Genome-Wide Analysis of the ERF Gene Family in Arabidopsis and Rice[W]

Toshitsugu Nakano[1], Kaoru Suzuki[1], Tatsuhito Fujimura, and Hideaki Shinshi*

Molecular and Cellular Breeding Research Group, Institute for Biological Resources and Functions, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki 305–8566, Japan (T.N., K.S., H.S.); and Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki 305–8672, Japan (T.N., T.F.)

Genes in the ERF family encode transcriptional regulators with a variety of functions involved in the developmental and physiological processes in plants. In this study, a comprehensive computational analysis identified 122 and 139 ERF family genes in Arabidopsis (*Arabidopsis thaliana*) and rice (*Oryza sativa* L. subsp. *japonica*), respectively. A complete overview of this gene family in Arabidopsis is presented, including the gene structures, phylogeny, chromosome locations, and conserved motifs. In addition, a comparative analysis between these genes in Arabidopsis and rice was performed. As a result of these analyses, the ERF families in Arabidopsis and rice were divided into 12 and 15 groups, respectively, and several of these groups were further divided into subgroups. Based on the observation that 11 of these groups were present in both Arabidopsis and rice, it was concluded that the major functional diversification within the ERF family predated the monocot/dicot divergence. In contrast, some groups/subgroups are species specific. We discuss the relationship between the structure and function of the ERF family proteins based on these results and published information. It was further concluded that the expansion of the ERF family in plants might have been due to chromosomal/segmental duplication and tandem duplication, as well as more ancient transposition and homing. These results will be useful for future functional analyses of the ERF family genes.

The ERF family is a large gene family of transcription factors and is part of the AP2/ERF superfamily, which also contains the AP2 and RAV families (Riechmann et al., 2000). The AP2/ERF superfamily is defined by the AP2/ERF domain, which consists of about 60 to 70 amino acids and is involved in DNA binding. These three families have been defined as follows. The AP2 family proteins contain two repeated AP2/ERF domains, the ERF family proteins contain a single AP2/ERF domain, and the RAV family proteins contain a B3 domain, which is a DNA-binding domain conserved in other plant-specific transcription factors, including VP1/ABI3, in addition to the single AP2/ERF domain. The ERF family is sometimes further divided into two major subfamilies, the ERF subfamily and the CBF/DREB subfamily (Sakuma et al., 2002). The AP2 domain was first identified as a repeated motif within the Arabidopsis (*Arabidopsis thaliana*) AP2 protein, which is involved in flower development (Jofuku et al., 1994). The ERF domain was first identified as a conserved motif in four DNA-binding proteins from tobacco (*Nicotiana tabacum*), namely, ethylene-responsive element-binding proteins 1, 2, 3, and 4 (EREBP1, 2, 3, and 4,

currently renamed ERF1, 2, 3, and 4), and was shown to specifically bind to a GCC box, which is a DNA sequence involved in the ethylene-responsive transcription of genes (Ohme-Takagi and Shinshi, 1995). In the case of the RAV family, *RAV1* and *RAV2* were first identified as full-length cDNAs encoding proteins that contain a B3-like domain and an AP2/ERF domain in Arabidopsis (Kagaya et al., 1999).

It has been demonstrated that the AP2/ERF proteins have important functions in the transcriptional regulation of a variety of biological processes related to growth and development, as well as various responses to environmental stimuli. Genes in the AP2 family have been shown to participate in the regulation of developmental processes, e.g. flower development (Elliott et al., 1996), spikelet meristem determinacy (Chuck et al., 1998), leaf epidermal cell identity (Moose and Sisco, 1996), and embryo development (Boutilier et al., 2002). Recently, the involvement of members of the RAV family in ethylene response (Alonso et al., 2003) and in brassino-steroid response (Hu et al., 2004) was reported. After finding the tobacco ERFs (Ohme-Takagi and Shinshi, 1995), many proteins in the ERF family were identified and implicated in many diverse functions in cellular processes, such as hormonal signal transduction (Ohme-Takagi and Shinshi, 1995), response to biotic (Yamamoto et al., 1999; Gu et al., 2000) and abiotic stresses (Stockinger et al., 1997; Liu et al., 1998; Dubouzet et al., 2003), and regulation of metabolism (van der Fits and Memelink, 2000; Aharoni et al., 2004; Broun et al., 2004; Zhang et al., 2005), and in developmental processes (van der Graaff et al., 2000; Banno et al., 2001; Chuck et al., 2002) in various plant species.

After the sequencing of the Arabidopsis genome was completed (Arabidopsis Genome Initiative, 2000), 145 genes were postulated to encode proteins containing the AP2/ERF domain, with 83% (121 genes) of these genes belonging to the ERF family (Sakuma et al., 2002). To date, most of the members of the ERF family have yet to be studied, despite the likelihood that these genes play important roles in many physiological aspects in plants. A great deal of experimental work will be required to determine the specific biological function of each of these genes. On the basis of phylogenetic analyses, it has become apparent that a large gene family of transcription factors consists of subgroups of genes that are closely related to each other (Kranz et al., 1998; Pãrenicova et al., 2003; Toledo-Ortiz et al., 2003; Reyes et al., 2004; Tian et al., 2004). A functional analysis of each transcription factor belonging to the ERF family should be done, taking into account functional redundancy. As a part of this process, an assessment of the structural relationships between all Arabidopsis ERF family proteins would provide a guide for predicting the functions of genes, which remains to be studied in this family. Moreover, the current availability of the rice (*Oryza sativa*) genome sequences also allows a comparative analysis between Arabidopsis and rice within the ERF family, which is useful in terms of studying the functional and evolutional diversity of the transcription factor family in plants.

In this study, the establishment of a complete picture of the ERF gene family in Arabidopsis was attempted. To this end, genes in the AP2/ERF superfamily in the Arabidopsis genome were surveyed again, resulting in the identification of 147 genes in this superfamily, including 122 genes in the ERF family. Phylogenetic analyses were performed, as well as exon/intron and protein motif structural analyses of the ERF family genes. Genes encoding proteins in the ERF family in rice genomic and cDNA databases were also surveyed, and comparative analyses of the phylogeny and conserved motifs in the rice and Arabidopsis ERF families were performed. The resulting classification of groups and identification of putative functional motifs will be useful in studies on the biological functions of each gene in the ERF families.

RESULTS AND DISCUSSION

Identification of the ERF Family Genes in Arabidopsis

To identify the ERF family genes in Arabidopsis, BLAST (Altschul et al., 1990) searches of the Arabidopsis databases were performed using the AP2/ERF domain (59 amino acids) of the tobacco ERF2 protein as a query sequence. One hundred forty-seven genes were identified as possibly encoding AP2/ERF domain(s) (Table I). The individual genes are listed in Supplemental Tables I and II. Fourteen genes were predicted to encode proteins containing two AP2/ERF domains. Six genes were predicted to encode one AP2/ERF domain together with one B3 domain. Based on these results, the former and the latter genes were assigned to the AP2 and the RAV families, respectively. One hundred twenty-seven genes were predicted to encode proteins containing a single AP2/ERF domain. One hundred twenty-two of these 127 genes were assigned to the ERF family. The remaining four genes, At2g41710, At2g39250, At3g54990, and At5g60120, also encode an AP2/ERF domain, but are distinct from the ERF type and are instead more closely related to that of the AP2 type. For this reason, these genes were assigned to a subclass of the AP2 family (Supplemental Table I). The remaining gene, At4g13040, includes an AP2/ERF-like domain sequence, but its homology appears quite low in comparison with the other AP2/ERF genes. Therefore, this gene was designated as a soloist.

Two reports indicated the number of genes in the AP2/ERF superfamily in Arabidopsis. Riechmann et al. (2000) proposed the existence of 124 ERF family genes, 14 AP2 family genes, and six RAV family genes. However, they did not present any information regarding the specifics of the individual genes. After this report, Sakuma et al. (2002) reported 145 genes that are classified as members of the AP2/ERF superfamily. Their classification process did not indicate locus

Table I. *Summary of the AP2/ERF superfamily*
Total for each family is shown in bold.

	This Study			Sakuma et al. (2002)	
Classification	Group	No.	Classification	Subgroup	No.
AP2 family		**18**	AP2 subfamily		**17**
	Double AP2/ERF domain	14		Double AP2/ERF domain	14
	Single AP2/ERF domain	4		Single AP2/ERF domain	3
ERF family		**122**	DREB, ERF subfamily		**121**
	Groups I to IV	57	DREB subfamily	A-1 to 6	56
	Groups V to X	58	ERF subfamily	B-1 to 6	65
	Groups VI-L and Xb-L	7		B-6	
At4g13040		**1**	AL079349		**1**
RAV family		**6**	RAV subfamily		**6**
	Total	**147**		Total	**145**

identifiers (such as the Arabidopsis Genome Initiative [AGI] code) or accession numbers for the individual genes and/or cDNAs. Of these genes, 121 were classified as part of the ERF subfamily and CBF/DREB subfamily, 17 were classified as part of the AP2 family, six were classified as part of the RAV family, and one remaining gene, AL079349, was unclassified. The gene AL079349 seems to be identical to the soloist gene At4g13040 in this study. Our BLAST search also included a new gene, At5g60120, to a group of proteins containing a single AP2-type AP2/ERF domain. This group includes three other genes, At2g41710, At2g39250, and At3g54990, which might be identical to the genes AC002339, AC004697, and AL132970, respectively, reported by Sakuma et al. (2002). All of the members in the CBF/DREB subfamily (group A) and in the ERF subfamily (group B) described by Sakuma would be included in the ERF family in this study (Table I; Supplemental Table II). In addition, our BLAST search included a gene, At1g22190, in the ERF family. Taken together, it was concluded that the Arabidopsis AP2/ERF superfamily is composed of 147 genes: 146 divided into three families, the ERF family (122 genes), the AP2 family (18 genes), and the RAV family (six genes), and a soloist gene, At4g13040, as shown in Table I.

Given the above classification, the 122 genes of the ERF family were subjected to further analyses. A generic name (AtERF#001–AtERF#122) was provisionally given to distinguish each gene (Supplemental Table II), to avoid confusion in this study. This numbering system provides a unique identifier for each ERF gene as pro-

posed for the MYB, WRKY, bZIP, and bHLH transcription factors in Arabidopsis (Kranz et al., 1998; Romero et al., 1998; Eulgem et al., 2000; Jakoby et al., 2002; Heim et al., 2003). For the genes named in previous publications, the definitive names were put with their generic name. This numbering system was also used to distinguish each rice ERF gene (Supplemental Table III).

Phylogenetic Relationships between the ERF Family Genes in Arabidopsis

To clarify the phylogenetic relationships between the genes in the Arabidopsis ERF family, multiple alignment analyses were performed using amino acid sequences of the AP2/ERF domain. The alignment indicated that the residues Gly-4, Arg-6, Glu-16, Trp-28, Leu-29, Gly-30, and Ala-38 are completely conserved among the 122 proteins in the ERF family (Supplemental Fig. 1). In addition, more than 95% of the ERF family members contain Arg-8, Gly-11, Ile-17, Arg-18, Arg-26, Ala-39, Asp-43, and Asn-57 residues (Supplemental Fig. 1). These observations are generally consistent with earlier reports on this topic (Riechmann and Meyerowitz, 1998; Sakuma et al., 2002). However, this same alignment indicated that the C-terminal regions of the AP2/ERF domains of seven proteins (AtERF#116–AtERF#122) possess a very low homology to the consensus sequence (Fig. 1). This region corresponds to the C-terminal half of the α-helix (Allen et al., 1998), which includes the highly conserved residues Asp-43 and Asn-57.

Figure 1. Alignment of the AP2/ERF domains from Arabidopsis ERF proteins. Black and light gray shading indicate identical and conserved amino acid residues, respectively. Dark gray shading indicates conserved amino acid residues in group VI-L or group Xb-L. The black bar and arrows represent predicted α-helix and β-sheet regions, respectively, within the AP2/ERF domain (Allen et al., 1998). Asterisks represent amino acid residues that directly make contact with DNA (Allen et al., 1998).

Based on these observations, a phylogenetic tree based on the alignment of the AP2/ERF domains of 115 of the AtERF proteins (excluding the seven proteins AtERF#116–AtERF#122) was constructed. As shown in Figure 2, this phylogram distinguished 10 groups, namely, groups I to X, rather than the two major subfamilies used previously, ERF (group B) and CBF/DREB (group A). Although the bootstrapping values for the nodes corresponding to the 10 groups were not high in every case (data not shown), the reliability of this clustering was supported by the presence and position of an intron and the common amino acid sequence motifs outside of the AP2/ERF domain, as described later. The members of group V and the seven proteins (AtERF#116–AtERF#122) were classified into a group, B-6, by Sakuma et al. (2002). Based on the alignment of amino acid sequences of AP2/ERF domains, however, these seven proteins are distinct from the other members of the ERF family, including the group V, as mentioned above (Fig. 1). In addition, as will be discussed later, a motif analysis indicated that

