/EREBP#021	RAP2.1
AT3G54320.3	AtAP2/EREBP#009	3	1369944	1366400	0.724	11	LOC_Os11g03540.1	OsAP2/EREBP#131	WRL1
AT1G25560.1	AtAP2/EREBP#010	1	27240408	27244907	0.63	5	LOC_Os05g47650.1	OsAP2/EREBP#096	RAV
AT1G79700.2	AtAP2/EREBP#011	1	15361747	15367611	0.747	9	LOC_Os09g25600.1	OsAP2/EREBP#153	AP2.1
AT1G68840.1	AtAP2/EREBP#012	1	28619028	28619859	0.654	1	LOC_Os01g49830.1	OsAP2/EREBP#127	RAV

Supplementary File 1. Conserved motifs discovered using the MEME Suite version 4.7.0.

MOTIF-1	
	WLGT[FY]DTAE[EA]AARAY
MOTIF-2	
	RPWG[KR]W[AV][AS]EIR
MOTIF-3	
	DRAA[LR][RK][LF]RGAxA[VRK][LT]NFPxSAx
MOYIF-4	
	xRxS[KR][YF]RGVR[RQ]
MOTIF-5	
	HRWTGR[YF]EAHLWD[NK][SN]
MOTIF-6	
	AA[LI][KR]Y[WR]GPN[AT]V[TL]NF[PD]
MOTIF-7	
	M[KE][NH]M[TS][RK][EQ]E[FY][VL]ASLRR

MOTIF-8	
	[DE]PRK[KR][TSA]R[VI]
MOTIF-9	
	[YR][LQ]G[GA][YF]D[TK]E[ED][KA]AARAY
MOTIF-10	
	PEM[QE]QLDF[ST]EAPWDE[AS][ED][AG]F[AH]L[RT]KYPS[WY]EIDWD[SA][IL]LS
MOTIF-11	
	[LR]PR[PA]A[ST]A[SA]PKD[VI][QR]AAAA[LA]AAA[AM]ARPPP
MOTIF-12	
	PPPPPP
MOTIF-13	
	GGGGGG
MOTIF-14	
	[RS][RS]G[PF][RS][SR][RG][SA]S[QI]YRGVT

MOTIF-15	
	HHHHH[QH][HQ]QQQx
MOTIF-16	
	KGVLLNFED[AG][ADE]GKVW[RK]FRYSYWNSS[QS][SY][YV][VL]
MOTIF-17	
	[AW]AR[ED][HP]LF[DE]K[AT]VTPSDVGKLNRLV[IV]PKQ[HQ]AE[KR]HFP[LF]
MOTIF-18	
	MCGGAI[LI]AD[LF]IP
MOTIF-19	
	D[DE]D[FW]EA[AD]F[ER]EF[EDL][DSV][DR][DS][DGH]D[DS][DE]D[ED]

Supplementary File 2. AP2/ERF family genes in rice.

