GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

A genome-wide analysis of the *ERF* gene family in sorghum

H.W. Yan, L. Hong, Y.Q. Zhou, H.Y. Jiang, S.W. Zhu, J. Fan and B.J. Cheng

Key Laboratory of Crop Biology of Anhui Province,
Anhui Agricultural University, Hefei, China

Corresponding author: B.J. Cheng
E-mail: beijiucheng@ahau.edu.cn

Genet. Mol. Res. 12 (2): 2038-2055 (2013)
Received November 9, 2012
Accepted March 8, 2013
Published May 13, 2013
DOI http://dx.doi.org/10.4238/2013.May.13.1

ABSTRACT. The ethylene response factor (ERF) family are members of the APETALA2 (AP2)/ERF transcription factor superfamily; they are known to play an important role in plant adaptation to biotic and abiotic stress. *ERF* genes have been studied in *Arabidopsis*, rice, grape, and maize; however, there are few reports of *ERF* genes in sorghum. We identified 105 sorghum *ERF* (*SbERF*) genes, which were categorized into 12 groups (A-1 to A-6 and B-1 to B-6) based on their sequence similarity, and this new method of classification for ERF genes was then further characterized. A comprehensive bioinformatic analysis of *SbERF* genes was performed using a sorghum genomic database, to analyze the phylogeny of *SbERF* genes, identify other conserved motifs apart from the AP2/ERF domain, map *SbERF* genes to the 10 sorghum chromosomes, and determine the tissue-specific expression patterns of *SbERF* genes. Gene clustering indicates that *SbERF* genes were generated by tandem duplications. Comparison of *SbERF* genes with maize *ERF* homologs suggests lateral gene transfer between monocot species. These results can contribute to our understanting of

the evolution of the *ERF* gene family.

Key words: *SbERF*; Genomic database; Phylogeny; Expression

INTRODUCTION

Plants can respond to environmental stresses such as drought, high salinity, and low temperature, which are the major factors affecting plant growth and crop production. Previous research has demonstrated that overexpression of certain genes is required to adapt to stress at the physiological and biochemical levels. The transcription factor APETALA2/ethylene response factor (AP2/ERF), characterized by a 57-66-amino acid AP2/ERF DNA-binding domain, plays an important role in the regulation of biotic and abiotic stress-responsive gene expression (Okamuro et al., 1997; Sakuma et al., 2002; Zhang et al., 2008). Based on sequence similarity and the structure of the AP2/ERF domains, the AP2/ERF superfamily, apart from one gene (At4g13040), can be divided into 3 main families: 1) AP2, 2) ERF, and 3) RAV (Nakano et al., 2006). AP2 family proteins contain two AP2/ERF domain repeats, and have important functions in the regulation of developmental processes, including leaf epidermal cell identity (Moose and Sisco, 1996), flower development (Elliott et al., 1996), spikelet meristem determinacy (Chuck et al., 1998), and embryo development (Boutilier et al., 2002). ERF family proteins contain a single conserved AP2/ERF domain and play crucial role in hormonal signal transduction (Shinshi et al., 1995), response to biotic and abiotic stresses (Dubouzet et al., 2003), and the regulation of metabolism during developmental processes (van der Graaff, et al., 2000; Banno et al., 2001; Chuck et al., 2002). RAV family proteins contain a conserved AP2/ERF domain and a B3-like domain, a plant-specific DNA-binding domain that is conserved within other transcription factors (Kagaya et al., 1999). The RAV family genes play a significant function in ethylene (Alonso et al., 2003), brassinosteroid (Hu et al., 2004), and biotic and abiotic stress responses.

The ERF family, which is the main focus of this study, can be further grouped into 2 major subfamilies: the CBF/dehydration response element binding (DREB) and the ethylene-responsive transcription factor (Sakuma et al., 2002). The ERF subfamily encodes pathogenesis-related proteins, which can specifically bind to *cis*-acting elements via a GCC box (AGCCGCC) (Shinshi, et al., 1995) of which the G2, G5, and C7 are essential (Buttner and Singh, 1997; Hao et al., 2002). In contrast, the DREB-type transcription factors play a vital role in conformation to abiotic stress and are known to bind to a dehydration-responsive element [(DRE), TACCGACAT] in the promoter regions of many dehydration and low-temperature stress-inducible genes (Baker et al., 1994; Yamaguchi-Shinozaki and Shinozaki, 1994; Stockinger et al., 1997; Thomashow, 1999; Stolf-Moreira et al., 2010). The central CCGAC sequence of the DREB is the minimal sequence motif, and the C4, G5, and C7 are essential for binding (Baker et al., 1994; Hao et al., 2002; Sakuma et al., 2002). The ERF-associated amphiphilic repression (EAR) motif belonging to transcriptional repressors has been examined in the subgroup members of many species, such as *Arabidopsis thaliana*, soybean, and maize. Recently, the EAR motif has been reported to participate in the modulation of plant stress and defense responses (Chuck et al., 2002). The 3 dimensional structure of the AP2/ERF domain consists of a 3-stranded anti-parallel β-sheet and an α-helix approximately parallel to the β-sheet (Allen et al., 1998). Allen et al. (1998) also demonstrated that arginine and tryptophan residues in the β-sheet contact the major groove of DNA using heteronuclear multidimensional NMR, indicating that the β-sheet of the AP2/ERF domain

functions in formation of the GCC box binding domain. The entire *ERF* family has been identified in some plants, and contains 147 and 157 genes in *Arabidopsis* and rice, respectively (Sakuma et al., 2002; Nakano et al., 2006). To date, 98 and 200 *AP2/ERF* genes have been indentified in soybean and *Populus trichocarpa*, respectively (Zhuang et al., 2008).