AtERF#116 to #119 and AtERF#120 to #122 are related to the genes in groups VI and X, respectively. Therefore, AtERF#116 to #119 and AtERF#120 to #122 were classified into groups VI-like (VI-L) and Xb-like (Xb-L), respectively. In this analysis, AtERF#021 to AtERF#043 were branched into a single clade (group III), which includes members of groups A-1, A-4, and A-5 (Sakuma et al., 2002). On the other hand, Sakuma et al. (2002) classified AtERF#052 (ABI4), AtERF#022, and AtERF#109 into groups A-3, A-5 (identical to group II), and B-3 (identical to group IX), respectively. However, the phylogenetic relationships and the common amino acid sequence motifs suggested that AtERF#052 (ABI4), AtERF#022, and AtERF#109 should be classified into group IV (identical to A-2), group III (including A-4 and A-1), and group X (identical to B-4), respectively (Fig. 3). Taking these into consideration, the final result of these analyses was the classification of the Arabidopsis ERF family into 12 groups, groups I to X, VI-L, and Xb-L (Fig. 3; Supplemental Table II). The relationship between the present classification and the
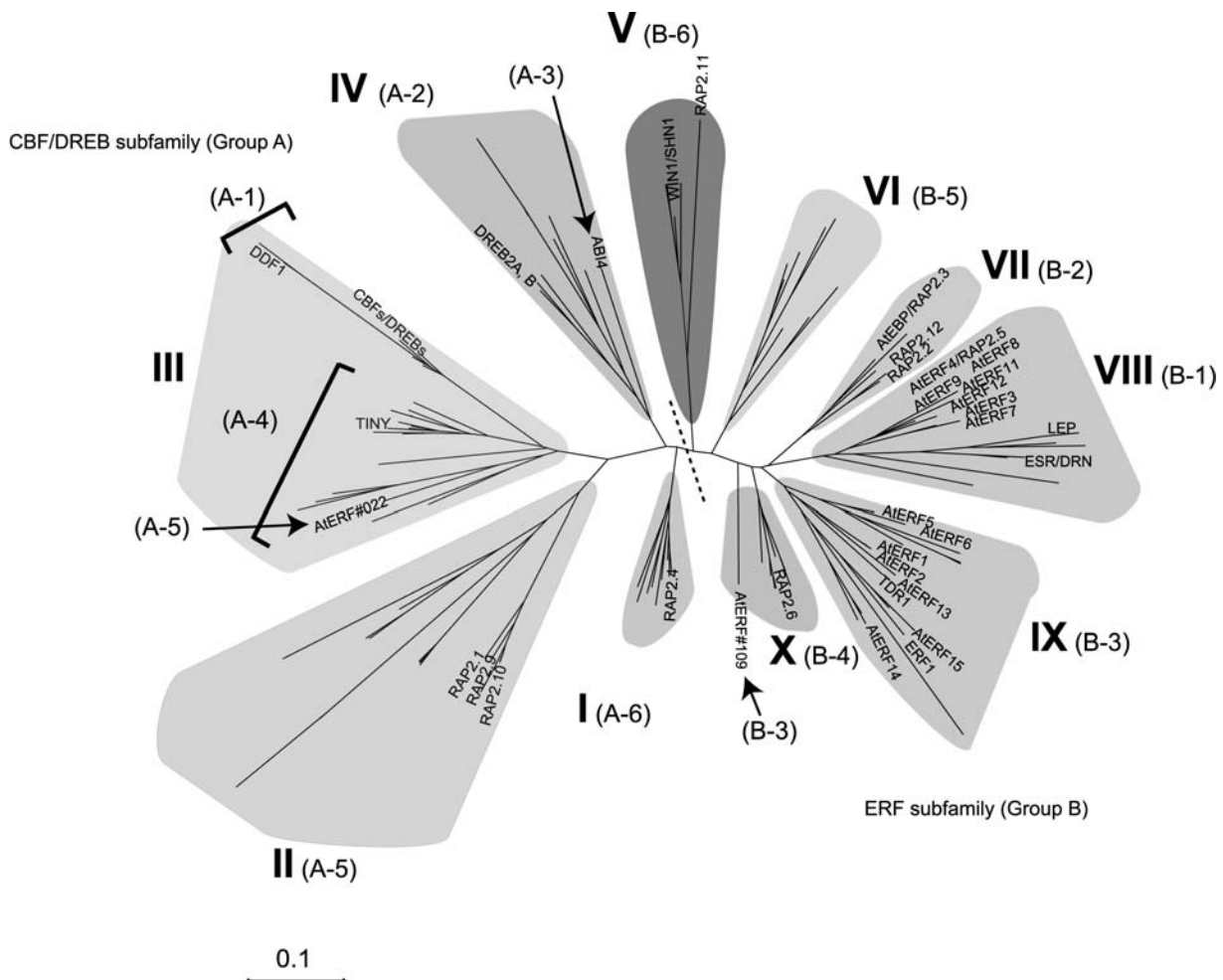
Figure 2. An unrooted phylogenetic tree of Arabidopsis ERF proteins. The amino acid sequences of the AP2/ERF domain, except members of group VI-L and Xb-L, were aligned by ClustalW (Supplemental Fig. 1), and the phylogenetic tree was constructed using the NJ method. The names of the ERF genes that have already been reported are indicated. The so-called CBF/DREB and ERF subfamilies are divided with a broken line. Classification by Sakuma et al. (2002) is indicated in parentheses.
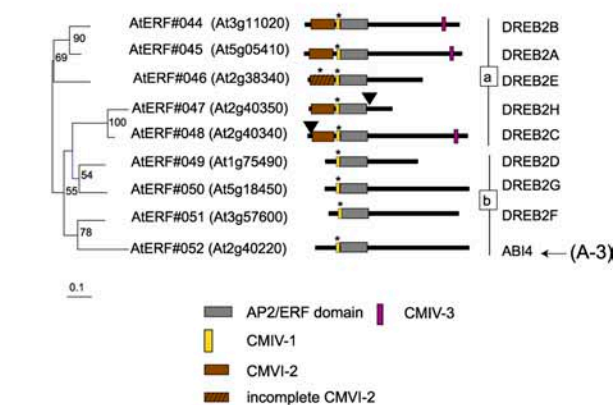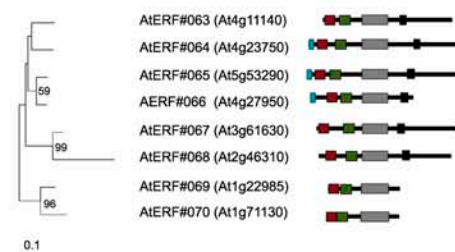
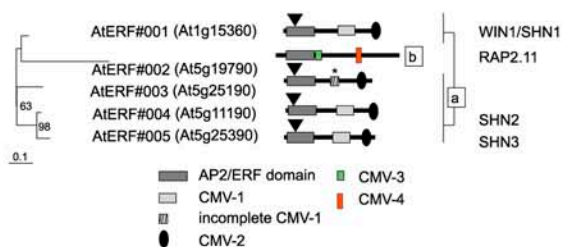# A Group I (A-6)
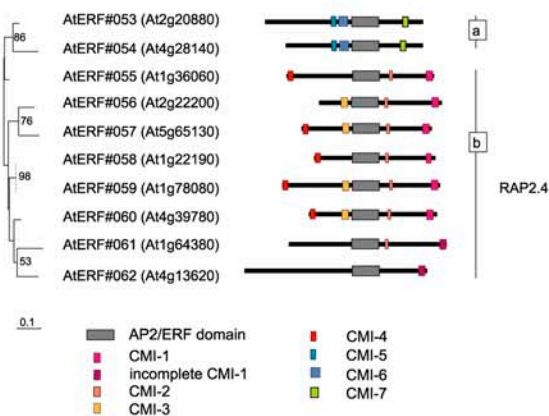
B Group II (A-5)
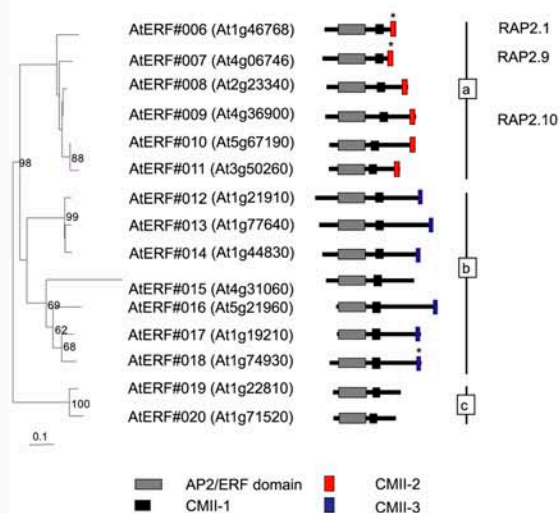
C Group III
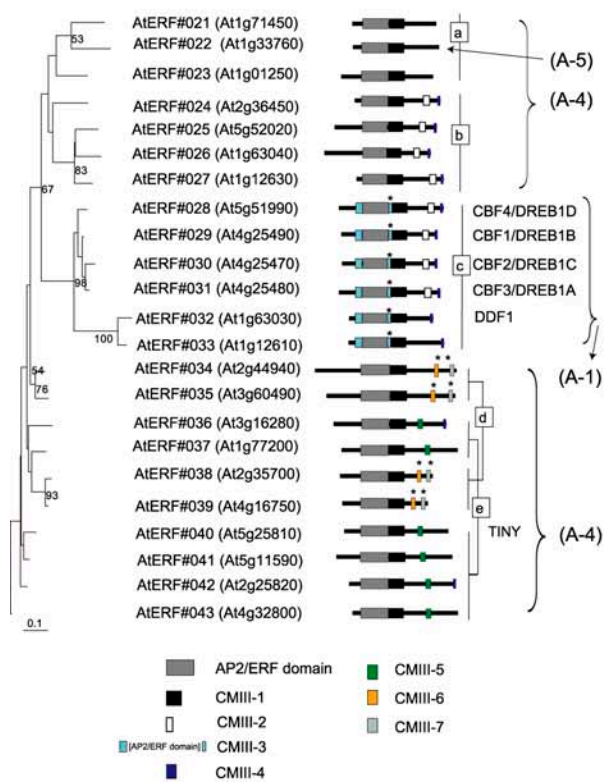
D Group IV (A-2)
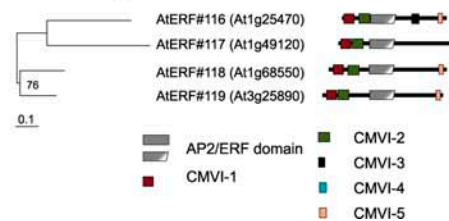

F Group VI (B-5)

# E Group V (B-6)

G Group VI-L (B-6)

Figure 3. Phylogenetic relationships among the Arabidopsis ERF genes, from group I (A), group II (B), group III (C), group IV (D), group V (E), group VI (F), group VI-L (G), group VII (H), group VIII (I), group IX (J), group X (K), and group Xb-L (L) in the Arabidopsis ERF family. Bootstrap values from 100 replicates were used to assess the robustness of the trees. Bootstrap values >50 are
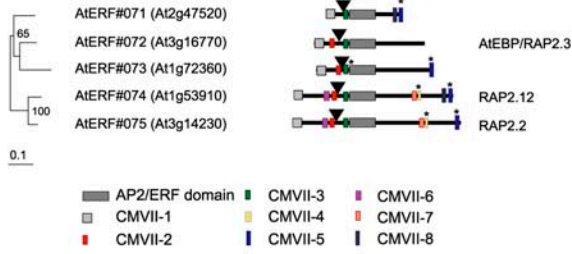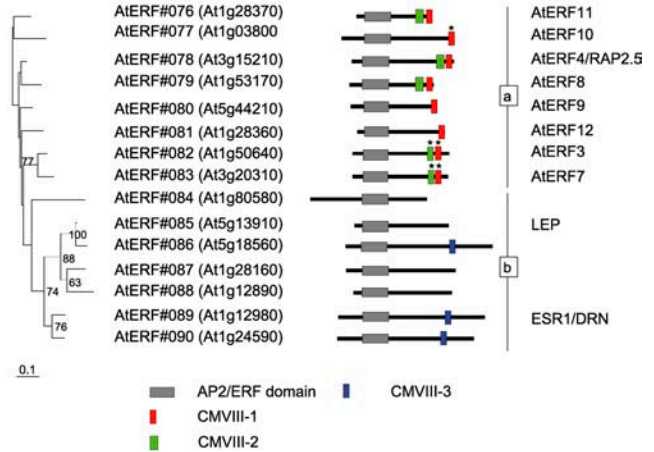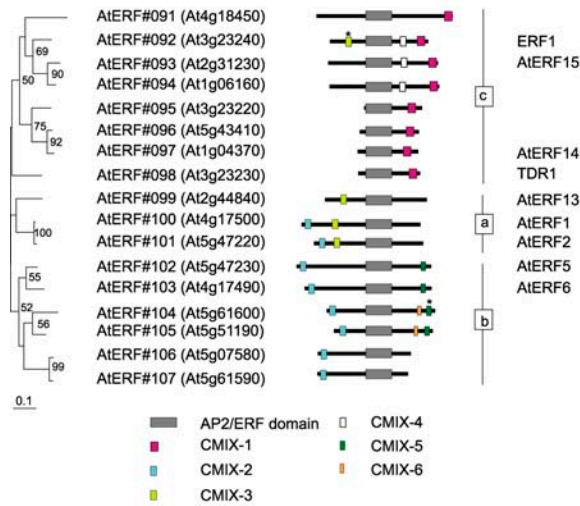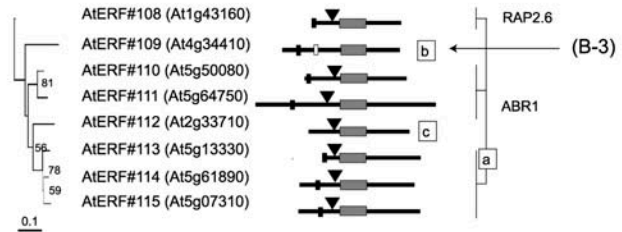
Figure 3. (*Continued.*)
shown. The phylogenetic tree, location of the intron (arrowhead), and a schematic diagram of the protein structures of every group, I to VI, VI to L, VII to X, and Xb-L, are shown in A to L, respectively. Each colored box represents the AP2/ERF domain and conserved motifs, as indicated below the tree. The amino acid sequences of the conserved motifs are summarized in Supplemental Table IV. The asterisk indicates that these motifs were defined by multiple alignments with manual correction rather than an MEME search. Classification by Sakuma et al. (2002) is indicated in parentheses.

previous classification by Sakuma et al. (2002) is indicated in Figures 2 and 3, Table II, and Supplemental Table II.

The Relationship between Gene Structure and Phylogenetic Classification

It was reported previously that most of the genes in the ERF family of Arabidopsis possess no introns and only four of these genes have an intron (Sakuma et al., 2002). These four genes likely correspond to four of the

five genes in group V (Fig. 3E). In this study, 20 genes, including these four genes, were found to contain a single intron in their open reading frame regions (Fig. 3, D, E, H, K, and L). As shown in Figure 3, E, H, K, and L, most of genes in groups V, VII, X, and Xb-L contain a single intron, with the position of the intron being conserved in each group. This further validates the classification of the ERF family genes of Arabidopsis in this study. In addition, groups V and X were further classified into two subgroups based on the existence of an intron.