This study					Nakano et al[10]
Group	**Generic name**	**MSU (LOC_OS ID)**	**RAP (Os ID)**	**Gene name**	
Ia	AP2/EREBP#022	LOC_Os02g51300.12b	Os02g0747600	BBM1	
Ia	AP2/EREBP#153	LOC_Os09g25600.12b	Os09g0423800	AP2-1	
Ia	AP2/EREBP#068	LOC_Os08g07440.12b	Os08g0171100	WRINKLED1	
Ia	AP2/EREBP#155	LOC_Os12g03290.12b	Os12g0126300	AP2-1	
Ia	AP2/EREBP#131	LOC_Os11g03540.12b	Os11g0129700	AP2-1	
Ia	AP2/EREBP#079	LOC_Os01g59780.12b	Os01g0813300		
Ia	AP2/EREBP#049	LOC_Os05g32270.1	Os05g0389000	AP2-1	
Ia	AP2/EREBP#109	LOC_Os08g34360.12b	Os08g0442400	BBM	
Ia	AP2/EREBP#052	LOC_Os05g45954.12b	Os05g0536250	AP2D4	
Ia	AP2/EREBP#099	LOC_Os06g05340.12b	Os06g0145700		
Ib	AP2/EREBP#044	LOC_Os04g55970.12b	Os04g0653600	ANT	
Ib	AP2/EREBP#086	LOC_Os03g19900.12b	Os03g0313100	ANT	
Ib	AP2/EREBP#088	LOC_Os03g56050.12b	Os03g0770700	ANT	
Ib	AP2/EREBP#137	LOC_Os03g12950.12b	Os03g0232200	BBM	
Ib	AP2/EREBP#141	LOC_Os07g03250.12b	Os07g0124700	PLETHORA	
Ib	AP2/EREBP#026	LOC_Os03g07940.12b	Os03g0176300	WRINKLED1	
Ib	AP2/EREBP#118	LOC_Os11g19060.12b	Os11g0295900	BBM1	
Ib	AP2/EREBP#005	LOC_Os01g67410.12b	Os01g0899800	ERF	
Ib	AP2/EREBP#060	LOC_Os06g44750.12b	Os06g0657500	ANT	
Ib	AP2/EREBP#018	LOC_Os02g40070.1	Os02g0614300	ANT	
Ib	AP2/EREBP#037	LOC_Os04g42570.1	Os04g0504500	ERF	
Ic	AP2/EREBP#092	LOC_Os05g28800.1	Os05g0356201		
Ic	AP2/EREBP#043	LOC_Os04g55560.12b	Os04g0649100	ERF	
Ic	AP2/EREBP#059	LOC_Os06g43220.12b	Os06g0639200	ERF	
Ic	AP2/EREBP#090	LOC_Os05g03040.22b	Os05g0121600	RAP2.7	
Ic	AP2/EREBP#033	LOC_Os03g60430.12b	Os03g0818800	FHP	
Ic	AP2/EREBP#073	LOC_Os07g13170.12b	Os07g0235800	ERF	
II	AP2/EREBP#085	LOC_Os03g06920.1	Os03g0165200		
II	AP2/EREBP#077	LOC_Os01g04750.1-B3	Os01g0140700	RAV	
II	AP2/EREBP#129	LOC_Os01g04800.1-B3	Os01g0141000	RAV	
II	AP2/EREBP#096	LOC_Os05g47650.1-B3	Os05g0549800	RAV	
II	AP2/EREBP#127	LOC_Os01g49830.1-B3	Os01g0693400	RAV	
IIIa	AP2/EREBP#045	LOC_Os04g56150.1	Os04g0655700		
IIIa	AP2/EREBP#089	LOC_Os04g32620.1	Os04g0398000	ERF	Os04g32620
IIIa	AP2/EREBP#023	LOC_Os02g55380.1	Os02g0797100		Os02g55380
IIIa	AP2/EREBP#012	LOC_Os02g10760.1	Os02g0202000	ERF	Os02g10760
IIIa	AP2/EREBP#057	LOC_Os06g40150.1	Os06g0604000	ERF1	Os06g40150
IIIa	AP2/EREBP#159	LOC_Os12g39330.1	Os12g0582900	RAP2.11	Os12g39330
IIIa	AP2/EREBP#075	LOC_Os07g38750.1	Os07g0575000		Os07g38750
IIIa	AP2/EREBP#072	LOC_Os07g10410.1	Os07g0204000		Os07g10410
IIIa	AP2/EREBP#008	LOC_Os12g41030.1	Os12g0602966		Os12g41030
IIIa	AP2/EREBP#009	LOC_Os12g41060.1	Os12g0603300	ERF3	Os12g41060
IIIa	AP2/EREBP#122	LOC_Os09g11460.1	Os09g0286600	ERF	Os09g11460
IIIa	AP2/EREBP#166	LOC_Os09g11480.1	Os09g0287000	EREBP5	Os09g11480
IIIa	AP2/EREBP#164	LOC_Os05g25260.1	Os05g0316800	EREBP9	Os05g25260
IIIa	AP2/EREBP#032	LOC_Os03g60120.1	Os03g0815800	EREBPP5	Os03g60120
IIIa	AP2/EREBP#145	LOC_Os07g12510.1	Os07g0227600	ERF	Os07g12510
IIIa	AP2/EREBP#134	LOC_Os01g12440.1	Os01g0224100		Os01g12440
IIIa	AP2/EREBP#162	LOC_Os01g46870.1	Os01g0657400	EREBP5	Os01g46870
IIIa	AP2/EREBP#146	LOC_Os06g06540.1	Os06g0160500		Os06g06540
IIIa	AP2/EREBP#058	LOC_Os06g42990.1	Os06g0636000	ANT	Os06g42990
IIIa	AP2/EREBP#014	LOC_Os02g32040.1	Os02g0520000		Os02g32040
IIIa	AP2/EREBP#035	LOC_Os04g32790.1	Os04g0399800	EREBP12	Os04g32790