In this study, we identified 126 *AP2/ERF* superfamily unigenes, including 105 *ERF* family unigenes, 16 *AP2* family unigenes, 4 *RAV* unigenes, and a solo gene in sorghum. We performed a comprehensive survey and computational analysis of sorghum *ERF* (*SbERF*) genes. In contrast to the method of Sakuma et al. (2002), we classified the *ERF* gene family into 12 groups (A-1 to A-6, B-1 to B-6) on the basis of conserved domain sequence analysis and phylogenetic tree analysis. Data were further investigated to determine the specific function of a large number of the AP2/ERF transcription factor superfamily. We also performed a phylogenetic analysis of the AP2/ERF domain family and comparative phylogeny analysis in sorghum and maize, in order to evaluate heterogeneity in the evolutionary rates of the AP2/ERF family (Hu et al., 2010).The complete *ERF* classification, comparative species analysis, and predicted expression patterns of *SbERF* genes reported in this study will be helpful for future investigation of the biological functions of individual *ERF* genes and the evolution of *ERF* genes in different species.

MATERIAL AND METHODS

Database searching for the *ERF* gene family

Multiple databases were interrogated to identify members of the sorghum *AP2/ERF* superfamily. Sorghum genome sequences were downloaded from the website (http://genome.jgi-psf.org/Sorbi1/Sorbi1.download.ftp.html) and a sequence database was constructed using the DNATOOLS software. Initially, members of the *AP2/ERF* subfamily from *Arabidopsis* were used as query sequences and submitted to the PFAM database (http://pfam.janelia.org/search/sequence) to determine a conserved AP2 domain motif. Next, the AP2 domain HMM (hidden Markov model) from the PFAM database was applied to the sorghum sequence database to screen for homologous *AP2/ERF* gene sequences. The E-value used for the BlastP analysis was le+1. Finally the PFAM database was used to examine whether each candidate *SbERF* gene encoded the ERF domain and ERF protein sequence. To avoid identification of overlapping genes, all of the identified *SbERF* genes were aligned using Clustal W (Thompson et al., 1994) in MEGA v 4.0 (Tamura et al., 2007). On the basis of these searches, we were able to identify all members of the *SbERF* family in the current available genomic databases.

Phylogenetic analysis of *ERF* genes

The aligned SbERF protein sequences were merged using the Genestudio software and a phylogenetic tree was constructed using Clustal X version 1.83 (Thompson et al., 1997). This method was also applied for the phylogenetic comparison of sorghum and maize *ERF* genes.

Analysis of SbERF proteins

On the basis of the phylogenetic analysis, the conserved motifs encoded in ERF proteins were identified using MEME (http://meme.sdsc.edu/meme/meme.html) (Bailey et al.,

2006) to identify structural divergence between different subgroups of the ERF family. Bioinformatic analysis of *SbERF* genes was performed using ExPASy (http://www.expasy.ch/tools) to determine the number of amino acids (length), molecular weight (MW), isoelectric point (pI), and length of the open reading frame (ORF) for each gene.

Chromosomal location of *ERF* genes

In order to determine the chromosomal location of *SbERF* genes, the chromosomal starting position of each gene was confirmed by BLASTN searching, using local sequence databases of the complete sorghum genome. A chromosomal map was generated using the Genome Pixelizer software (http://www.niblrrs.ucdavis.edu/GenomePixelizer/GenomePixelizer_Welcome.html) and the *SbERF* genes were named according to their chromosomal position from the top to bottom of each chromosome.

Expression pattern analysis of *SbERF* genes

The expression patterns of each *SbERF* gene may provide a theoretical basis for further gene functional verification. Sorghum EST databases were acquired and the sorghum expression data were obtained using the DNATOOLS BLAST program. The expression data were queried using the BLASTN program with a maximum identity >95% and length >200 bp.

RESULTS

Identification of unigenes containing the AP2/ERF domain in sorghum

Using BLASTP searches, a total of 126 sorghum genes predicted to include one or more AP2/ERF domains were identified (Table 1). According to the PFAM database analysis, 16 genes were identified to encode proteins containing 2 AP2/ERF domains and could be assigned to the AP2 subfamily. Four genes were predicted to encode one AP2/ERF domain and one B3-like domain, and could be assigned to the RAV subfamily. One gene (*Sb09g002080.1*) was defined as a soloist as it contained a complete AP2/ERF domain, but exhibited a low homology with other *AP2/ERF* genes. The remaining 105 genes were predicted to encode proteins containing either a complete or incomplete AP2/ERF domain, and we named these genes *SbERF* genes (*SbERF 1-105*). The individual unigenes are listed in Table 2, including the ORF and gene lengths, MW, and pI predicted for each protein. The length of the *SbERF* ORFs ranged from 279 (*SbERF39*) to 1407 bp (*SbERF48*) and the MW ranged from 11.57 (*SbERF59*) to 98.10 kDa (*SbERF39*; Table 2).

Table 1. Summary of the AP2/ERF superfamily in *Arabidopsis*, rice, maize, and sorghum.