Table II. *Comparison of group/subgroup size between Arabidopsis and rice ERF families*

Group	Subgroup	AtERF Genes	OsERF Genes
I (A-6)[a]	a	2	2
	b	8	7
II (A-5)	a	6	2
	b	7	11
	c	2	2
III (A-1, -4, -5)	a	3	–
	b	4	6
	c	6	11
	d	4	7
	e	6	2
IV (A-2, -3)	a	5	1
	b	4	5
V (B-6)	a	4	5
	b	1	3
VI (B-5)	–	8	6
VI-L (B-6)	–	4	3
VII (B-2)	a	5	14
	b	–	1
VIII (B-1)	a[b]	8	8
	b[b]	7	5
IX (B-3)	a[c]	3	3
	b[c]	6	4
	c[c]	8	11
X (B-3, -4)	a	6	7
	b	1	3
	c	1	3
Xb-L (B-6)	–	3	–
XI	–	–	4
XII	–	–	1
XIII	–	–	1
XIV	–	–	1
Total		122	139

[a]Classification by Sakuma et al. (2002). [b]Subclassification by McGrath et al. (2005). [c]Subclassification by Gutterson and Reuber (2004).

Conserved Motifs outside of the AP2/ERF Domain

Generally, regions outside the DNA-binding domain in transcription factors contain functionally important domains involved in transcriptional activity, protein-protein interactions, and nuclear localization (Liu et al., 1999). Such functional domains, or amino acid sequence motifs, are often conserved among members of a subgroup in large families of transcription factors in plants, such as MYB, WRKY, NAC, Dof, GATA, and GRAS (Kranz et al., 1998; Eulgem et al., 2000; Lijavetzky et al., 2003; Ooka et al., 2003; Reyes et al., 2004; Tian et al., 2004). Proteins within a subgroup that share these motifs are likely to share similar functions.

An investigation of the conserved motifs in the proteins of each group in the ERF family of Arabidopsis was carried out via a multiple alignment analysis with ClustalW (Thompson et al., 1994). The conserved motifs found in the AtERF family are summarized in Supplemental Table IV. Most of the motifs are selectively distributed among the specific clades in the

phylogenetic tree, demonstrating structural similarities among proteins within the same group (Fig. 3). Based on the conservation of these motifs, most of the groups in the ERF family can be further divided into several distinct subgroups (Fig. 3). Although the functions of most of these conserved motifs have not been investigated, it is plausible that some may play important roles in transcriptional regulation.

For instance, the motif analysis revealed the motif CMVIII-1 in group VIIIa (Figs. 3I and 4A). This motif is identical to an ERF-associated amphiphilic repression (EAR) motif, which has been shown to function as a repression domain (Fujimoto et al., 2000; Ohta et al., 2001). The EAR motif was identified as a conserved sequence, (L/F)DLN(L/F)xP, in the C-terminal regions of the repressor-type ERF proteins and also in TFIIIA-type C_2H_2 zinc-finger proteins (Ohta et al., 2001). Recently, it was reported that similar motifs, LxLxL within domain I in AUX/IAA proteins and AtERF4 (AtERF#078; Tiwari et al., 2004) and xLxLxL within the repression domains in SUPERMAN and its related TFIIIA-type C_2H_2 zinc-finger proteins (Hiratsu et al., 2004), are important in conferring repression. Interestingly, it was found that the CMII-2 motif conserved in group IIa is composed of a DLNxxP sequence that is also conserved in the CMVIII-1/EAR motif (Fig. 4B). The conservation of the DLNxxP sequence in these groups suggests it may function in transcriptional regulation. A recent study showed that a novel B3 domain repressor protein harbors the DLNxxP motif in the C-terminal region and that the mutation of the motif reduced the repression activity of the protein (Tsukagoshi et al., 2005).

Regions of acidic amino acid-rich, Gln-rich, Pro-rich, and/or Ser/Thr-rich amino acid sequences are often designated as transcriptional activation domains (Liu et al., 1999). Most of the conserved motifs identified in this study have such features in their amino acid compositions (Supplemental Table IV), whereas the functions of these motifs have not been rigorously demonstrated.

A unique motif, CMX-2, is conserved in the N-terminal region of the proteins in groups Xb and Xb-L (Fig. 3, K and L). This motif contains a characteristic consensus sequence, $CX_2CX_4CX_{2\sim4}C$ (Fig. 5). The Cys repeat feature, which may be a zinc-finger motif, suggests its function may be in binding to DNA or in protein-protein interactions, although there is no experimental evidence at this time.

Several motifs related to putative phosphorylation sites are conserved in proteins in groups VI, VII, and IXb (Fig. 6). The functional site prediction tool (see "Materials and Methods") predicted that mitogen-activated protein (MAP) kinase and/or casein kinase I may potentially phosphorylate the CMVI-3/SP(T/V) SVL motif (Fig. 6A) in group VI. Interestingly, Patmatch, a peptide pattern search tool available at The Arabidopsis Information Resource (TAIR) Web site (http://www.arabidopsis.org/cgi-bin/patmatch/nph-patmatch.pl), showed that the SP(T/V)SVL motif

Nakano et al.

Figure 4. The EAR motif-like sequences conserved in the C-terminal region of subgroup VIIIa and subgroup IIa ERF proteins. A, An alignment of the sequences of the C-terminal regions of subgroup VIIIa proteins. B, An alignment of the sequences of the C-terminal regions of subgroup IIA proteins. The conserved motifs are underlined. Black and gray shading indicate identical and conserved amino acid residues present in more than 50% of the aligned sequences, respectively. Consensus amino acid residues are given below the alignment. The "x" in the sequence indicates no conservation at this position. Bold letters in the sequence represent conserved amino acid residues in the original EAR motif (Ohta et al., 2001). Asterisks indicate proteins with demonstrated repression activity (Fujimoto et al., 2000; Ohta et al., 2001).

is conserved within another 33 putative proteins, some of which are MYB-like transcription factors and GATA transcription factors (data not shown). Putative MAP kinase phosphorylation sites are conserved in the proteins of groups VII and IXb (Fig. 6, B and C).

Characteristics of Each Group in the Arabidopsis ERF Gene Family

The characteristics of each group in the Arabidopsis ERF family are described below. For reference, the current knowledge regarding the functions of the genes in the ERF family is summarized in Table III.

Group I

Group I was divided into two subgroups, Ia and Ib (Fig. 3A; Table II). At this time, the functions of these genes are unknown. DBF1 from maize (*Zea mays*) has been shown to activate the drought-responsive element 2 (DRE2)-dependent transcription of ABA-responsive *rab17* in transiently transformed maize callus (Kizis and Pages, 2002). Since DBF1 contains CMI-1 and CMI-2 motifs, it is a member of subgroup Ib in maize (data not shown). Recently, the overexpression of *Medicago truncatula WXP1* has been shown to activate wax production in transgenic alfalfa (*Medicago sativa*; Zhang et al., 2005). The WXP1 protein contains all conserved motifs identified in subgroup Ib, with

Figure 5. A putative zinc-finger motif conserved in group Xb and group Xb-L proteins. Black and gray shading indicate identical and conserved amino acid residues present in more than 50% of the aligned sequences, respectively. Consensus amino acid residues are given below the alignment. The "x" indicates no conservation at this position.

A   Subgroup VI/CMVI-3

```
AtERF#063   TENVEEA-------STEVKGVSD-FIIG---GGECLRSPVSVLESPFSGESTAVKEEFVGV
AtERF#064   KKKKKKN-------NKSKKSVTASSSISRSSSNDCLCSPVSVLRSPFA-----V-DEFSGI
AtERF#065   IEEKPVI-------MTTPTPTTSSSESTEE-DLQHLSSPTSVLNHR---SEE-IQQVQQPF
AtERF#066   QEQE----------QEPESNMSVSISESMD-DSQHLSSPTSVLNYQTYVSEEPIDSLIKPV
AtERF#067   ------T-------PVIDLQTVSACDYGRD-SRQSLHSPTSVLRFN---VNEETEHEIEAI
AtERF#068   TEVLPET-------PVIDLETVSGCDSARE-SQISLCSPTSVLRFS---HNDETEYRTEPT
OsERF#053   VPPEVAERVPQPPGASKNASPATSYDSGEE-SHAAASPTSVLRSFPP-SAVVATADTANK
OsERF#054   DP----------------VAVGYESGAE-SSPAVSSPTSVLRKVP--S-LCSLAEDKDD
OsERF#055   PSRATYA------------GAASGYDESSD-ESQLVGSPVSVLRPMP--ARATAKKEAKEA
OsERF#056   ASVPDPTPAPPEEDRANDS----S-------FNP-FASPTSVLRYAAADEVTAPS------
AtERF#116   --------------------------VVTVTSQCLRSPTSASVSCVSADDLSKEK---TS
OsERF#108   TPT---------------PTASSSSSTSAAPFAHPSPSSVLEATKPAPKPESPPLPEQA
                                          ++
```

B   Subgroup VII/CMVII-4

```
OsERF#071            NSFGCSDFSQ-ENDIKTPDITSMLAP--TMTGVDDSAFLQNNASDAM-------VPP--VM
OsERF#072            NSFGSSDFGW-EDDTLTPDYTSVFVPNAAMPAYGEPAYLTGGAPKRMRNNYGIAVPQGNGM
OsERF#070 (OsEREBP1) NSFGCSDFGW-ENDTKTPDITSIAPIS-TIAEVDESAFIKS-STNPM-------VPP--VM
AtERF#074            NSFDCSEFGWSDQAPITPDISSAVIN-----NNNSALFFEEANPAKK-------LK-----
AtERF#075            NSFDCSEFGWSDHGPKTPEISSMLVN-----NNE-ASFVEETNAAKK-------LKP---N
                                    ++
```

C   Subgroup IXb/CMIX-5 and 6

```
AtERF#102 (AtERF5)   -------PFVTSNLTELCDWDLTG-------------FLNFPLLSPLSPHPPFGYSQLTVV..
AtERF#103 (AtERF6)   ----------VESGLTAIDDWDLTE-------------FLSMPLLSPLSPHPPFGYPQLTVV..
OsERF#096            --------------------AAAAAGDIWDELKGICSLPPLSPLSPHPHMAFPQLFVIDL
AtERF#104            ---PVVPLTPSSWMGFWDVGAGD-------------GIFSIPPLSPTSPNFSVISVT......
AtERF#105            -----CPLTPSSWKGFWDGADSKDM-----------GIFSVPLLSPCPSLGHSQLVVT.....
OsERF#094            AAAESSPATPSSWSWEQY---------------WEALLGGLPPLSPLSPHPALGFPQLTVN..
OsERF#095            -----TSPSPSSEAAACGGGGGELFVPPMPSSWSWDQLEGFFGILSPLSPHPQMGFPEVAVN..
OsERF#097            EPTSFPVVTPSSWSWDQYWD--------------GLPPLSPLSPHPALGFPQLTVS..
                            ++                          ++ ++
```

Figure 6. Putative protein kinase phosphorylation sites conserved in ERF proteins. A, Putative MAP kinase and/or casein kinase I phosphorylation sites conserved in group VI and VI-L proteins. B, Putative MAP kinase phosphorylation sites conserved in group VII proteins. C, Putative MAP kinase phosphorylation sites conserved in subgroup IXb proteins. The conserved motifs are underlined. Black and gray shading indicate identical and conserved amino acid residues present in more than 50% of the aligned sequences, respectively. The "+" indicates an amino acid matching the motif patterns provided by the ELM (http://elm.eu.org/browse.html).

the closest related protein in Arabidopsis being AtERF#059.

Group II

Group II consists of three subgroups, IIa, IIb, and IIc (Fig. 3B; Table II). All of the genes in this group contain the CMII-1 motif in the C-terminal region adjacent to the AP2/ERF domain. This motif is similar to the CMIII-1 motif found in group III (Supplemental Table IV). In addition, subgroups IIa and IIb, but not IIc, contain additional motifs at the C terminus. The proteins in group IIa contain a CMII-2 motif that is similar to the EAR motif as described above. The members of group IIb, except AtERF#015, have the CMII-3 motif at the C terminus. This motif is also found in subgroups IIIb, IIIc, AtERF#036 of IIId, and AtERF#042 of IIIe, in subgroup III, and group VII (except AtERF#072), as described below.

Group III

The proteins in group III commonly contain a CMIII-1 motif that is similar to the CMII-1 motif conserved in proteins in group II. Possession of these motifs and the other phylogenetic relationships suggest a strong similarity between groups II and III. Based on other conserved motifs and phylogeny, group III was divided into five subgroups, IIIa to IIIe (Fig. 3C).

Subgroups IIIb and IIIc contain two consensus motifs, CMIII-2 and CMIII-4, in the C-terminal region (Fig. 3C). The CMIII-4 motif has also been reported as a LWSY motif conserved in OsDREB1A, B, and C (OsERF#024, #031, and #026) and in CBF3/DREB1A (AtERF#031; Dubouzet et al., 2003). In addition, there are highly conserved regions on both sides of the AP2/ERF domain in the proteins in subgroup IIIc, with both regions collectively referred to as the CMIII-3 motif. The presence of conserved sequences in these two regions of this motif, PKK/RPAGRxKFxETRHP (region I) and DSAWR (region II), were reported previously in CBFs/DREB1s of Arabidopsis and their homologs of several other plant species (Jaglo et al., 2001; Haake et al., 2002). The remaining genes in group III were divided into two additional subgroups, IIId and IIIe (Fig. 3C). In the phylogenetic tree in Figure 3C, AtERF#036, AtERF#037, AtERF#038, and AtERF#039 are branched into a single clade. However, AtERF#038, AtERF#039, AtERF#034, and AtERF#035 were assigned to subgroup IIId because these proteins commonly contain the CMIII-6 and CMIII-7 motifs. Similarly, AtERF#036 and AtERF#037 were assigned to subgroup IIIe based on the presence of the CMIII-5 motif in these proteins, which is conserved in subgroup IIIe.