(*Continued*)

Supplementary File 2. (*Continued*)

This study					Nakano et al[10]
Group	**Generic name**	**MSU (LOC_OS ID)**	**RAP (Os ID)**	**Gene name**	
IIIa	AP2/EREBP#069	LOC_Os08g07700.1	Os08g0173700	EREBP4	Os08g07700
IIIa	AP2/EREBP#133	LOC_Os07g47330.1	Os07g0669500	BSL	Os07g47330
IIIa	AP2/EREBP#017	LOC_Os02g38090.1	Os02g0594300	ESR1	Os02g38090
IIIa	AP2/EREBP#149	LOC_Os04g52090.1	Os04g0610400	RAP2.6	Os04g52090
IIIa	AP2/EREBP#051	LOC_Os05g41780.1	Os05g0497300	ERF2	Os05g41780
IIIa	AP2/EREBP#004	LOC_Os01g58420.1	Os01g0797600	ERF3	Os01g58420
IIIa	AP2/EREBP#165	LOC_Os05g41760.1	Os05g0497200	EREBP11	Os05g41760
IIIa	AP2/EREBP#046	LOC_Os04g57340.1	Os04g0669200	ERF2	Os04g57340
IIIa	AP2/EREBP#011	LOC_Os02g09650.1	Os02g0189600	ERF2	Os02g09650
IIIa	AP2/EREBP#061	LOC_Os06g47590.1	Os06g0691100	ERF	Os06g47590
IIIa	AP2/EREBP#010	LOC_Os02g06330.1	Os02g0158000	ERF2	Os02g06330
IIIb	AP2/EREBP#151	LOC_Os05g37640.1	Os05g0448700		Os05g37640
IIIb	AP2/EREBP#062	LOC_Os05g37640.1	Os05g0448675	ERF	
IIIb	AP2/EREBP#067	LOC_Os09g39850.1	Os09g0572000	ERF	Os09g39850
IIIb	AP2/EREBP#106	LOC_Os07g22730.1	Os07g0410300	ERF	Os07g22730
IIIb	AP2/EREBP#074	LOC_Os07g22770.1	Os07g0410700	EREBP1	Os07g22770
IIIb	AP2/EREBP#163	LOC_Os03g64260.1	Os03g0860100	EREBP2	Os03g64260
IIIb	AP2/EREBP#097	LOC_Os05g49010.1	Os05g0564700		Os05g49010
IIIb	AP2/EREBP#168	LOC_Os04g46220.1	Os04g0546800	ERF	Os04g46220
IIIb	AP2/EREBP#147	LOC_Os02g43790.1	Os02g0654700	EREBP5	Os02g43790
IIIb	AP2/EREBP#034	LOC_Os04g18650.1	Os04g0257500	TSRF1	Os04g18650
IIIb	AP2/EREBP#024	LOC_Os03g05590.1	Os03g0150200	EREBP5	Os03g05590
IIIb	AP2/EREBP#154	LOC_Os10g30840.1	Os10g0445100		Os10g30840
IIIb	AP2/EREBP#066	LOC_Os09g39810.1	Os09g0571700		Os09g39810
IIIb	AP2/EREBP#071	LOC_Os08g44960.1	Os08g0563600	ERF	Os08g44960