Classification	Conserved domain	Plant			
		Arabidopsis	Rice[a]	Maize	Sorghum
AP2 subfamily	Double AP2/ERF domain	18	26	31	16
RAV subfamily	Single AP2/ERF domain and one B3 domain	6	7	2	4
ERF subfamily	Single AP2/ERF domain	122	131	151	105
Soloist		1	0	1	1
Total		147	164	185	126

[a]Nakano et al., 2006.

Table 2. Sorghum *ERF* genes and their predicted features.

Gene	Sequence ID	Chromosome	ORF length (bp)	Predicted features		
				Length (aa)	MW (kDa)	pI
SbERF1	Sb01g000352.1	1	984	327	34.77	5.57
SbERF2	Sb01g003670.1	1	723	240	25.87	10.28
SbERF3	Sb01g014810.1	1	1353	450	47.55	4.72
SbERF4	Sb01g020555.1	1	411	136	13.96	7.84
SbERF5	Sb01g028930.1	1	834	277	30.10	7.13
SbERF6	Sb01g029065.1	1	912	303	31.71	4.72
SbERF7	Sb01g040280.1	1	741	246	25.48	5.13
SbERF8	Sb01g044410.1	1	954	317	33.75	7.10
SbERF9	Sb01g045040.1	1	975	324	34.46	4.65
SbERF10	Sb01g045050.1	1	807	268	28.77	6.01
SbERF11	Sb01g045060.1	1	1008	335	36.66	5.60
SbERF12	Sb01g045070.1	1	1008	335	36.06	6.99
SbERF13	Sb01g046920.1	1	441	146	15.28	5.54
SbERF14	Sb01g049400.1	1	774	257	35.74	5.39
SbERF15	Sb02g005810.1	2	1023	340	36.35	5.57
SbERF16	Sb02g005840.1	2	1017	338	35.96	5.34
SbERF17	Sb02g006023.1	2	900	299	31.78	8.84
SbERF18	Sb02g006780.1	2	963	320	34.83	9.15
SbERF19	Sb02g020900.1	2	867	288	32.84	4.82
SbERF20	Sb02g020910.1	2	849	282	31.50	4.75
SbERF21	Sb02g023230.1	2	915	304	32.14	6.75
SbERF22	Sb02g025530.1	2	1173	390	41.91	4.89
SbERF23	Sb02g026630.1	2	885	294	31.02	5.58
SbERF24	Sb02g029550.1	2	684	227	23.75	9.41
SbERF25	Sb02g030300.1	2	699	232	25.07	4.78
SbERF26	Sb02g030310.1	2	828	275	28.56	5.76
SbERF27	Sb02g030320.1	2	723	240	25.53	4.74
SbERF28	Sb02g030330.1	2	708	235	25.11	4.96
SbERF29	Sb02g030340.1	2	735	244	26.40	4.81
SbERF30	Sb02g030350.1	2	732	243	25.40	8.83
SbERF31	Sb02g039300.1	2	963	320	35.90	5.01
SbERF32	Sb02g042760.1	2	705	234	25.41	5.89
SbERF33	Sb03g002630.1	3	756	251	26.34	5.61
SbERF34	Sb03g004980.1	3	789	262	28.64	5.52
SbERF35	Sb03g007200.1	3	1035	344	36.63	4.75
SbERF36	Sb03g012890.1	3	684	227	24.22	8.59
SbERF37	Sb03g034780.1	3	543	180	18.29	9.87
SbERF38	Sb03g037080.1	3	729	242	24.93	9.15
SbERF39	Sb03g037085.1	3	279	92	98.10	11.34
SbERF40	Sb03g042060.1	3	576	191	19.27	10.20
SbERF41	Sb03g042110.1	3	888	295	31.36	5.80
SbERF42	Sb03g047170.1	3	729	242	25.81	6.63
SbERF43	Sb04g006080.1	4	645	214	22.11	6.89
SbERF44	Sb04g021160.1	4	744	247	25.42	8.13
SbERF45	Sb04g022245.1	4	792	263	27.87	10.12
SbERF46	Sb04g022943.1	4	588	195	20.73	5.72
SbERF47	Sb04g024690.1	4	933	310	33.10	6.62
SbERF48	Sb04g027180.1	4	1407	468	49.41	9.16
SbERF49	Sb04g027660.1	4	996	331	35.73	6.17
SbERF50	Sb04g031950.1	4	687	228	24.38	5.12
SbERF51	Sb04g031960.1	4	393	130	13.57	1.74
SbERF52	Sb04g032940.1	4	705	234	24.32	4.84
SbERF53	Sb04g032960.1	4	855	284	29.59	5.66
SbERF54	Sb04g033105.1	4	936	311	32.43	5.45
SbERF55	Sb04g034290.1	4	726	241	24.89	4.71
SbERF56	Sb04g035160.1	4	1107	368	40.09	4.79
SbERF57	Sb05g001280.1	5	723	240	26.06	6.07
SbERF58	Sb05g004170.1	5	1152	383	39.45	7.03
SbERF59	Sb05g006275.1	5	330	109	11.57	11.45
SbERF60	Sb05g007135.1	5	585	194	20.77	5.18

Continued on next page

Table 2. Continued.