The functions of the genes in subgroup IIIc have been studied extensively. These genes have been shown to play crucial roles in low-temperature-, salt-, and/or drought-stress-responsive gene expression (Gilmour

Table III. *ERF genes whose biological function has been reported*

Group	Genes	Functions	Method	Species[a]	References[b]
Ia	WXP1	Wax accumulation	Overexpression	Mt	1
IIIc	CBF1 to 4/DREB1A to D	Freezing, drought, salt tolerance	Overexpression	At	2, 3, 4
	DDF1	Salt tolerance, GA biosynthesis regulation	Activation tagging	At	5
IIIe	TINY	Growth regulation	Activation tagging	At	6
IVb	ABI4	Abscisic acid response, sugar signaling	Knockout mutant	At	7, 8, 9
Va	WIN1/SHNs	Wax accumulation	Overexpression	At	10, 11
VI	Pti6	Disease resistance	Overexpression	Le	12, 13
	Tsi1	Salt tolerance, disease resistance	Overexpression	Nt	14
	CaERFLP1	Salt tolerance, disease resistance	Overexpression	Ca	15
VII	JERF3	Salt tolerance	Overexpression	Le	16
	CaPF1	Freezing tolerance, disease resistance	Overexpression	Ca	17
VIIIa	AtERF4	Ethylene, jasmonic acid, and abscisic acid response	Overexpression, knockout mutant	At	18, 19
	AtERF7	Abscisic acid response	Overexpression, RNAi	At	20
VIIIb	ESR1/DRN	Organ identity	Activation tagging	At	21, 22
	BD1	Floral meristem identity	Knockout mutant	Zm	23
	FZP	Floral meristem identity	Knockout mutant	Os	24
	LEP	Leaf petiole development	Activation tagging	At	25
IXc	ERF1	Disease resistance	Overexpression	At	26, 27
	Pti5	Disease resistance	Overexpression	Le	13, 28
	NtERF5	Disease resistance	Overexpression	Nt	29
	TERF1	Salt tolerance	Overexpression	Le	30
IXa	ORCA3	Indole alkaloid biosynthesis	Activation tagging	Cr	31
	OPBP1	Salt tolerance, disease resistance	Overexpression	Nt	32
	Pti4	Disease resistance	Overexpression	Le	12, 13
Xa	ABR1	Abscisic acid response	Knockout mutant	At	33

[a]Prefixes At, Ca, Cr, Le, Mt, Nt, Os, and Zm are *Arabidopsis thaliana, Capsicum annuum, Catharanthus roseus, Lycopersicon esculentum, Medicago truncatula, Nicotiana tabacum, Oryza sativa,* and *Zea mays,* respectively.   [b]1, Zhang et al. (2005); 2, Liu et al. (1998); 3, Gilmour et al. (2000); 4, Haake et al. (2002); 5, Magome et al. (2004); 6, Wilson et al. (1996); 7, Finkelstein et al. (1998); 8, Arenas-Huertero et al. (2000); 9, Huijser et al. (2000); 10, Aharoni et al. (2004); 11, Broun et al. (2004); 12, Zhou et al. (1997); 13, Gu et al. (2002); 14, Park et al. (2001); 15, Lee et al. (2004); 16, Wang et al. (2004); 17, Yi et al. (2004); 18, Yang et al. (2005); 19, McGrath et al. (2005); 20, Song et al. (2005); 21, Banno et al. (2001); 22, Kirch et al. (2003); 23, Chuck et al. (2002); 24, Komatsu et al. (2003); 25, van der Graaff et al. (2000); 26, Solano et al. (1998); 27, Berrocal-Lobo et al. (2002); 28, He et al. (2001); 29, Fischer and Droge-Laser (2004); 30, Huang et al. (2004); 31, van der Fits and Memelink (2000); 32, Guo et al. (2004); 33, Pandey et al. (2005).

et al., 1998; Liu et al., 1998; Haake et al., 2002; Dubouzet et al., 2003; Magome et al., 2004). Recently, the C-terminal region of 98 amino acids of CBF1/DREB1B (AtERF#029) was shown to function as a transactivation domain (Wang et al., 2005). This region includes CMIII-2 and CMIII-4. Although the functions of the subgroup IIIb proteins are unknown, these proteins may also function as transcriptional activators in gene expression as a response to abiotic stress based on the conservation of the acidic amino acid-rich regions, which is also a feature in the proteins in subgroup IIIc. Maize DRE-binding factor, DBF2 (Kizis and Pages, 2002), is a homolog of subgroup IIId, sharing conserved motifs as well as similarity in the AP2/ERF domain. Arabidopsis *TINY* (AtERF#040) belongs to subgroup IIIe.

Group IV

Group IV was divided into two subgroups, IVa and IVb (Fig. 3D). High homology is present throughout the N-terminal region outside the AP2/ERF domain of AtERF#044 (DREB2B), AtERF#045 (DREB2A), AtERF#046 (DREB2E), AtERF#047 (DREB2H), and AtERF#048 (DREB2C), which were assigned to subgroup IVa. This conserved region is divided into two blocks, referred to as motifs CMIV-1 and CMIV-2 (Fig. 3D). The CMIV-2 motif includes a putative nuclear localization signal (Liu et al., 1998). The genes for AtERF#047 and AtERF#048 contain a single intron at the N-terminal and C-terminal halves of the protein, respectively, as shown in Figure 3D. DREB2A (AtERF#045) and DREB2B (AtERF#044) were identified as transcription factors involved in DRE-mediated transcription (Liu et al., 1998). ORCA1 (Menke et al., 1999) and OsDREB2A (OsERF#040; Dubouzet et al., 2003) belong to subgroup IVa in *Catharanthus roseus* and rice, respectively. The CMIV-1 motif is completely conserved in the proteins in group IV (Fig. 3D). Therefore, it is reasonable to assign AtERF#049, AtERF#050, and AtERF#051, which correspond to the DREB2-related proteins DREB2D, DREB2G, and DREB2F, respectively (Sakuma et al., 2002), and AtERF#052 (ABI4) to the same subgroup, namely, subgroup IVb. ABI4 has been shown to be involved in germination-related ABA signaling (Finkelstein et al., 1998) and sugar response (Arenas-Huertero et al., 2000; Huijser et al., 2000).

Group V

Group V consists of two subgroups, Va and Vb (Fig. 3E; Table II). The four genes in subgroup Va are closely related to each other, sharing two motifs, CMV-1 and CMV-2, in the C-terminal regions. AtERF#003 contains CMV-2 and part of CMV-1. Only a single gene, AtERF#002, which does not contain these motifs, was assigned to subgroup Vb. Two motifs, CMV-3 and CMV-4, were identified in AtERF#002 through comparison with the rice ERF genes in subgroup Vb (Supplemental Table III). Recently, two research groups showed that the overexpression of *WIN1/SHN1* (AtERF#001) results in the enhanced accumulation of epidermal wax (Aharoni et al., 2004; Broun et al., 2004). These authors showed that SHN2 (AtERF#004) and SHN3 (AtERF#005) shared a similar function with WIN1/SHN1 (AtERF#001; Aharoni et al., 2004; Broun et al., 2004). Aharoni et al. (2004) also predicted that these three ERF proteins would have two conserved motifs corresponding to motifs CMV-1 and CMV-2, respectively. Their preliminary results also showed that the overexpression of AtERF#003 (At5g25190) did not result in the typical morphological *shn* phenotype (Aharoni et al., 2004). This might be due to the partial CMV-1 motif in AtERF#003. Thus, the results of two studies (Aharoni et al., 2004; Broun et al., 2004) are consistent with the results of our phylogenetic study and motif analysis. There is no information regarding the function of AtERF#002. In total, these results support our concept that the assessment of the structural relationships between all Arabidopsis ERF family proteins should provide information that assists in predicting the functions of unknown genes.

Group VI

Group VI consists of proteins that share two conserved motifs, CMVI-1 and CMVI-2, in the N-terminal region (Fig. 3F; Supplemental Table II). The C-terminal regions of AtERF#069 and AtERF#070 were shorter than the others (AtERF#063–AtERF#068), which shared the CMVI-3 motif in the C-terminal region. The tobacco Tsi1 (Park et al., 2001) and tomato (*Lycopersicon esculentum*) Pti6 proteins (Zhou et al., 1997) exhibit characteristic features of group VI. Tsi1 and Pti6 have been shown to play a role in abiotic and/or biotic stress-responsive gene expression (Zhou et al., 1997; Park et al., 2001).

Group VI-L

As previously described, proteins encoded by the genes AtERF#116 to #119 are characterized by their imperfect AP2/ERF domain. Since these proteins all have two conserved motifs, CMVI-1 and CMVI-2, characteristic features of group VI, these genes were classified as group VI-L. For AtERF#116, the Munich Information Center for Protein Sequences Arabidopsis Database (MAtDB) predicts a sequence of 364 amino acids including two introns in the coding region. However, TAIR and The Institute for Genomic Research (TIGR) predict a protein of 287 amino acids with no introns in the coding region. In the latter cases, two introns are located within the 5′-untransrated region of the gene. Because one full-length cDNA (RAFL21-49-G19) matched the gene annotation given by TAIR and TIGR, a 287-amino acid sequence was used for the analyses.

Group VII

A characteristic feature of group VII is a MCGGAI(I/L) motif (Tournier et al., 2003) referred to as CMVII-1 (Fig. 3H). Two additional motifs, CMVII-2 and CMVII-3, were also identified (Fig. 3H). In addition, AtERF#074 and AtERF#075 contain another motif, CMVII-4 (Fig. 3H). *AtEBP* (AtERF#072) was the first gene identified in this group (Büttner and Singh, 1997). All of the genes in this group have a single intron in the 5′-flanking region of the AP2/ERF domain (Fig. 3H). Close inspection of the sequences indicated that an LWS(I/L/Y) sequence, designated as the CMVII-5 motif, was retained at the C terminus in this group. The CMVII-5 motif is similar to the CMII-3 and CMIII-4 motifs.

It was found that the At1g72360 locus, which was assigned to AtERF#073, is differently annotated in MAtDB and TAIR/TIGR. MAtDB and TAIR/TIGR predicted that At1g72360 would encode a protein of 262 and 211 amino acids, respectively. In the latter case, the predicted protein lacks an N-terminal CMVII-1 motif. A cDNA (BT002063), corresponding to At1g72360, encodes a sequence of 262 amino acids. Given that this agrees with the MAtDB result, this sequence was used in this study.

AtEBP (AtERF#072) has been shown to interact in vitro with OBF4, a bZIP transcription factor, although the functional importance of this interaction is unknown (Büttner and Singh, 1997). In the case of rice OsEBP89 (OsERF#70), this gene interacts with OsBP-5, a Myc transcription factor, and coregulates the expression of the *waxy* gene via a 31-bp cis-acting sequence (Zhu et al., 2003).

Group VIII

Group VIII consists of two subgroups, VIIIa and VIIIb (Fig. 3I; Table II). The predicted proteins of subgroup VIIIa have a conserved motif, CMVIII-1, at the C terminus. As mentioned above, this motif has also been designated as the EAR motif (Ohta et al., 2001). AtERF#076, AtERF#078, AtERF#079, AtERF#082, and AtERF#083 contain another motif, CMVIII-2. The ERFs of tobacco, Arabidopsis, and rice, which contain the CMVIII-1 and CMVIII-2 motifs, have been shown to repress GCC box-mediated transcription via a transient assay (Fujimoto et al., 2000; Ohta et al., 2000, 2001).

Recently, AtERF4 (AtERF#078) was shown to be a negative regulator in the expression of ethylene-, jasmonate-, and ABA-responsive genes (McGrath et al., 2005; Yang et al., 2005). In addition, AtERF7 (AtERF#083) was shown to play an important role in ABA response in plants (Song et al., 2005).

The remaining genes of group VIII were assigned to the subgroup VIIIb (Fig. 3I). Of these genes, AtERF#086, #089, and #090 contain the CMVIII-3 motif in the C-terminal region. This group includes genes such as *LEP* (AtERF#085; van der Graaff et al., 2000) and *ESR1/DRN* (AtERF#089; Banno et al., 2001; Kirch et al., 2003) that are involved in the differentiation and development of organs.

Group IX

Group IX consists of three subgroups, IXa to IXc (Table II). Generally, these subgroups, IXa, IXb, and IXc, are characterized by the motifs CMIX-3, CMIX-2, and CMIX-1, respectively (Fig. 3J).

The subgroup IXc is made up of eight genes. The predicted amino acid sequences of AtERF#095, AtERF#096, AtERF#097, and AtERF#098 are relatively small in length, ranging from 128 to 139 residues. AtERF#092 (ERF1), AtERF#093 (AtERF15), and AtERF#094 also contain a CMIX-4 motif. AtERF#092 (ERF1) alone contains a CMIX-3 motif, which was not detected by the MEME program (Bailey and Elkan, 1994), which discovers conserved motifs within given data set. This motif is also conserved in putative orthologs of AtERF#092 (ERF1) in other plant species, including tobacco S25XP1, tomato Pti5, and rice OsERF#091 (data not shown). AtERF#100 (AtERF1) and AtERF#101 (AtERF2) of subgroup IXa also share a CMIX-2 motif, which is a characteristic feature of subgroup IXb. The CMIX-2 and CMIX-3 motifs are putative acidic regions that might function as transcriptional activation domains (Fujimoto et al., 2000). The CMIX-3 motif corresponds to a conserved sequence that was previously referred to as a 24-amino acid (DMLV) motif (Gutterson and Reuber, 2004).

Tobacco ERF2 and ERF4 were assigned to subgroups IXa and IXb, respectively. The N-terminal regions of tobacco ERF2 and ERF4 have been shown to contain possible transactivation domains (Ohta et al., 2000).