IIIb	AP2/EREBP#078	LOC_Os01g54890.1	Os01g0752500	EREBP6	Os01g54890
IIIb	AP2/EREBP#116	LOC_Os10g41330.1	Os10g0562900	ERF	Os10g41330
IIIb	AP2/EREBP#020	LOC_Os02g43820.1	Os02g0655200	ERF3	Os02g43820
IIIb	AP2/EREBP#038	LOC_Os04g46240.1	Os04g0547500		Os04g46240
IIIb	AP2/EREBP#039	LOC_Os04g46250.1	Os04g0547600	ERF	Os04g46250
IIIb	AP2/EREBP#027	LOC_Os03g08460.1	Os03g0182800	ERF3	Os03g08460
IIIb	AP2/EREBP#030	LOC_Os03g22170.1	Os03g0341000	RAP2.2	Os03g22170
IIIb	AP2/EREBP#076	LOC_Os07g47790.1	Os07g0674800	EREBP	Os07g47790
IIIb	AP2/EREBP#003	LOC_Os01g21120.1	Os01g0313300	EREBP3	Os01g21120
IIIb	AP2/EREBP#093	LOC_Os05g29810.1	Os05g0361700	EREBP2	Os05g29810
IIIb	AP2/EREBP#125	LOC-Os03g08470.1	Os03g0183000	RAP2.6	Os03g08470
IIIb	AP2/EREBP#107	LOC_Os07g42510.1	Os07g0617000	ERF2	Os07g42510
IIIb	AP2/EREBP#064	LOC_Os09g26420.1	Os09g0434500	ERF2	Os09g26420
IIIb	AP2/EREBP#124	LOC_Os06g09390.1	Os06g0194000	ERF3	Os06g09390
IIIb	AP2/EREBP#084	LOC_Os02g54160.1	Os02g0782700	EREBP1	Os02g54160
IIIb	AP2/EREBP#029	LOC_Os03g08500.1	Os03g0183300	CHP	Os03g08500
IIIb	AP2/EREBP#028	LOC_Os03g08490.1	Os03g0183200	CHP	Os03g08490
IIIb	AP2/EREBP#006	LOC_Os10g25170.1	Os10g0390800	EREBP3	Os10g25170
IIIc	AP2/EREBP#063	LOC_Os09g13940.1	Os09g0309700	TINY	Os09g13940
IIIc	AP2/EREBP#108	LOC_Os08g27220.1	Os08g0360800		Os08g27220
IIIc	AP2/EREBP#110	LOC_Os08g41030.1	Os08g0521600	ERF	Os08g41030
IIIc	AP2/EREBP#015	LOC_Os02g32140.1	Os02g0521100		Os02g32140
IIIc	AP2/EREBP#001	LOC_Os01g04020.1	Os01g0131600	RAP2.6	Os01g04020
IIIc	AP2/EREBP#111	LOC_Os08g42550.1	Os08g0537900	ERF3	Os08g42550
IIIc	AP2/EREBP#152	LOC_Os08g36920.1	Os08g0474000	RAP2.6	Os08g36920
IIIc	AP2/EREBP#123	LOC_Os09g28440.1	Os09g0457900	EREBPP LIKE	Os09g28440
IIIc	AP2/EREBP#130	LOC_Os02g52670.1	Os02g0764700	EREBP4	Os02g52670

(*Continued*)

Supplementary File 2. (*Continued*)