Gene	Sequence ID	Chromosome	ORF length (bp)	Predicted features		
				Length (aa)	MW (kDa)	pI
SbERF61	Sb05g022110.1	5	960	319	34.18	6.33
SbERF62	Sb06g014800.1	6	861	286	29.66	4.47
SbERF63	Sb06g016710.1	6	693	230	23.63	9.00
SbERF64	Sb06g023313.1	6	1323	440	46.77	8.00
SbERF65	Sb06g024355.1	6	882	293	30.96	5.47
SbERF66	Sb06g024360.1	6	972	323	34.27	5.72
SbERF67	Sb06g024380.1	6	1041	346	36.93	4.91
SbERF68	Sb06g024390.1	6	858	285	30.70	6.20
SbERF69	Sb06g024530.1	6	966	321	33.35	4.90
SbERF70	Sb06g024540.1	6	765	254	26.90	4.83
SbERF71	Sb06g025890.1	6	651	216	22.81	6.52
SbERF72	Sb06g025900.1	6	684	227	24.59	5.54
SbERF73	Sb06g028090.1	6	696	231	23.79	9.21
SbERF74	Sb06g030660.1	6	768	255	27.69	9.78
SbERF75	Sb07g006190.1	7	684	227	22.90	9.97
SbERF76	Sb07g006195.1	7	657	218	22.39	7.77
SbERF77	Sb07g006200.1	7	633	210	21.98	9.90
SbERF78	Sb07g006210.1	7	645	214	22.72	8.47
SbERF79	Sb07g015070.1	7	876	291	31.67	6.76
SbERF80	Sb07g020090.1	7	816	271	29.27	5.70
SbERF81	Sb07g022180.1	7	1116	371	41.63	5.83
SbERF82	Sb07g022265.1	7	684	227	24.09	4.66
SbERF83	Sb07g023030.1	7	891	296	31.94	6.25
SbERF84	Sb07g023575.1	7	1167	388	41.25	6.67
SbERF85	Sb07g025210.1	7	723	240	24.75	5.31
SbERF86	Sb07g025843.1	7	837	278	29.58	9.98
SbERF87	Sb07g027260.1	7	537	178	19.40	10.65
SbERF88	Sb07g027280.1	7	495	164	17.51	11.35
SbERF89	Sb08g004260.1	8	1263	420	43.38	5.84
SbERF90	Sb08g019530.1	8	1269	422	44.28	4.53
SbERF91	Sb09g016150.1	9	798	265	28.34	4.41
SbERF92	Sb09g016550.1	9	750	249	27.04	7.22
SbERF93	Sb09g017950.1	9	600	199	20.49	5.38
SbERF94	Sb09g020690.1	9	657	218	21.80	9.55
SbERF95	Sb09g021540.1	9	528	175	18.73	7.92
SbERF96	Sb09g023190.1	9	717	238	25.43	6.83
SbERF97	Sb09g024400.1	9	774	257	25.92	8.79
SbERF98	Sb09g029070.1	9	909	302	32.36	9.27
SbERF99	Sb10g001490.1	10	510	169	18.13	7.80
SbERF100	Sb10g001620.1	10	750	249	26.23	4.75
SbERF101	Sb10g004580.1	10	606	201	20.36	10.00
SbERF102	Sb10g006260.1	10	777	258	39.51	4.94
SbERF103	Sb10g007780.1	10	1104	367	38.78	5.83
SbERF104	Sb10g012735.1	10	606	201	22.02	11.15
SbERF105	Sb10g028110.1	10	504	167	17.78	5.23

ORF = open reading frame; MW = molecular weight; pI = isoelectric point.

Phylogenetic analysis of the *ERF* genes in sorghum

To distinguish target sequences of the ERF family and analyze the evolution of each subfamily, a phylogenetic tree of all full-length sorghum ERF protein sequences was generated using the MEGA program by the neighbor-joining method (Figure 1). The phylogenetic tree showed that the 105 *SbERFs* distributed into 33 sister groups, indicating 2 taxa on either side of a split, with a common ancestor and no additional descendents.

Based on analysis of the phylogenetic tree, the ERF family proteins in sorghum were divided into 12 subgroups, namely: A-1 to A-6 and B-1 to B-6. Previously, the ERF