It has been suggested that tobacco ERF4 (Ohta et al., 2000) and AtERF5 (AtERF#102; Fujimoto et al., 2000) contain a putative MAP kinase phosphorylation site in the C-terminal region. This site is designated as a CMIX-5 motif in this study, and was found in AtERF#103 (AtERF6), AtERF#104, and AtERF#105, as well as AtERF#102 (AtERF5; Figs. 3J and 6C). In addition, it was found that AtERF#104 and AtERF#105 have an additional putative MAP kinase phosphorylation site, which was designated as a CMIX-6 motif (Fig. 6C). In contrast, AtERF#106 and AtERF#107 have no MAP kinase phosphorylation sites. Thus, the genes

in subgroup IXb were classified into three types based on their putative MAP kinase phosphorylation site constitution. This suggests that the three types of ERF genes share roles in transcriptional regulation in response to distinct extracellular signals.
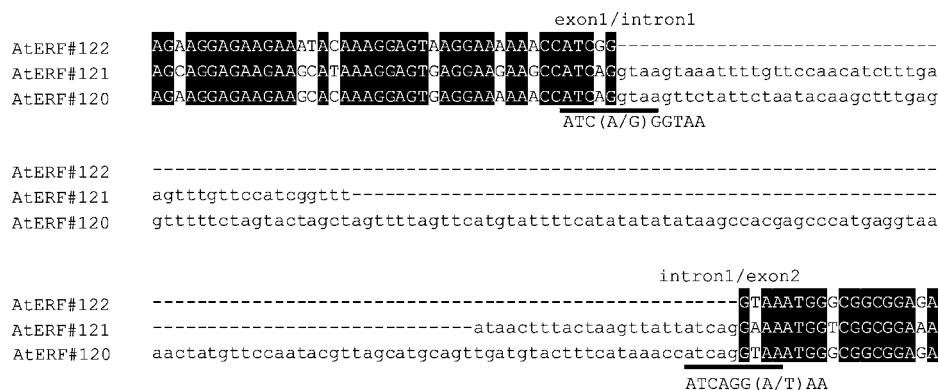
The genes in group IX have often been linked in defensive gene expression in response to pathogen infection. For example, the overexpression of Arabidopsis *ERF1* (AtERF#092) and tomato *Pti4* enhanced resistance to necrotic fungi and bacteria and biotrophic fungi, respectively (Berrocal-Lobo et al., 2002; Gu et al., 2002). Furthermore, defense-related phytohormones such as ethylene, jasmonate, and salicylic acid have been shown to differentially induce the expression of genes in group IX (Gu et al., 2000; Oñate-Sánchez and Singh, 2002).

Group X

Group X consists of eight genes (Fig. 3K; Table II). With the exception of AtERF#112, the products of these genes commonly contain one conserved motif, CMX-1, in the N-terminal region. In addition, with the exception of AtERF#109, the proteins of these genes contain a single intron (Fig. 3K). AtERF#109 and the genes in group IX were assigned to the B-3 group in a previous report (Sakuma et al., 2002). In this study, however, the results of the phylogenetic analysis (Fig. 3K) and the presence of the conserved CMX-1 motif indicate that AtERF#109 can be reasonably assigned to group X. Since AtERF#109 has some distinct features, e.g. an additional CMIX-2 motif and no intron, from the other genes in this group, this gene was selectively assigned to subgroup Xb. AtERF#112 was assigned to subgroup Xc because it has no conserved motifs (Fig. 3K; Supplemental Table II). Recently, Arabidopsis *ABR1* (AtERF#111) was identified as a repressor of ABA response. Disruption of the *ABR1* gene led to hypersensitivity response to ABA in seed germination and root growth assays (Pandey et al., 2005).

Group Xb-L

Group Xb-L consists of three genes (Fig. 3L; Table II). Like group VI-L, these genes are also characterized by an imperfect AP2/ERF domain (Fig. 1). Since the proteins of this group contain two conserved motifs, CMX-1 and CMX-2 (Fig. 3, K and L), which are characteristic features of AtERF#109 of subgroup Xb, these genes were designated as group Xb-L. Two of these genes, AtERF#120 and AtERF#121, contain a single intron at a common location, but this location is different from the genes of group X. Short direct repeats of nucleotide sequences were found around the exon/intron junction in these genes (Fig. 7). Because the insertion of transposable element results in sequence duplication at the target site, these introns might have been generated by the insertion of a transposable element into an AtERF#109-like ancestor gene.

```
                                          exon1/intron1
AtERF#122    AGAAGGAGAAGAAATACAAAGGAGTAAGGAAAAAACCATCGG------------------------------
AtERF#121    AGCACGGAGAAGAACCATAAAGCAGTGAGGAACGAAGCCATCACgtaagtaaattttgttccaacatctttga
AtERF#120    AGAAGGAGAAGAACGCCACAAAGCAGTGAGGAAAAAACCATCACgtaagttctattctaatacaagctttgag
                                                     ATC(A/G)GGTAA

AtERF#122    ----------------------------------------------------------------------
AtERF#121    agtttgttccatcggtttt-----------------------------------------------------
AtERF#120    gtttttctagtactagctagtttttagttcatgtattttcatatatatataagccacgagcccatgaggtaa


                                          intron1/exon2
AtERF#122    ------------------------------------------------------GTAAATGGCCGGCGGAGA
AtERF#121    ---------------------------ataacttttactaagttattatcagGAAAATGGTCGGCGGAAA
AtERF#120    aactatgttccaatacgttagcatgcagttgatgtactttcataaaccatcagCTAAATGGGCGGCGGAGA
                                                              ATCAGG(A/T)AA
```

Figure 7. The nucleotide sequence alignment of the exon1/intron1/exon2 region in group Xb-L genes. Intron sequences are in lowercase. Putative duplication sites are underlined, and conserved nucleotide sequences are shown. Identical nucleotides are highlighted with black shading.

Identification of ERF Genes in Rice in Silico and a Comparative Analysis with Arabidopsis

Multiple BLAST searches were performed in rice databases using the protein sequence of the AP2/ERF domain as a seed (see "Materials and Methods"), resulting in the identification of 139 ERF family genes (Tables II and IV). During the assessment of our results, we found that the Os04g48330 locus, which was assigned to OsERF#023, has been differently annotated in the National Center for Biotechnology Information (NCBI) and TIGR rice genome annotation databases. Os04g48330 was predicted to encode a protein of 157 amino acids with an incomplete AP2/ERF domain in contrast to the 237 amino acids of the result of our searches. Since the assessment of genomic sequence using the GENSCAN program (http://genes.mit.edu/GENSCAN.html) agreed with our result, we used the 237 amino acids in this study.

Full-length cDNA clones corresponding to 60 genes were identified in the Knowledge-based Oryza Molecular biological Encyclopedia (KOME) Web site (Supplemental Table III). Recently, the International Rice Genome Sequencing Project (IRGSP; International Rice Genome Sequencing Project, 2005) announced the completion of a high-quality rice genome sequence and proposed the existence of 157 AP2/ERF family genes that encoded a domain matched to InterPro ID IPR001471 (see supplemental table VII in the International Rice Genome Sequencing Project, 2005), whereas the specifics of the individual genes were unclear.

To determine the phylogenetic relationships among the ERF family genes in rice, a multiple alignment analysis was performed using amino acid sequences in the AP2/ERF domain. This analysis revealed that amino acid residues that may be involved in some form of physical contact with DNA are also conserved among most of the OsERF proteins (Supplemental Fig. 2). Three rice ERF proteins, OsERF#108, OsERF#109, and OsERF#138, have very low homology regarding the consensus sequence in the C-terminal region of the AP2/ERF domain, much like the groups VI-L and Xb-L in Arabidopsis. Since these OsERF proteins contain the conserved motifs CMVI-1 and CMVI-2 in the

N-terminal region, characteristic features of group VI and VI-L, these genes were designated as part of group VI-L in rice. In contrast, there is no gene that could be assigned to group Xb-L. Based on the above results, a phylogenetic tree was constructed using the sequence alignments of the AP2/ERF domains of the OsERF genes with the exception of OsERF#108, OsERF#109, and OsERF#138. The conserved motifs identified in the Arabidopsis ERF family were also examined in the form of the deduced amino acid sequences of the OsERF genes. The comparative analyses of the phylogenetic tree and the conserved motifs of the rice ERF family with those of Arabidopsis suggested that the classification used for Arabidopsis is in many ways applicable to the rice ERF family with several exceptions (Fig. 8; Tables II and IV).

One of the characteristic features of the rice ERF family is that the number of genes in group VII is 3 times larger than that in Arabidopsis (Table II). An MEME search using amino acid sequences of both AtERFs and OsERFs identified three additional conserved motifs, CMVII-6, -7, and -8, in this group (Figs. 3H and 9; Supplemental Table IV). OsERF#073 was assigned to a distinct subgroup, VIIb, because it does not have the diagnostic motif CMVII-1 of group VII (Supplemental Table III). Recently, OsEREBP1 (OsERF#070) was shown to be a possible substrate of a MAP kinase, BWMK1, in rice (Cheong et al., 2003). The phosphorylation of OsEREBP1 (OsERF#070) by BWMK1 results in the enhancement of its binding to the GCC box and the transactivation activity of GCC box-mediated transcription (Cheong et al., 2003). The CMVII-4 motif is identical to a putative MAP kinase phosphorylation site in OsEREBP1 (OsERF#070), which was suggested previously (Cheong et al., 2003). OsERF#071 and OsERF#072 in rice, and AtERF#074 (RAP2.12) and AtERF#075 (RAP2.2) in Arabidopsis, also contain the CMVII-4 motif (Fig. 6B; Supplemental Table III).

Gutterson and Reuber (2004) reported that the carboxyl-terminal half of the 24-amino acid (DMLV) motif (CMIX-3) was conserved in B-3a (group IXa) in both monocots and dicots, but that the amino-terminal half of the DMLV motif was specifically conserved in dicots. Our motif analysis revealed that the amino-

Table IV. *ERF family genes in rice*

Group Name	Generic Name	TIGR ID	Gene Name
Ia	OsERF#045	Os04g44670	
Ia	OsERF#046	Os02g42580	
Ib	OsERF#047	Os03g09170	
Ib	OsERF#048	Os08g31580	
Ib	OsERF#049	Os02g51670	
Ib	OsERF#050	Os09g20350	
Ib	OsERF#120	Os06g11860	
Ib	OsERF#051	Os10g22600	
Ib	OsERF#052	Os05g49700	
IIa	OsERF#007	Os06g07030	
IIa	OsERF#008	Os04g55520	
IIb	OsERF#009	Os03g15660	
IIb	OsERF#010	Os06g09690	
IIb	OsERF#011	Os02g54050	
IIb	OsERF#012	Os08g35240	
IIb	OsERF#013	Os06g11940	
IIb	OsERF#014	Os06g09810	
IIb	OsERF#015	Os06g09790	
IIb	OsERF#016	Os06g09760	
IIb	OsERF#119	Os06g10780	
IIb	OsERF#126	Os02g52880	
IIb	OsERF#139	Os06g09730	
IIc	OsERF#017	Os01g66270	
IIc	OsERF#131	Os05g34730	
IIIb	OsERF#019	Os11g13840	
IIIb	OsERF#021	Os02g35240	
IIIb	OsERF#022	Os04g36640	
IIIb	OsERF#020	Os02g45420	
IIIb	OsERF#023	Os04g48330	
IIIb	OsERF#137	Os03g02650	
IIIc	OsERF#018	Os10g38000	
IIIc	OsERF#024	Os09g35030	OsDREB1A
IIIc	OsERF#025	Os02g45450	
IIIc	OsERF#026	Os06g03670	OsDREB1C
IIIc	OsERF#027	Os01g73770	
IIIc	OsERF#028	Os08g43200	
IIIc	OsERF#029	Os08g43210	
IIIc	OsERF#030	Os04g48350	
IIIc	OsERF#031	Os09g35010	OsDREB1B
IIIc	OsERF#116	BAD67595.1[a]	OsDREB1D
IIIc	OsERF#133	Os09g35020	
IIId	OsERF#032	Os02g43940	
IIId	OsERF#033	Os04g46400	
IIId	OsERF#034	Os04g46440	
IIId	OsERF#035	Os02g43970	
IIId	OsERF#036	Os10g41130	
IIId	OsERF#037	Os04g46410	
IIIe	OsERF#122	Os06g36000	
IIIe	OsERF#038	Os02g13710	
IIId	OsERF#039	Os01g10370	
IVa	OsERF#040	Os01g07120	OsDREB2A
IVb	OsERF#041	Os03g07830	
IVb	OsERF#042	Os05g27930	
IVb	OsERF#043	Os05g39590	
IVb	OsERF#044	Os08g45110	
IVb	OsERF#117	Os05g28350	
Va	OsERF#001	Os06g40150	
Va	OsERF#002	Os06g08340	
Va	OsERF#003	Os02g10760	
Va	OsERF#127	Os02g55380	
Va	OsERF#129	Os04g56150	
Vb	OsERF#004	Os12g39330	

Table IV. (*Continued.*)

Group Name	Generic Name	TIGR ID	Gene Name
Vb	OsERF#005	Os07g10410	
Vb	OsERF#006	Os07g38750	
VI	OsERF#053	Os01g12440	
VI	OsERF#054	Os01g46870	
VI	OsERF#055	Os06g06540	
VI	OsERF#056	Os05g25260	
VI	OsERF#057	Os07g12510	
VI	OsERF#058	Os03g60120	
VIIa	OsERF#059	Os10g25170	
VIIa	OsERF#060	Os03g08460	OsEBP89
VIIa	OsERF#061	Os05g29810	
VIIa	OsERF#062	Os03g08470	
VIIa	OsERF#063	Os09g11480	
VIIa	OsERF#064	Os03g08500	
VIIa	OsERF#065	Os07g42510	
VIIa	OsERF#066	Os03g22170	
VIIa	OsERF#067	Os07g47790	
VIIa	OsERF#068	Os01g21120	
VIIa	OsERF#069	Os03g08490	
VIIa	OsERF#070	Os02g54160	OsEREBP1
VIIa	OsERF#071	Os06g09390	
VIIa	OsERF#072	Os09g26420	
VIIb	OsERF#073	Os09g11460	
VIIIa	OsERF#074	Os05g41780	
VIIIa	OsERF#075	Os01g58420	OsERF3
VIIIa	OsERF#076	Os04g57340	
VIIIa	OsERF#077	Os04g52090	
VIIIa	OsERF#121	Os06g47590	
VIIIa	OsERF#130	Os05g41760	
VIIIa	OsERF#132	Os02g06330	
VIIIa	OsERF#134	Os02g09650	
VIIIb	OsERF#078	Os07g47330	FZP
VIIIb	OsERF#079	Os02g38090	
VIIIb	OsERF#080	Os08g07700	
VIIIb	OsERF#081	Os02g32040	
VIIIb	OsERF#082	Os04g32790	
IXc	OsERF#083	Os03g64260	
IXc	OsERF#084	Os05g49010	
IXc	OsERF#085	Os05g37640	
IXc	OsERF#086	Os07g22770	
IXc	OsERF#087	Os09g39850	
IXc	OsERF#088	Os03g05590	
IXc	OsERF#089	Os10g30840	
IXc	OsERF#090	Os08g44960	
IXc	OsERF#123	Os09g39810	
IXc	OsERF#128	Os04g18650	
IXc	OsERF#136	Os07g22730	
IXa	OsERF#091	Os02g43790	
IXa	OsERF#092	Os01g54890	
IXa	OsERF#093	Os04g46220	
IXb	OsERF#094	Os04g46250	
IXb	OsERF#095	Os02g43820	
IXb	OsERF#096	Os10g41330	
IXb	OsERF#097	Os04g46240	
Xa	OsERF#098	Os02g34260	
Xa	OsERF#099	Os01g64790	
Xa	OsERF#100	Os04g34970	
Xa	OsERF#101	Os04g32620	
Xa	OsERF#118	Os11g06770	
Xa	OsERF#124	Os12g07030	
Xa	OsERF#125	Os02g34270	