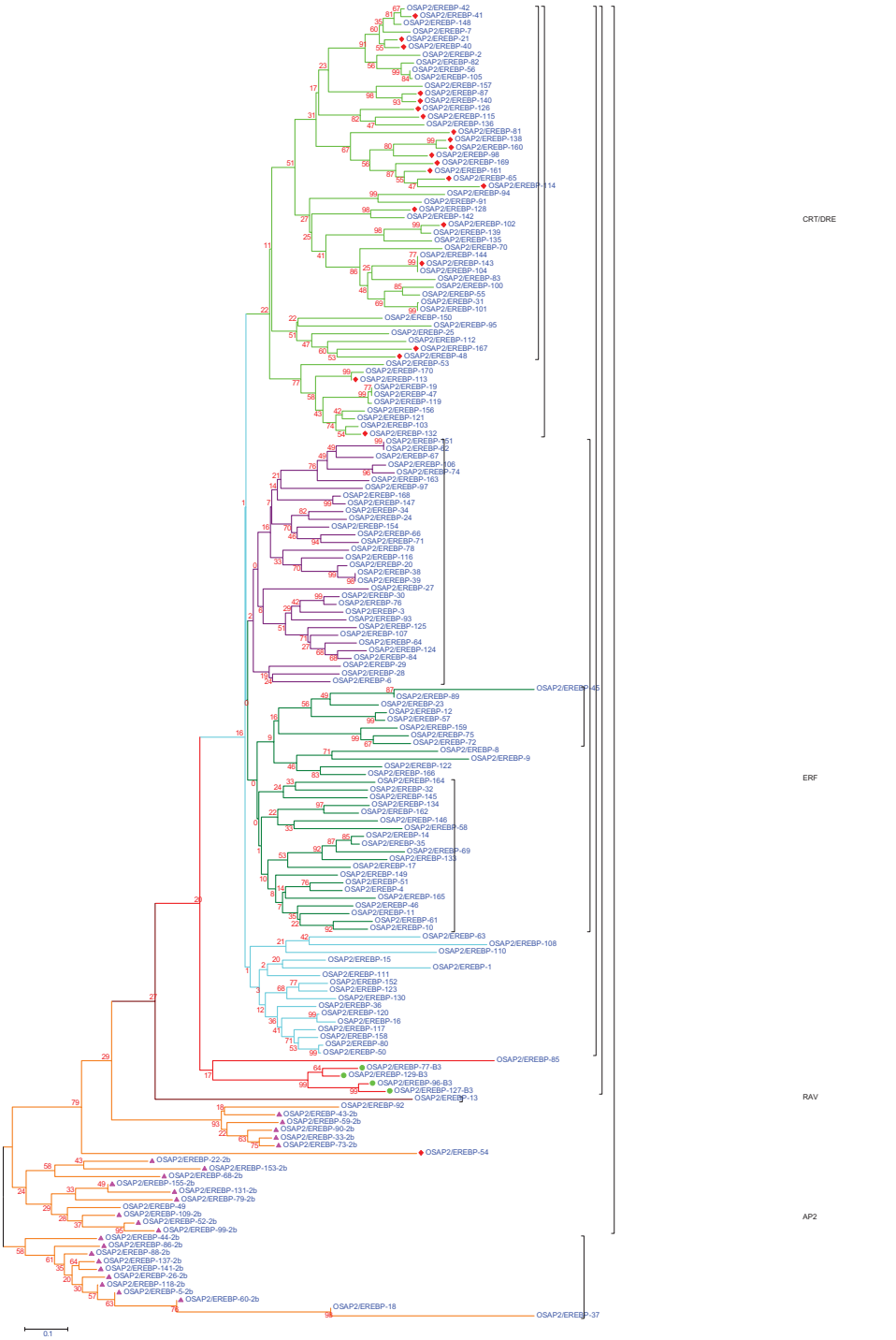
This study					Nakano et al[10]
Group	**Generic name**	**MSU (LOC_OS ID)**	**RAP (Os ID)**	**Gene name**	
IIIc	AP2/EREBP#036	LOC_Os04g34970.1	Os04g0429050		Os04g34970
IIIc	AP2/EREBP#120	LOC_Os02g34260.1	Os02g0546600	ERF	Os02g34260
IIIc	AP2/EREBP#016	LOC_Os02g34270.1	Os02g0546800	ERF	Os02g34270
IIIc	AP2/EREBP#117	LOC_Os11g06770.1	Os110168500	RAP2.6	Os11g06770
IIIc	AP2/EREBP#158	LOC_Os12g07030.1	Os12g0168100	RAP2.6	Os12g07030
IIIc	AP2/EREBP#080	LOC_Os01g64790.1	Os01g0868000	ERF	Os01g64790
IIIc	AP2/EREBP#050	LOC_Os05g36100.1	Os05g36100	ERF	Os05g36100
IV-a	AP2/EREBP#053	LOC_Os05g49700.1	Os05g0572000	CHP	Os05g49700
IV-a	AP2/EREBP#170	LOC_Os08g31580.1	Os08g0408500	ERF	Os08g31580
IV-a	AP2/EREBP#113	LOC_Os09g20350.1	Os09g0369000	DREBF1	Os09g20350
IV-a	AP2/EREBP#019	LOC_Os02g42585.1	Os02g0638650	ERF	
IVa	AP2/EREBP#047	AP004768.3			AP004768.3
IV-a	AP2/EREBP#119	LOC_Os04g44670.1	Os04g0529100	ERF	Os04g44670
IV-a	AP2/EREBP#156	LOC_Os10g22600.1	Os10g0371100	ERF	Os10g22600
IV-a	AP2/EREBP#121	LOC_Os03g09170.1	Os03g0191900	ERF	Os03g09170
IV-a	AP2/EREBP#103	LOC_Os06g09810.1	Os06g0198400		Os06g09810
IV-a	AP2/EREBP#132	LOC_Os02g51670.1	Os02g0752800	DREB2B	Os02g51670
IV-b	AP2/EREBP#150	LOC_Os05g28350.1	Os05g0351200	ERF	Os05g28350
IV-b	AP2/EREBP#095	LOC_Os05g39590.1	Os05g0473300	ERF1	Os05g39590
IV-b	AP2/EREBP#025	LOC_Os03g07830.1	Os03g0174400	CHP	Os03g07830
IV-b	AP2/EREBP#112	LOC_Os08g45110.1	Os08g0565200	ERF	Os08g45110
IV-b	AP2/EREBP#167	LOC_Os01g07120.1	Os01g0165000	DREB2	Os01g07120
IV-b	AP2/EREBP#048	LOC_Os05g27930.1	Os05g0346200	DRE	Os05g27930
IV-c	AP2/EREBP#094	LOC_Os05g34730.1	Os05g0420300		Os05g34730
IVc	AP2/EREBP#091	LOC_Os01g66270.1	Os01g0885900		Os01g66270