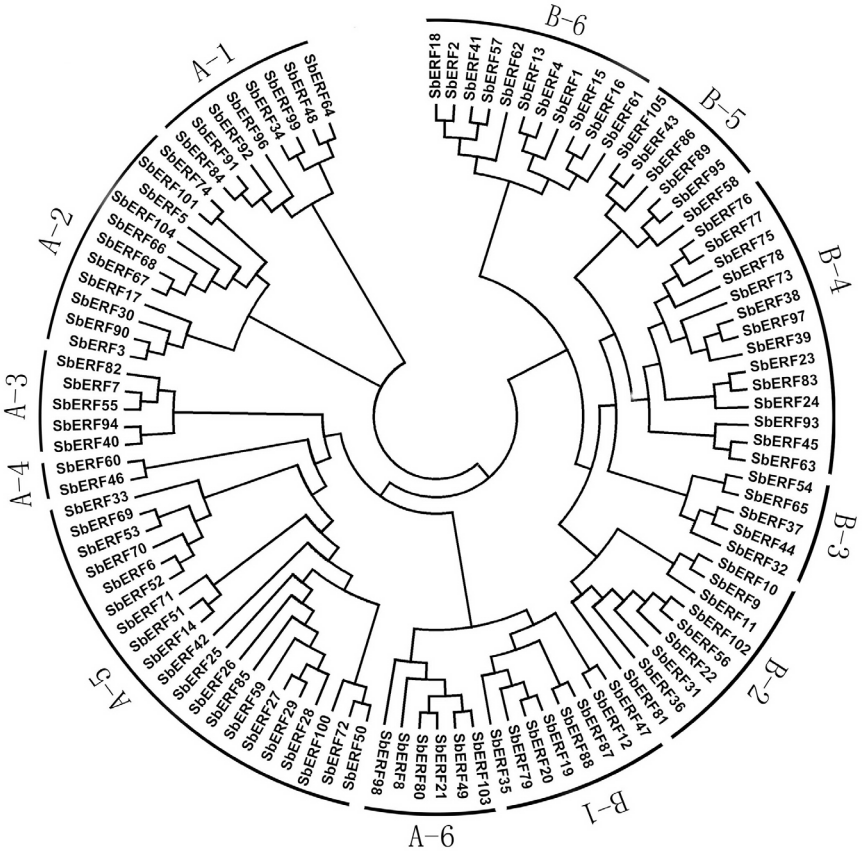

Figure 1. Phylogenetic tree of sorghum *ERF* genes. An unrooted tree was generated using the MEGA program by the neighbor-joining method.

family had been divided into 2 subfamilies in *Arabidopsis*: CBF/DREB and ERF. We referred to the method of classification described by Sakuma et al. (2002) and proposed a similar grouping method: subgroups A-1 to A-6 encode CBF/DREB proteins (group A), and subgroups B-1 to B-6 encode ERF proteins (group B; Figure 2). The alignments indicated that more than 88% of group A contained Val-7, and in addition, Glu-12 is incompletely conserved. Interestingly, residues Ala-7 and Asp-12 are highly conserved among the 53 group B members (Figure 2).

In order to better evaluate the phylogenetic relationship between the same gene families in different species, we compared sorghum with maize as a model monocot plant. The phylogenetic trees containing 105 *SbERF* genes and 151 maize *ERF* genes (*ZmERF*) were aligned and reconstructed (Figure 3). All sorghum and maize *ERF* genes could be classified into 96 sister groups, including 73 *SbERF-ZmERF*, 3 *SbERF-SbERF*, and 20 *ZmERF-ZmERF* sister groups. The high evolutionary rate of *ERF* genes indicates the influence of genome-wide factors, most likely those associated with Poaceae life history. These results show that genome-wide surveys are a promising approach towards a further understanding of the molecular clock during evolution at the genomic level.

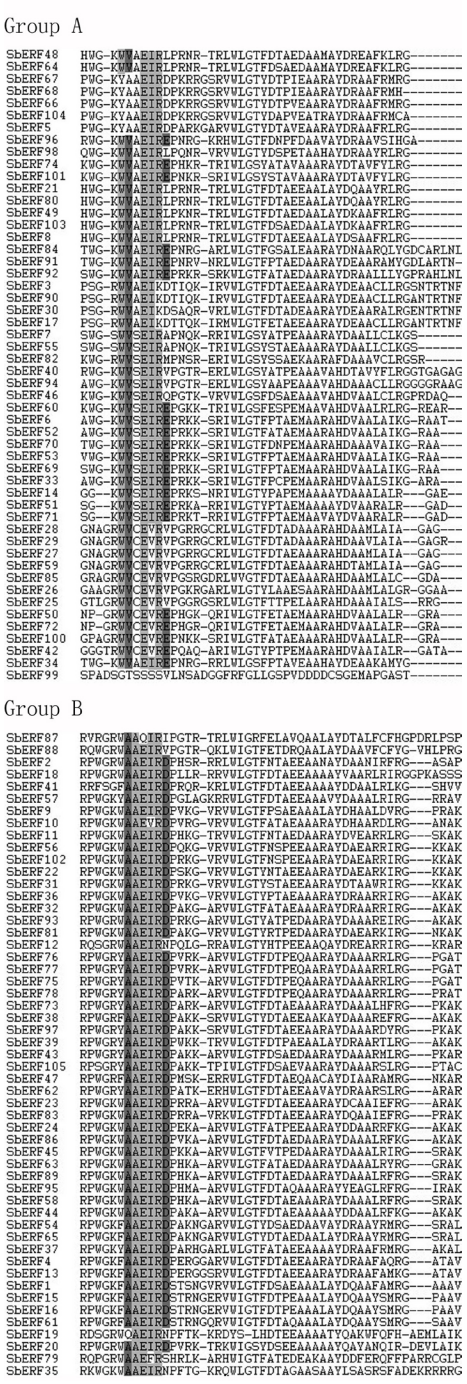

Figure 2. Comparison of the deduced amino acid sequences of DREB (group A) and ERF proteins (group B) in sorghum. The amino acid sequences of the AP2/ERF domains were aligned using ClustalW.