(*Table continues on following page.*)

Table IV. (*Continued from previous page.*)

Group Name	Generic Name	TIGR ID	Gene Name
Xb	OsERF#102	Os09g28440	
Xb	OsERF#103	Os02g52670	
Xb	OsERF#104	Os08g36920	
Xc	OsERF#105	Os05g36100	
Xc	OsERF#106	Os08g42550	
Xc	OsERF#107	Os02g32140	
VI-L	OsERF#108	Os01g04020	
VI-L	OsERF#109	Os09g13940	
VI-L	OsERF#138	Os08g27220	
XI	OsERF#110	Os12g41030	
XI	OsERF#111	Os12g41040	
XI	OsERF#112	Os12g41060	
XI	OsERF#114	Os06g42910	
XII	OsERF#113	Os06g42990	
XIII	OsERF#135	BAC99579.1[a]	
XIV	OsERF#115	Os08g41030	
		Total 139 genes	

[a]The GenBank accession number is indicated because OsERF#116 and OsERF#135 do not have the TIGR locus identifier.

terminal half of CMIX-3 is also specifically conserved within the rice group IXa (data not shown). In addition, Gutterson and Reuber (2004) showed that a second short motif (NSGEPDPVRIKSKRS in AtERF1) was highly conserved only in dicots. Consistently with this, our results showed that this motif was not found in the rice ERF family and was limited to AtERF1 (AtERF#100) and AtERF2 (AtERF#101) in the Arabidopsis ERF family (data not shown).

Seven OsERF genes, OsERF#110 to #115 and OsERF#135, could not be assigned to any of the groups designated in Arabidopsis. These proteins were divided into three additional groups, XI to XIV (Fig. 8; Tables II and IV; Supplemental Table III). Group XI includes OsERF#110, #111, #112, and 114. Groups XII, XIII, and XIV consist of OsERF#113, #135, and #115, respectively. It is interesting to postulate the possible rice-specific functions of these OsERF genes. By contrast, there is no gene assigned to subgroup IIIa in the rice ERF family (Tables II and IV).

Two monocotyledonous genes of the ERF family, maize *BRANCHED SILKLESS1* (*BD1*; Chuck et al., 2002) and rice *FRIZZY PANICLE* (*FZP*; OsERF#078; Komatsu et al., 2003), have been shown to play crucial roles in the establishment of floral meristem identity. The structural features of the proteins encoded by *BD1* and *FZP* (OsERF#078) reveal that these genes are assigned to subgroup VIIIb. The subgroup VIIIb in Arabidopsis also includes genes that are involved in the differentiation and development of organs as described above. Komatsu et al. (2003) showed that FZP (OsERF#078) acts as a transcriptional activator in transiently transformed Arabidopsis cells. It has been shown that truncation of 10 amino acids at the C terminus affects the function of BD1 (Chuck et al., 2002). Truncation of this C-terminal region in FZP (OsERF#078) also resulted in a loss of function

(Komatsu et al., 2003). Since the phylogenetic relationships among the subgroup VIIIb proteins show that AtERF#086 is the closest homolog of BD1 and FZP (OsERF#078; data not shown), AtERF#086 would be a functional ortholog of BD1 and FZP (OsERF#078) in Arabidopsis. However, no apparent conserved sequence could be detected at the C termini of AtERF#086, BD1, and FZP (OsERF#078). The importance of the C terminus region in the function of FZP/BD1 may have specifically evolved in gramineous species such as rice and maize. By contrast, the motif CMVIII-3 was conserved in BD1, FZP (OsERF#078), AtERF#086, ESR1/DRN (AtERF#089), and AtERF#090.

Evolution and Divergence of the ERF Family Genes

The Arabidopsis genome has undergone several rounds of genome-wide duplication events, including polyploidy (Vision et al., 2000; Blanc et al., 2003; Bowers et al., 2003), which has great impact on the amplification of members of a gene family in the genome. To further investigate the relationship between the genetic divergence within the ERF family and gene duplication in Arabidopsis, the chromosomal location of each ERF gene was determined from the genomic sequences of Arabidopsis (Fig. 10). One hundred twenty-two ERF genes are distributed over all of the chromosomes except on the short arms of chromosomes II and IV. Ninety ERF genes are found in previously identified duplicated segmental regions on chromosomes (Fig. 10) that are the result of a polyploidy that occurred around 24 to 40 million years ago, probably close to the emergence of the crucifer family (Blanc et al., 2003). Although Blanc et al. (2003) did not include AtERF#017 and #018, AtERF#051 and #052, AtERF#069 and #070, and AtERF#077 and #080 as duplicated pair genes in recently duplicated segmental chromosomes, the phylogenetic relationships (Fig. 3, B, D, F, and I) suggest that they are closely related to each other. Of these, AtERF#017 and #018, AtERF#069 and #070, and AtERF#077 and #080 were listed as duplicated genes in the segmental duplications dataset maintained at TIGR (http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis_genome_duplication.shtml). Therefore, they were considered to be duplicated genes. Consequently, about 75% of ERF genes, which lie within recently duplicated segmental chromosomes, have a clear relative in these regions. Since the density of the duplicated genes in recently duplicated segmental chromosomes was reported to be $28.0\% \pm 7.8\%$ (Blanc et al., 2003), the duplicated pairs of ERF genes have been preferentially retained compared with other genes. This finding is consistent with a previous report demonstrating that duplicated genes involved in signal transduction and transcription are preferentially retained (Blanc and Wolfe, 2004). In particular, the number of ERF genes belonging to group III and group IX is relatively large in both Arabidopsis and rice, as shown in Table II. As shown in Table III, members of group III and group IX have been shown
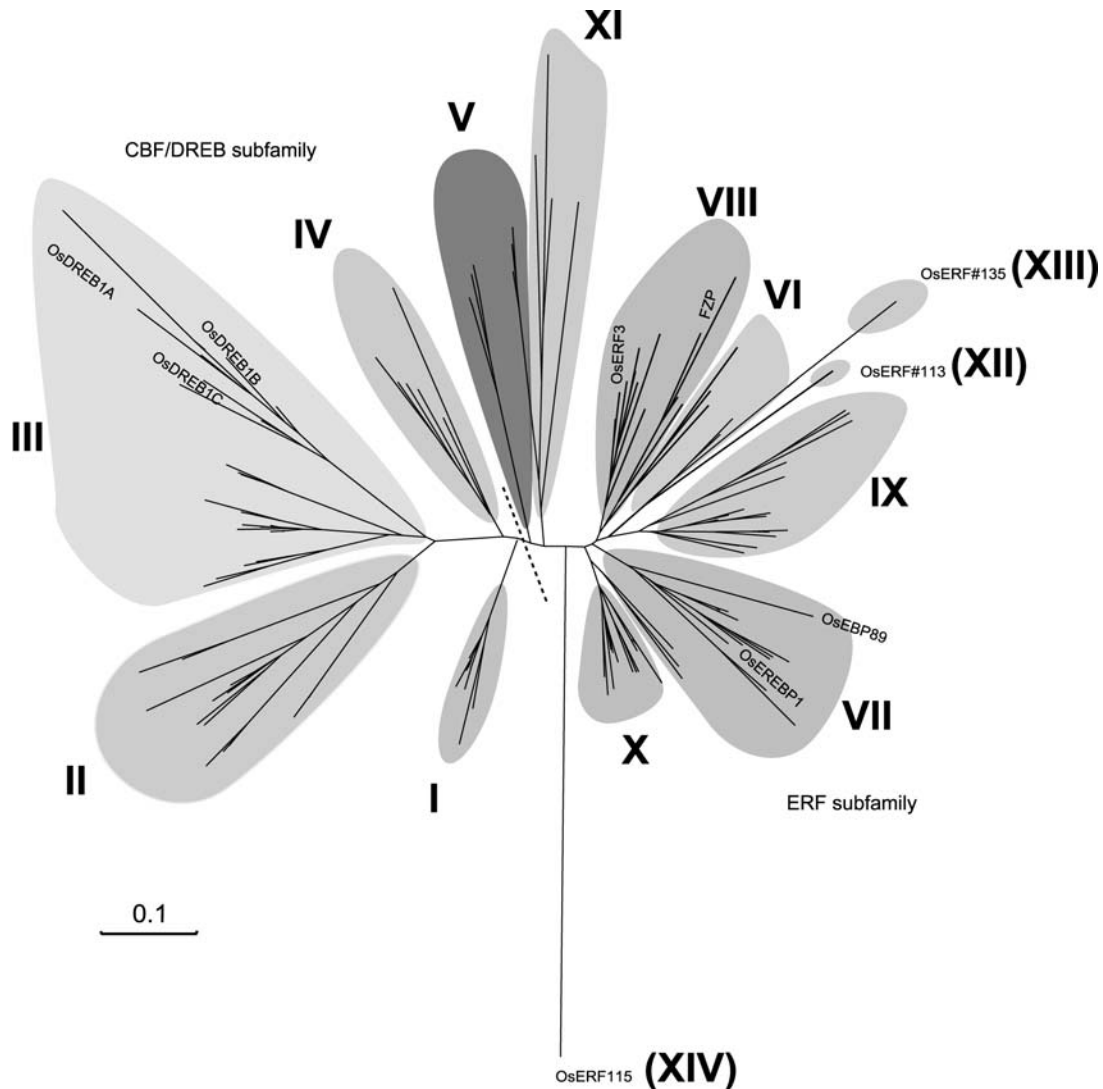
Figure 8. An unrooted phylogenetic tree of the ERF proteins in rice. The amino acid sequences of the AP2/ERF domain, except members of group VI-L, were aligned using ClustalW (Supplemental Fig. 2), and the phylogenetic tree was constructed using the NJ method. The names of the ERF genes that have been reported previously are indicated. The so-called CBF/DREB and ERF subfamilies are divided with a broken line.

to play crucial roles in abiotic and/or biotic stress responses. Therefore, the increased number of genes in these groups might be the evolutionary consequence of adapting to various environmental changes.

By comparison, the following genes appear to have undergone a tandem duplication event: AtERF#033 and #027; AtERF#081 and #076; AtERF#032 and #026; AtERF#048 and #047; AtERF#095, #098, and #092; AtERF#103 and #100; AtERF#030, #031, and #029; AtERF#101 and #102; AtERF#107 and #104; and AtERF#122 and #121 (Fig. 10). Based on the chromosomal locations and phylogenetic relationships, the history of some of the clades shown in Figure 3 can be somewhat explained. For example, the presence of the tandem arrays of AtERF#100 and #103, and AtERF#101 and #102, located on the recently duplicated segmental chromosome in chromosomes IV and V, respectively,

suggest that the tandem duplication of the ancestor of these genes predated the most recent polyploidy. Interestingly, the rice genome also contains two pairs of genes, OsERF#091 and #095, and OsERF#093 and #094, at two loci in chromosomes II and IV, respectively. OsERF#091 and #093, and OsERF#095 and #094, are closely related paralogs of groups IXa and IXb, respectively. This indicates that these gene pairs originated prior to monocot/eudicot divergence and that, since that point, the two pairs were duplicated in the Arabidopsis and rice genome in parallel. Thus, the pairwise divergence and evolution of ERF genes suggests that these genes might coordinately regulate certain biological processes common to these species.

On the other hand, we found that the DNA sequences for OsERF#010 (Os06g09690) and OsERF#139 (Os06g09730) genes were completely identical. These

Figure 9. Conserved amino acid sequence motifs in group VII ERF proteins. A, CMVII-6; B, CMVII-7; C, CMVII-8. The conserved motifs are underlined. Consensus sequences calculated by MEME program are given below the underlines. These regions were identified by MEME search using all members of Arabidopsis and rice group VII proteins. Black and gray shading indicate identical and conserved amino acid residues present in more than 50% of the aligned sequences, respectively.

genes are located on the 3′ end of PAC clone AP003510. A close inspection demonstrated that sequences of the genomic regions including them (at least 7.5 kb in length) were completely identical.

In the course of this study, ERF family genes in moss (*Physcomitrella patens* [BJ196641.1, BQ041358.1, BJ194243.1, BQ826584.1, BJ183188.1, BJ189648.1, and BQ040739.1]) and in unicellular green alga (*Chlamydomonas reinhardtii* [BQ823895 and BI718194]) were also identified. Because these sequences, which were derived from expressed sequence tags (ESTs), are incomplete, the diagnostic conserved motifs to identify phylogeny used in this study could not reliably be detected. However, a preliminary examination using the BLAST program suggested that BQ041358.1, BJ196641.1, BJ194243.1, BJ189648.1, BQ826584.1, BI718194, BQ040739.1, and BJ183188.1 encode ERF proteins belonging to group II or III, III, III, V, VIII, VIII, IX, and X, respectively (data not shown), suggesting that the basis of the phylogenetic topology of the ERF family had already been established before the divergence of vascular plants.