IV-c	AP2/EREBP#128	LOC_Os04g55520.1	Os04g0648900	DREB2F	Os04g55520
IV-c	AP2/EREBP#142	LOC_Os06g07030.1	Os06g0166400	TINY	Os06g07030
IV-c	AP2/EREBP#102	LOC_Os06g09790.1	Os06g0198200	CBF LIKE	Os06g09790
IV-c	AP2/EREBP#139	-------	Os02g0767800		
IV-c	AP2/EREBP#135	LOC_Os03g15660.1	Os03g0263000	TINY	Os03g15660
IV-c	AP2/EREBP#070	LOC_Os08g35240.1	Os08g0454000		Os08g35240
IV-c	AP2/EREBP#144	LOC_Os04g48330.1	Os04g0572200	ERF027	
IV-c	AP2/EREBP#143	LOC_Os06g11940.1	Os06g0223100	DREB	Os06g11940
IV-c	AP2/EREBP#104	LOC_Os06g10780.1	Os06g0210300	ERF	Os06g10780
IV-c	AP2/EREBP#083	LOC_Os02g54050.1	Os02g0781300		Os02g54050
IV-c	AP2/EREBP#100	LOC_Os06g08340.1	Os06g0181700	ERF	Os06g08340
IV-c	AP2/EREBP#055	LOC_Os06g09760.1	Os06g197900		Os06g09760
IVc	AP2/EREBP#031	AP006056	------		AP006056
IV-c	AP2/EREBP#101	LOC_Os06g09717.1	Os06g0197200		
IV-d	AP2/EREBP#042	LOC_Os04g46440.1	Os04g0550200		Os04g46440
IV-d	AP2/EREBP#041	LOC_Os04g46410.1	Os04g0549800	DREB1C	Os04g46410
IV-d	AP2/EREBP#148	LOC_Os02g43970.1	Os02g0657000	RAP211	Os02g43970
IV-d	AP2/EREBP#007	LOC_Os10g41130.1	Os10g0560700	TINY	Os10g41130
IV-d	AP2/EREBP#021	LOC_Os02g43940.1	Os02g0656600	DREB2B	Os02g43940
IV-d	AP2/EREBP#040	LOC_Os04g46400.1	Os04g0549700	DREB1F	Os04g46400
IV-d	AP2/EREBP#002	LOC_Os01g10370.1	Os01g0200600		Os01g10370
IV-d	AP2/EREBP#082	LOC_Os02g13710.1	Os02g0231000	TINY	Os02g13710
IV-d	AP2/EREBP#056	LOC_Os06g36000.1	Os06g0553700		Os06g36000
IV-d	AP2/EREBP#105	LOC_Os06g11860.1	Os06g0222400	CHP	Os06g11860
IV-d	AP2/EREBP#157	LOC_Os11g13840.1	Os11g0242300		Os11g13840
IV-d	AP2/EREBP#087	LOC_Os04g36640.1	Os04g0443800	DREB	Os04g36640
IV-d	AP2/EREBP#140	LOC_Os02g35240.1	Os02g0558700	DREB	Os02g35240