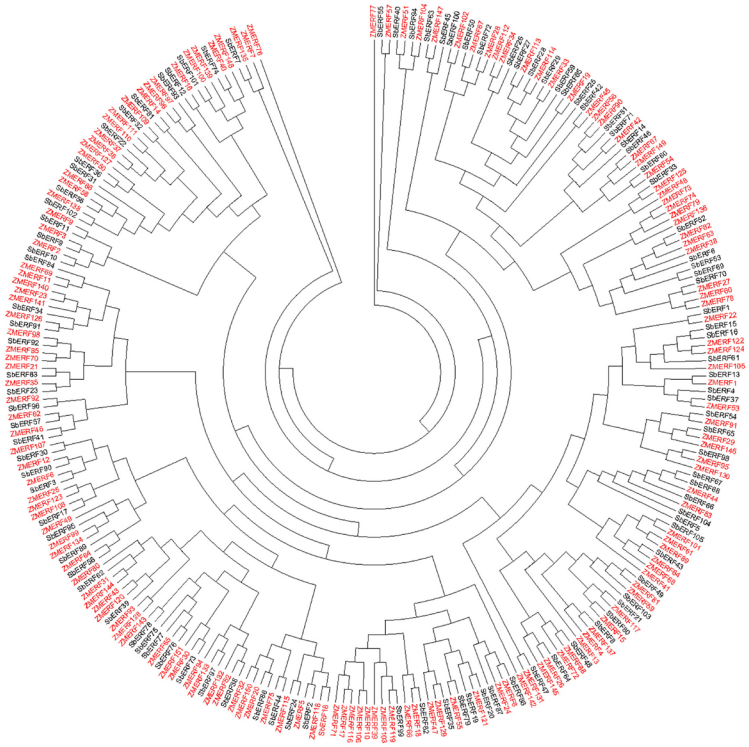

Figure 3. Phylogenetic tree of maize and sorghum *ERF* genes. An unrooted tree was generated using Clustal W by the neighbor-joining method.

Analysis of conserved motifs in the ERF family

Besides the AP2/ERF domain, other domains in *AP2/ERF* genes can also play an important role in transcriptional regulation. All sorghum AP2/ERF proteins were subjected to MEME motif analysis to identify conserved motifs amongst proteins in the same group (Figure 4). Conserved motifs can provide evidence for further classification into different subgroups as it is possible that proteins within a subgroup, which share identical motifs, are likely to exhibit similar functions; however, unfortunately it is still not possible to predict the function of these proteins. As shown in Figure 4, where motif 1 represents the AP2/ERF domain, several other motifs with important functions in transcriptional activity, protein-protein interactions, and nuclear localization were identified. Interestingly, all SbERF proteins contain at least one motif from motif 2, 3, 4, 5, or 6. This result is particularly significant as these major motifs may play a pivotal role in regulating expression of ERF proteins. The LWSY motif and EAR-like motif are found in some AP2/ERF subgroups in *Arabidopsis*, rice, maize, and sorghum (Figure 5), and these acidic sites may act as activation domains for transcription. All of the genes in the A-2 and B-3 groups contain motif 2 in the C-terminal region adjacent to the AP2/ERF domain; however, the position of this motif varied in other groups. Additionally, high levels of homology are present throughout the N-terminal region outside the AP2/ERF domain of group B-2.

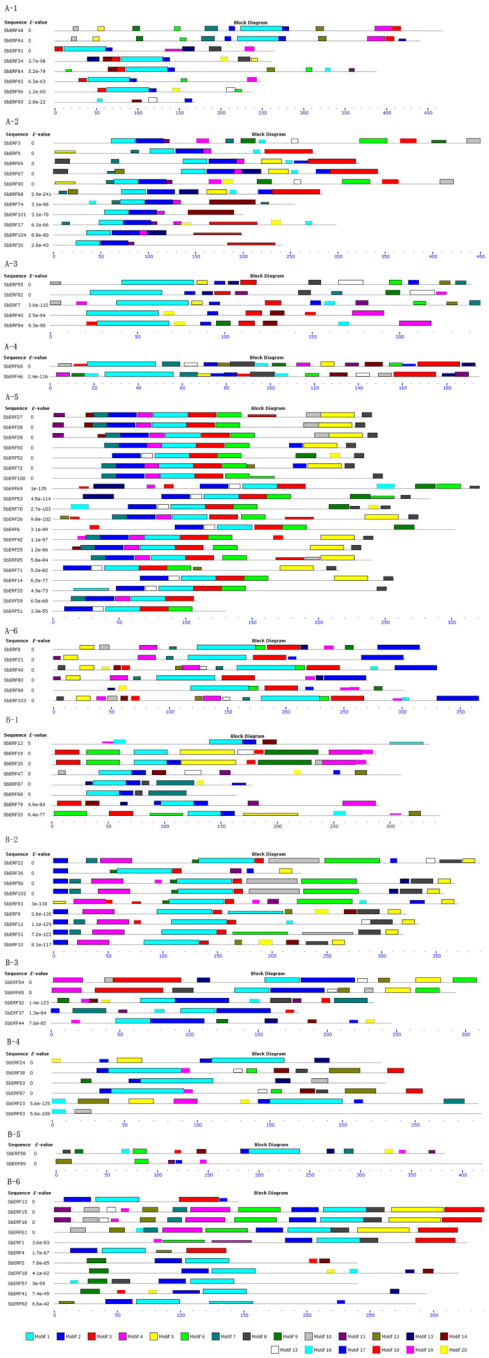

Figure 4. Conserved motifs identified in the sorghum ERF family using the MEME search tool. A = Conserved motifs in group A proteins. B = Conserved motifs in group B proteins. Gene names and combined P values (E-value) are indicated on the left, each color represents a distinct motif and the length of each motif is represented.