Recently, the AP2/ERF domain-encoding gene was reported in bacteria, a bacteriophage, and a ciliate genome as a part of homing endonuclease genes, mobile genetic elements that replicate and move in the genome (Magnani et al., 2004). In this report, it was also demonstrated that an AP2/ERF domain in a cyanobacterium, *Trichodesmium erythraeum*, recognizes stretches of poly(G)/poly(C), and that an Arabidopsis ERF protein, AtERF#060 (At4g39780), contains a region similar to the HNH domain in the cyanobacterium AP2/ERF protein (Magnani et al., 2004). Our

Figure 10. The locations of the ERF family genes on the Arabidopsis chromosomes. The chromosomal positions of the ERF genes are indicated by their generic names (Supplemental Table II). Group/subgroup names are shown in parentheses ahead of the generic name. The chromosome number is indicated at the top of each chromosome. The blue boxes indicate the duplicated segmental regions resulting from the most recent polyploidy (Blanc et al., 2003). Only the duplicated regions containing ERF genes are shown. Identical colored circles or squares indicate duplicated gene pairs, deduced by Blanc et al. (2003). The thick lines join tandem repeated genes. Colored fonts indicate ERF genes located on ancient segmental duplications (Blanc et al., 2003). AtERF#077 and #080 (black triangle), AtERF#017 and #018 (red triangle), AtERF#069 and #070 (blue triangle), and AtERF#051 and #052 (green triangle) are potential duplicated gene pairs, as described in the text.
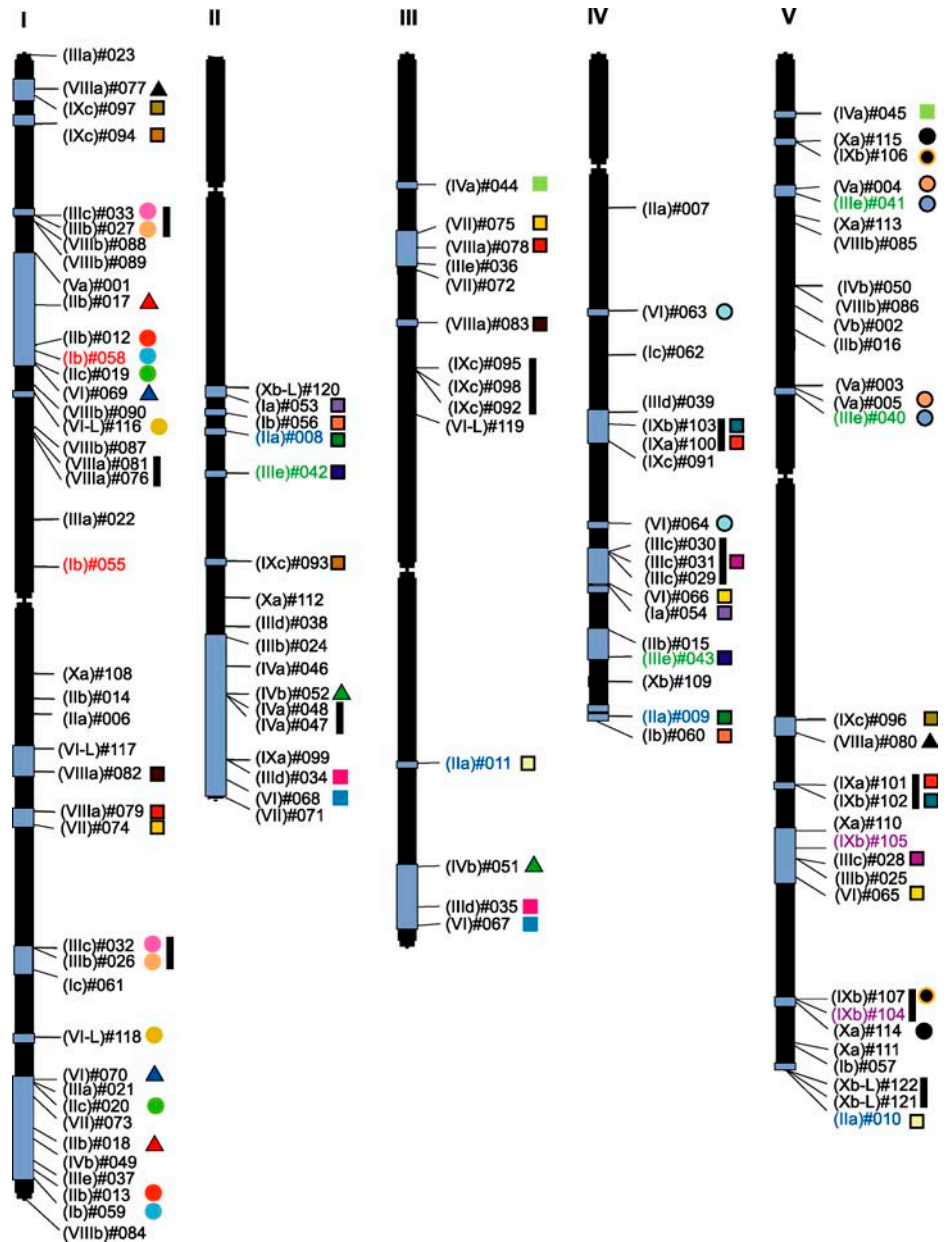
analysis showed that the HNH domain-like region of Arabidopsis ERF proteins corresponds to part of the CMI-3 motif that is shared with four members of subgroup Ib.

CONCLUSION

In this study, 122 and 139 ERF genes were identified in Arabidopsis and rice, respectively, and a comparative analysis between the phylogenetic relationships among the genes was performed. The results revealed a great deal about the diversification and conservation of the ERF family in plants. Chromosomal/segmental duplication, tandem gene duplication, as well as a more ancient transposition and homing might have contrib-

uted to the expansion of the ERF gene family. During the expansion of the ERF gene family, many groups and subgroups have evolved, resulting in a high level of functional divergence. Most of these groups/subgroups are present both in Arabidopsis and rice, suggesting that the appearance of many of the genes in these species predates monocot/eudicot divergence. Likewise, some groups/subgroups are present in only one species, suggesting that they have evolved or have been lost in one species after this divergence. Since rice is a cultivated species, selection either during domestication from its wild ancestor or during agricultural improvement in the subsequent time may also have been important for the evolution of rice ERF family. Members within a given group/subgroup may have recent common evolutionary origins and may

possess specific conserved motifs that have related molecular functions. Paralogous genes in a group/subgroup might have redundant functions. This may explain the low success rate of classical forward genetic strategies in the elucidation of the functions of ERF genes in plants (Table III). Phylogenetic and comparative analyses of ERF genes in Arabidopsis and rice will act as a first step toward a comprehensive functional characterization of the ERF gene family by reverse genetic approaches in the future. The results from the comparative study between Arabidopsis and rice will also provide useful information regarding the functions of ERF genes in agronomic, economic, and ecological traits in rice and possibly in other beneficial plant species.

MATERIALS AND METHODS

Database Search

Arabidopsis

Multiple database searches were performed to collect all members of the Arabidopsis (*Arabidopsis thaliana*) AP2/ERF superfamily. We used the BLAST programs (TBLASTN and BLASTP) available on the MAtDB, TAIR, and TIGR Arabidopsis databases and NCBI Arabidopsis genome database. As a query sequence, we first used the amino acid sequence of the AP2/ERF domain from tobacco (*Nicotiana tabacum*) ERF2. To increase the extent of the database search results, we also performed the position-specific iterated BLAST (Altschul et al., 1997) search against the Arabidopsis database on the NCBI Web site. We also performed the database searches using amino acid sequences of the AP2/ERF domain of some members of the Arabidopsis ERF family as a query sequence to confirm completion of the collection.

For the information regarding cDNA and ESTs, TAIR was searched using AGI ID. The exon/intron structures were investigated using SeqViewer at TAIR. The Arabidopsis ERF family is summarized in Supplemental Table II.

Rice

To identify members of the rice (*Oryza sativa* L. subsp. *japonica*) ERF family, multiple database searches were performed. First, we used the BLAST program (TBLASTN) available on the Rice Genome Database-japonica of the Rice Genome Research Program (http://rgp.dna.affrc.go.jp/) Web site. Based on this search, we identified 116 ERF family genes in the rice genome. The cDNA coding regions for the OsERF genes were predicted using the Rice Genome Automated Annotation System (http://ricegaas.dna.affrc.go.jp/rgadb/; Sakata et al., 2002). After the IRGSP (International Rice Genome Sequencing Project, 2005) announced completion of a high-quality rice genome sequence, we surveyed again the rice database using position-specific iterated BLAST (Altschul et al., 1997) program on the NCBI Web site. In addition, we surveyed the database of coding sequences from genes in the current version (December 30, 2004) of TIGR Rice Pseudomolecules for 12 chromosomes using a TBLASTN search at the TIGR Web site (http://tigrblast.tigr.org/euk-blast/index.cgi?project=osa1). In convenience for future analyses, if possible, we used the TIGR locus identifier. For these searches, we initially used the amino acid sequence of the AP2/ERF domain from tobacco ERF2 as a query sequence. Then, we surveyed the TIGR rice database again using the amino acid sequence of the AP2/ERF domain from OsERF#139 (Os06g09730) that has low homology to that of tobacco ERF2 (P value = 0.0020) as a query. Based on these searches, we collected all members of rice ERF family from the current available genomic database (Supplemental Table III).

The full-length rice cDNAs (Kikuchi et al., 2003) were searched at the KOME Web site (http://cdna01.dna.affrc.go.jp/cDNA/CDNA_main_front.html). The OsERF family gene is summarized in Supplemental Table III.

Other Plant Species

Physcomitrella patens and *Chlamydomonas reinhardtii* ERF genes were surveyed based on homology with the protein sequence of the AP2/ERF domain (Hao et al., 1998; Fujimoto et al., 2000). For *C. reinhardtii* ERF genes, a TBLASTN search was performed against the EST division at the DNA Data Bank of Japan Web site (http://www.ddbj.nig.ac.jp/search/blast-j.htmlat). For *P. patens* ERF genes, a TBLASTN search was performed on the Physcomitrella EST Project Web site (http://www.moss.leeds.ac.uk/). The obtained EST sequences were translated to protein sequences using the open reading frame finder at the NCBI Web site (http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi).

The Location of ERF Genes on Arabidopsis Chromosomes

To determine the location of AtERF genes on five chromosomes, Chromosome Map Tool (http://www.arabidopsis.org/jsp/ChromosomeMap/tool.jsp) at TAIR was used. Gene duplications and their presence on duplicated chromosomal segments were investigated using "Paralogons in Arabidopsis" (http://wolfe.gen.tcd.i.e./athal/dup) with the default parameters set to a minimum threshold for paired proteins per block above 7.

Sequences Analysis and Construction of the Phylogenetic Tree

A multiple alignment analysis was performed with ClustalW using DNASIS Pro software and/or DNASIS DNASpace (Hitachi Software). Phylogenetic trees were constructed using the neighbor-joining (NJ) method (Saitou and Nei, 1987) based on DNASIS DNASpace (Hitachi Software). The weight matrix used was BLOSUM 30. To predict the phosphorylation sites, a functional site prediction tool was used at the Eukaryotic Linear Motif resource for Functional Sites in Proteins (ELM; http://elm.eu.org/browse.html).

Determination of Conserved Motifs

Conserved motifs were investigated by multiple alignment analyses using ClustalW and MEME version 3.0 (Bailey and Elkan, 1994).

LITERATURE CITED

Aharoni A, Dixit S, Jetter R, Thoenes E, van Arkel G, Pereira A (2004) The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in Arabidopsis. Plant Cell **16:** 2463–2480

Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M, Suzuki M (1998) A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. EMBO J **17:** 5484–5496

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science **301:** 653–657

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **215:** 403–410

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:** 3389–3402

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana.* Nature **408:** 796–815

Arenas-Huertero F, Arroyo A, Zhou L, Sheen J, Leon P (2000) Analysis of *Arabidopsis* glucose insensitive mutants, *gin5* and *gin6,* reveals a central role of the plant hormone ABA in the regulation of plant vegetative development by sugar. Genes Dev **14:** 2085–2096

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *In* R Altman, D Brutlog, P Karp, R Lathrop, D Searls, eds, Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. American Association for Artificial Intelligence Press, Menlo Park, CA, pp 28–36

Banno H, Ikeda Y, Niu QW, Chua NH (2001) Overexpression of Arabidopsis *ESR1* induces initiation of shoot regeneration. Plant Cell **13:** 2609–2618

Berrocal-Lobo M, Molina A, Solano R (2002) Constitutive expression of *ETHYLENE-RESPONSE-FACTOR1* in *Arabidopsis* confers resistance to several necrotrophic fungi. Plant J **29:** 23–32

Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. Genome Res **13:** 137–144

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell **16:** 1679–1691

Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang L, Hattori J, Liu CM, van Lammeren AA, Miki BL, et al (2002) Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. Plant Cell **14:** 1737–1749

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422:** 433–438

Broun P, Poindexter P, Osborne E, Jiang CZ, Riechmann JL (2004) WIN1, a transcriptional activator of epidermal wax accumulation in *Arabidopsis.* Proc Natl Acad Sci USA **101:** 4706–4711

Büttner M, Singh KB (1997) *Arabidopsis thaliana* ethylene-responsive element binding protein (AtEBP), an ethylene-inducible, GCC box DNA-binding protein interacts with an ocs element binding protein. Proc Natl Acad Sci USA **94:** 5961–5966

Cheong YH, Moon BC, Kim JK, Kim CY, Kim MC, Kim IH, Park CY, Kim JC, Park BO, Koo SC, et al (2003) BWMK1, a rice mitogen-activated protein kinase, locates in the nucleus and mediates pathogenesis-related gene expression by activation of a transcription factor. Plant Physiol **132:** 1961–1972

Chuck G, Meeley RB, Hake S (1998) The control of maize spikelet meristem fate by the *APETALA2*-like gene *indeterminate spikelet1.* Genes Dev **12:** 1145–1154

Chuck G, Muszynski M, Kellogg E, Hake S, Schmidt RJ (2002) The control of spikelet meristem identity by the *branched silkless1* gene in maize. Science **298:** 1238–1241

Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) *OsDREB* genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. Plant J **33:** 751–763

Elliott RC, Betzner AS, Huttner E, Oakes MP, Tucker WQ, Gerentes D, Perez P, Smyth DR (1996) *AINTEGUMENTA,* an *APETALA2*-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth. Plant Cell **8:** 155–168

Eulgem T, Rushton PJ, Robatzek S, Somssich IE (2000) The WRKY superfamily of plant transcription factors. Trends Plant Sci **5:** 199–206