(*Continued*)

Supplementary File 2. (*Continued*)

This study					Nakano et al[10]
Group	**Generic name**	**MSU (LOC_OS ID)**	**RAP (Os ID)**	**Gene name**	
IV-d	AP2/EREBP#126	LOC_Os02g45420.1	Os02g0676800	DREB1E	Os02g45420
IV-d	AP2/EREBP#115	LOC_Os10g38000.1	Os10g0523900	DREB	Os10g38000
IV-d	AP2/EREBP#136	LOC_Os03g02650.1	Os03g0117900		
IV-d	AP2/EREBP#081	LOC_Os01g73770.1	Os01g0968800	DREB1F	Os01g73770
IV-d	AP2/EREBP#138	LOC_Os02g45450.1	Os02g0677300	CRT/DREB1	Os02g45450
IV-d	AP2/EREBP#160	LOC_Os04g48350.1	Os04g0572400	CRT/DREB1	Os04g48350
IV-d	AP2/EREBP#098	LOC_Os06g03670.1	Os06g0127100	CBF	Os06g03670
IV-d	AP2/EREBP#169	LOC_Os09g35010.1	Os09g0522000	CBF	
IV-d	AP2/EREBP#161	LOC_Os08g43210.1	Os08g0545500	DREB1I	Os08g43210
IV-d	AP2/EREBP#065	LOC_Os09g35020.1	Os09g0522100	CBF3	Os09g35020
IV-d	AP2/EREBP#114	LOC_Os09g35030.1	Os09g0522200	DREB1A	Os09g35030
Soloist	AP2/EREBP#054	LOC_Os06g06970.1	Os06g0165600	CRT/DRE	
	AP2/EREBP#013	LOC_Os02g29550.1	Os02g0499000	CHP	
		Total = 170			

Note: **The OS RAP (ID) number is mentioned because OsAP2/ERF#139 does not have the MSU locus identifier.


CRT/DRE
ERF
RAV
AP2
0.1