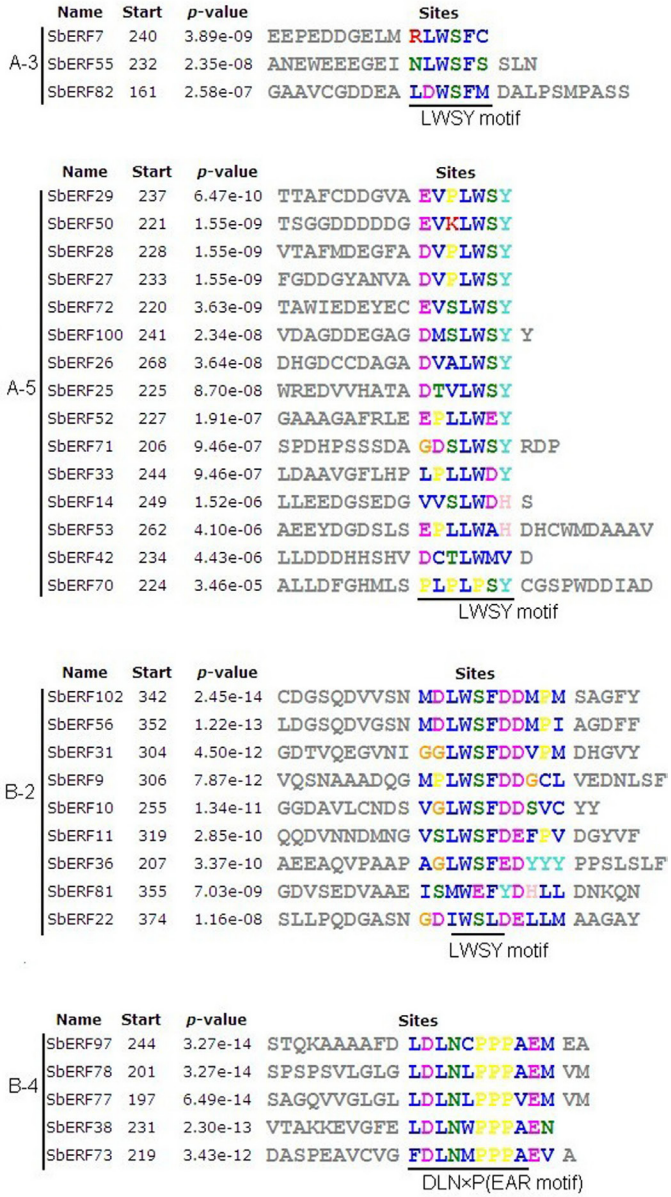

Figure 5. Conserved LWSY motif and ERF-associated amphiphilic repression (EAR) motif sequences in the C-terminal region of selected subgroups of ERF proteins. Consensus amino acid residues were identified using the MEME program and the conserved motifs are underlined.

Physical locations of *ERF* genes in sorghum

The chromosomal locations of the *SbERF* gene were determined using the Genome Pixelizer software, to further analyze evolution and divergence of *ERF* genes, based on the *SbERF*

classification. As shown in Figure 6, the 105 *SbERF* genes are unevenly distributed on the 10 sorghum chromosomes, with the largest number of *SbERF* genes located on chromosome 2 (N = 18). In total, 14 *SbERF* genes are located on chromosomes 1, 4, and 7, while chromosome 8 only contains 2 *SbERF* genes (Figure 6). Interestingly, almost all *SbERF* genes are located at either the top or bottom of chromosomes. Most group A-1 members are located on chromosome 9 (N = 3), with the others situated on chromosomes 3, 4, 6, 7, and 10. Four members of group A-2 are located on chromosome 6, with others on chromosomes 1, 2, 8, and 10. Similarly, group A-5 members are largely located on chromosome 2, group B-2 members are largely located on chromosome 1, and group B-4 members are largely located on chromosome 7. Moreover, the location of members of the same group within each chromosome is very close. This phenomenon is termed gene clustering, which results from the creation and deletion of tandem duplications.

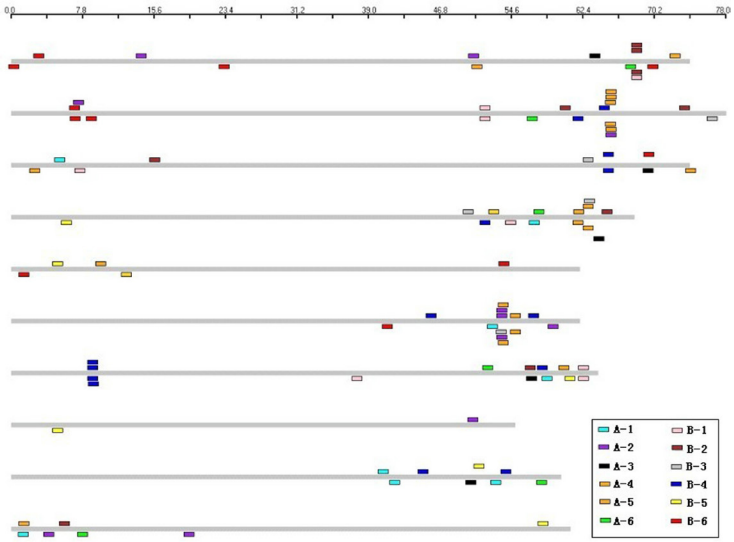

Figure 6. Genomic distribution of *ERF* genes on sorghum chromosomes. The 12 gene categories correspond to Figure 1. The colored boxes above and below the chromosomes (gray bars) designate the approximate chromosomal location of genes from each *ERF* gene category.