Finkelstein RR, Wang ML, Lynch TJ, Rao S, Goodman HM (1998) The Arabidopsis abscisic acid response locus *ABI4* encodes an APETALA 2 domain protein. Plant Cell **10:** 1043–1054

Fischer U, Droge-Laser W (2004) Overexpression of *NtERF5,* a new member of the tobacco ethylene response transcription factor family enhances resistance to *tobacco mosaic virus.* Mol Plant Microbe Interact **17:** 1162–1171

Fujimoto SY, Ohta M, Usui A, Shinshi H, Ohme-Takagi M (2000) Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. Plant Cell **12:** 393–404

Gilmour SJ, Sebolt AM, Salazar MP, Everard JD, Thomashow MF (2000) Overexpression of the Arabidopsis *CBF3* transcriptional activator mimics multiple biochemical changes associated with cold acclimation. Plant Physiol **124:** 1854–1865

Gilmour SJ, Zarka DG, Stockinger EJ, Salazar MP, Houghton JM, Thomashow MF (1998) Low temperature regulation of the *Arabidopsis* CBF family of AP2 transcriptional activators as an early step in cold-induced *COR* gene expression. Plant J **16:** 433–442

Gu YQ, Wildermuth MC, Chakravarthy S, Loh YT, Yang C, He X, Han Y, Martin GB (2002) Tomato transcription factors pti4, pti5, and pti6 activate defense responses when expressed in Arabidopsis. Plant Cell **14:** 817–831

Gu YQ, Yang C, Thara VK, Zhou J, Martin GB (2000) *Pti4* is induced by ethylene and salicylic acid, and its product is phosphorylated by the Pto kinase. Plant Cell **12:** 771–786

Guo ZJ, Chen XJ, Wu XL, Ling JQ, Xu P (2004) Overexpression of the AP2/EREBP transcription factor OPBP1 enhances disease resistance and salt tolerance in tobacco. Plant Mol Biol **55:** 607–618

Gutterson N, Reuber TL (2004) Regulation of disease resistance pathways by AP2/ERF transcription factors. Curr Opin Plant Biol **7:** 465–471

Haake V, Cook D, Riechmann JL, Pineda O, Thomashow MF, Zhang JZ (2002) Transcription factor CBF4 is a regulator of drought adaptation in Arabidopsis. Plant Physiol **130:** 639–648

Hao D, Ohme-Takagi M, Sarai A (1998) Unique mode of GCC box recognition by the DNA-binding domain of ethylene-responsive element-binding factor (ERF domain) in plant. J Biol Chem **273:** 26857–26861

He P, Warren RF, Zhao T, Shan L, Zhu L, Tang X, Zhou JM (2001) Overexpression of *Pti5* in tomato potentiates pathogen-induced defense gene expression and enhances disease resistance to *Pseudomonas syringae* pv. *tomato.* Mol Plant Microbe Interact **14:** 1453–1457

Heim MA, Jakoby M, Werber M, Martin C, Weisshaar B, Bailey PC (2003) The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. Mol Biol Evol **20:** 735–747

Hiratsu K, Mitsuda N, Matsui K, Ohme-Takagi M (2004) Identification of the minimal repression domain of SUPERMAN shows that the DLELRL hexapeptide is both necessary and sufficient for repression of transcription in *Arabidopsis.* Biochem Biophys Res Commun **321:** 172–178

Hu YX, Wang YX, Liu XF, Li JY (2004) *Arabidopsis* RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development. Cell Res **14:** 8–15

Huang Z, Zhang Z, Zhang X, Zhang H, Huang D, Huang R (2004) Tomato TERF1 modulates ethylene response and enhances osmotic stress tolerance by activating expression of downstream genes. FEBS Lett **573:** 110–116

Huijser C, Kortstee A, Pego J, Weisbeek P, Wisman E, Smeekens S (2000) The *Arabidopsis SUCROSE UNCOUPLED-6* gene is identical to *ABSCISIC ACID INSENSITIVE-4*: involvement of abscisic acid in sugar responses. Plant J **23:** 577–585

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature **436:** 793–800

Jaglo KR, Kleff S, Amundsen KL, Zhang X, Haake V, Zhang JZ, Deits T, Thimashow MF (2001) Components of the Arabidopsis C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in *Brassica napus* and other plant species. Plant Physiol **127:** 910–917

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

Jofuku KD, den Boer BG, Van Montagu M, Okamuro JK (1994) Control of Arabidopsis flower and seed development by the homeotic gene *APETALA2.* Plant Cell **6:** 1211–1225

Kagaya Y, Ohmiya K, Hattori T (1999) RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. Nucleic Acids Res **27:** 470–478

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, et al (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. Science **301:** 376–379

Kirch T, Simon R, Grunewald M, Werr W (2003) The *DORNROSCHEN/ENHANCER OF SHOOT REGENERATION1* gene of Arabidopsis acts in the control of meristem ccll fate and lateral organ development. Plant Cell **15:** 694–705

Kizis D, Pages M (2002) Maize DRE-binding proteins DBF1 and DBF2 are

involved in *rab17* regulation through the drought-responsive element in an ABA-dependent pathway. Plant J **30**: 679–689

Komatsu M, Chujo A, Nagato Y, Shimamoto K, Kyozuka J (2003) *FRIZZY PANICLE* is required to prevent the formation of axillary meristems and to establish floral meristem identity in rice spikelets. Development **130**: 3841–3850

Kranz HD, Denekamp M, Greco R, Jin H, Leyva A, Meissner RC, Petroni K, Urzainqui A, Bevan M, Martin C, et al (1998) Towards functional characterisation of the members of the *R2R3-MYB* gene family from *Arabidopsis thaliana*. Plant J **16**: 263–276

Lee JH, Hong JP, Oh SK, Lee S, Choi D, Kim WT (2004) The ethylene-responsive factor like protein 1 (CaERFLP1) of hot pepper (*Capsicum annuum* L.) interacts in vitro with both GCC and DRE/CRT sequences with different binding affinities: possible biological roles of CaERFLP1 in response to pathogen infection and high salinity conditions in transgenic tobacco plants. Plant Mol Biol **55**: 61–81

Lijavetzky D, Carbonero P, Vicente-Carbajosa J (2003) Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. BMC Evol Biol **3**: 17

Liu L, White MJ, MacRae TH (1999) Transcription factors and their genes in higher plants functional domains, evolution and regulation. Eur J Biochem **262**: 247–257

Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1998) 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. Plant Cell **10**: 1391–1406

Magnani E, Sjolander K, Hake S (2004) From endonucleases to transcription factors: evolution of the AP2 DNA binding domain in plants. Plant Cell **16**: 2265–2277

Magome H, Yamaguchi S, Hanada A, Kamiya Y, Oda K (2004) *dwarf and delayed-flowering 1*, a novel Arabidopsis mutant deficient in gibberellin biosynthesis because of overexpression of a putative AP2 transcription factor. Plant J **37**: 720–729

McGrath KC, Dombrecht B, Manners JM, Schenk PM, Edgar CI, Maclean DJ, Scheible WR, Udvardi MK, Kazan K (2005) Repressor- and activator-type ethylene response factors functioning in jasmonate signaling and disease resistance identified via a genome-wide screen of Arabidopsis transcription factor gene expression. Plant Physiol **139**: 949–959

Menke FLH, Champion A, Kijne JW, Memelink J (1999) A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene *Str* interacts with a jasmonate- and elicitor-inducible AP2-domain transcription factor, ORCA2. EMBO J **18**: 4455–4463

Moose SP, Sisco PH (1996) *Glossy15*, an *APETALA2*-like gene from maize that regulates leaf epidermal cell identity. Genes Dev **10**: 3018–3027

Ohme-Takagi M, Shinshi H (1995) Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell **7**: 173–182

Ohta M, Matsui K, Hiratsu K, Shinshi H, Ohme-Takagi M (2001) Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. Plant Cell **13**: 1959–1968

Ohta M, Ohme-Takagi M, Shinshi H (2000) Three ethylene-responsive transcription factors in tobacco with distinct transactivation functions. Plant J **22**: 29–38

Oñate-Sánchez L, Singh KB (2002) Identification of Arabidopsis ethylene-responsive element binding factors with distinct induction kinetics after pathogen infection. Plant Physiol **128**: 1313–1322

Ooka H, Satoh K, Doi K, Nagata T, Otomo Y, Murakami K, Matsubara K, Osato N, Kawai J, Carninci P, et al (2003) Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. DNA Res **10**: 239–247

Pandey GK, Grant JJ, Cheong YH, Kim BG, Li L, Luan S (2005) ABR1, an APETALA2-domain transcription factor that functions as a repressor of ABA response in Arabidopsis. Plant Physiol **139**: 1185–1193

Pǎrenicova L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B, et al (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: new openings to the MADS world. Plant Cell **15**: 1538–1551

Park JM, Park CJ, Lee SB, Ham BK, Shin R, Paek KH (2001) Over-expression of the tobacco *Tsi1* gene encoding an EREBP/AP2-type

transcription factor enhances resistance against pathogen attack and osmotic stress in tobacco. Plant Cell **13**: 1035–1046

Reyes JC, Muro-Pastor MI, Florencio FJ (2004) The GATA family of transcription factors in Arabidopsis and rice. Plant Physiol **134**: 1718–1732

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, et al (2000) *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science **290**: 2105–2110

Riechmann JL, Meyerowitz EM (1998) The AP2/EREBP family of plant transcription factors. Biol Chem **379**: 633–646

Romero I, Fuertes A, Benito MJ, Malpica JM, Leyva A, Paz-Ares J (1998) More than 80R2R3-MYB regulatory genes in the genome of *Arabidopsis thaliana*. Plant J **14**: 273–284

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4**: 406–425

Sakata K, Nagamura Y, Numa H, Antonio BA, Nagasaki H, Idonuma A, Watanabe W, Shimizu Y, Horiuchi I, Matsumoto T, et al (2002) RiceGAAS: an automated annotation system and database for rice genome sequence. Nucleic Acids Res **30**: 98–102

Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K (2002) DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochem Biophys Res Commun **290**: 998–1009

Solano R, Stepanova A, Chao Q, Ecker JR (1998) Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. Genes Dev **12**: 3703–3714

Song CP, Agarwal M, Ohta M, Guo Y, Halfter U, Wang P, Zhu JK (2005) Role of an Arabidopsis AP2/EREBP-type transcriptional repressor in abscisic acid and drought stress responses. Plant Cell **17**: 2384–2396

Stockinger EJ, Gilmour SJ, Thomashow MF (1997) *Arabidopsis thaliana* CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc Natl Acad Sci USA **94**: 1035–1040

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22**: 4673–4680

Tian C, Wan P, Sun S, Li J, Chen M (2004) Genome-wide analysis of the GRAS gene family in rice and *Arabidopsis*. Plant Mol Biol **54**: 519–532

Tiwari SB, Hagen G, Guilfoyle TJ (2004) Aux/IAA proteins contain a potent transcriptional repression domain. Plant Cell **16**: 533–543

Toledo-Ortiz G, Huq E, Quail PH (2003) The Arabidopsis basic/helix-loop-helix transcription factor family. Plant Cell **15**: 1749–1770

Tournier B, Sanchez-Ballesta MT, Jones B, Pesquet E, Regad F, Latche A, Pech JC, Bouzayen M (2003) New members of the tomato ERF family show specific expression pattern and diverse DNA-binding capacity to the GCC box element. FEBS Lett **550**: 149–154

Tsukagoshi H, Saijo T, Shibata D, Morikami A, Nakamura K (2005) Analysis of a sugar response mutant of Arabidopsis identified a novel B3 domain protein that functions as an active transcriptional repressor. Plant Physiol **138**: 675–685

van der Fits L, Memelink J (2000) ORCA3, a jasmonate-responsive transcriptional regulator of plant primary and secondary metabolism. Science **289**: 295–297

van der Graaff E, Dulk-Ras AD, Hooykaas PJ, Keller B (2000) Activation tagging of the *LEAFY PETIOLE* gene affects leaf petiole development in *Arabidopsis thaliana*. Development **127**: 4971–4980

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in *Arabidopsis*. Science **290**: 2114–2117

Wang H, Huang Z, Chen Q, Zhang Z, Zhang H, Wu Y, Huang D, Huang R (2004) Ectopic overexpression of tomato *JERF3* in tobacco activates downstream gene expression and enhances salt tolerance. Plant Mol Biol **55**: 183–192

Wang Z, Triezenberg SJ, Thomashow MF, Stockinger EJ (2005) Multiple hydrophobic motifs in Arabidopsis CBF1 COOH-terminus provide functional redundancy in trans-activation. Plant Mol Biol **58**: 543–559

Wilson K, Long D, Swinburne J, Coupland G (1996) A Dissociation

insertion causes a semidominant mutation that increases expression of *TINY*, an Arabidopsis gene related to *APETALA2*. Plant Cell **8:** 659–671

Yamamoto S, Suzuki K, Shinshi H (1999) Elicitor-responsive, ethylene-independent activation of GCC box-mediated transcription that is regulated by both protein phosphorylation and dephosphorylation in cultured tobacco cells. Plant J **20:** 571–579

Yang Z, Tian L, Latoszek-Green M, Brown D, Wu K (2005) Arabidopsis ERF4 is a transcriptional repressor capable of modulating ethylene and abscisic acid responses. Plant Mol Biol **58:** 585–596

Yi SY, Kim JH, Joung YH, Lee S, Kim WT, Yu SH, Choi D (2004) The pepper transcription factor CaPF1 confers pathogen and freezing tolerance in Arabidopsis. Plant Physiol **136:** 2862–2874

Zhang JY, Broeckling CD, Blancaflor EB, Sledge MK, Sumner LW, Wang ZY (2005) Overexpression of *WXP1*, a putative *Medicago truncatula* AP2 domain-containing transcription factor gene, increases cuticular wax accumulation and enhances drought tolerance in transgenic alfalfa (*Medicago sativa*). Plant J **42:** 689–707

Zhou J, Tang X, Martin GB (1997) The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a *cis*-element of pathogenesis-related genes. EMBO J **16:** 3207–3218

Zhu Y, Cai XL, Wang ZY, Hong MM (2003) An interaction between a MYC protein and an EREBP protein is involved in transcriptional regulation of the rice *Wx* gene. J Biol Chem **278:** 47803–47811