Expression patterns of *SbERF* genes

We detected the expression patterns of *SbERF* genes using a sorghum EST database. As illustrated in Table 3, we divided the sorghum EST data into 11 tissue groups: ear, pollen, embryo, callus, ovary, seeding, root, leaves, meristem, rhizome, or multiple tissues. Of the 105 *SbERF* genes, no apparent expression information was available for 15 *SbERF* genes. Based on the results of the BLASTN program analysis, 13 genes were identified to have tissue-specific expression: *SbERF105* in the ear; *SbERF5* in the embryo; S*bERF1*, *19*, *20*, *26*, *57*, *72*, and *74* in seedlings; *SbERF7* in the root; *SbERF14* and *71* in leaves; and *SbERF45* in the rhizome. The remainder of *SbERF* genes have a broad expression spectrum. The tissue organ-specific expression pattern of *ERF* genes indicates that *SbERF* transcripts are abundantly expressed in seedlings and relatively weakly expressed in meristem and rhizome.

Table 3. Expression patterns of *ERF* genes in sorghum.

Gene name	Ear	Pollen	Embryo	Callus	Ovary	Seeding	Root	Leaves	Meristem	Rhizome	Multiple
SbERF1						■					
SbERF2				●	●			●			●
SbERF3		●	●	●		●					
SbERF4			●	●	●	●		●			●
SbERF5			■								
SbERF6	●							●			
SbERF7							■				
SbERF8				●		●				●	
SbERF9				●	●	●		●			●
SbERF10			●					●	●		
SbERF11		●	●	●	●	●	●		●		●
SbERF12											
SbERF13	●				●	●		●	●		●
SbERF14								■			
SbERF15						●		●			●
SbERF16						●		●			●
SbERF17											
SbERF18	●			●	●	●	●	●	●		●
SbERF19						■					
SbERF20						■					
SbERF21	●			●	●	●	●	●	●	●	●
SbERF22	●		●	●	●	●	●	●		●	●
SbERF23		●	●	●		●					●
SbERF24		●	●			●					
SbERF25											
SbERF26						■					
SbERF27	●				●	●			●		●
SbERF28	●				●	●			●		●
SbERF29	●			●	●						●
SbERF30											
SbERF31			●	●		●					
SbERF32		●	●	●			●				
SbERF33		●				●			●		
SbERF34		●		●		●					
SbERF35	●	●	●	●	●	●	●				
SbERF36		●	●	●	●	●	●	●	●		
SbERF37	●	●	●		●	●		●		●	
SbERF38											
SbERF39											
SbERF40											
SbERF41	●				●				●		●
SbERF42											
SbERF43						●		●			●
SbERF44		●	●	●						●	
SbERF45										■	
SbERF46	●	●		●				●			
SbERF47	●			●	●						●
SbERF48	●				●	●	●	●	●	●	
SbERF49	●		●	●	●	●	●	●			
SbERF50											
SbERF51				●		●		●	●		
SbERF52						●		●			
SbERF53		●		●	●	●	●	●		●	●
SbERF54						●	●			●	
SbERF55				●			●				●
SbERF56				●		●	●			●	●
SbERF57						■					
SbERF58				●		●					
SbERF59						●					●
SbERF60					●		●		●	●	

Continued on next page

Table 3. Continued.

Gene name	Ear	Pollen	Embryo	Callus	Ovary	Seeding	Root	Leaves	Meristem	Rhizome	Multiple
SbERF61											
SbERF62			●	●	●	●	●	●		●	●
SbERF63				●		●					
SbERF64	●			●	●	●	●	●	●	●	●
SbERF65			●		●	●		●			
SbERF66			●			●					
SbERF67			●	●			●				
SbERF68			●	●							
SbERF69	●							●			
SbERF70						●		●			
SbERF71								■			
SbERF72						■					
SbERF73	●			●	●						●
SbERF74						■					
SbERF75							●			●	
SbERF76		●			●	●	●			●	
SbERF77							●			●	
SbERF78				●			●			●	
SbERF79											
SbERF80			●	●	●	●	●	●	●	●	●
SbERF81											
SbERF82											
SbERF83					●	●			●		
SbERF84		●		●							
SbERF85			●			●					
SbERF86		●	●			●					
SbERF87											
SbERF88					●				●		
SbERF89	●				●	●			●		●
SbERF90											
SbERF91	●			●		●				●	
SbERF92		●		●	●						
SbERF93					●				●		
SbERF94		●		●				●			
SbERF95		●	●	●					●		
SbERF96		●		●							
SbERF97			●	●	●	●		●			●
SbERF98		●	●	●		●	●	●			
SbERF99	●									●	
SbERF100			●					●			
SbERF101		●	●	●	●	●	●				
SbERF102	●		●		●	●	●	●	●	●	●
SbERF103	●		●			●		●			
SbERF104			●				●				
SbERF105	■										

Dots = positive expression of *ERF* genes in each sorghum tissue and organ. Squares = tissue organ-specific expression. Multiple represents expression data from mixed sorghum tissues and organs.

DISCUSSION

As crucial transcription factors, the ERF family have a significant impact on the evolution of plant-specific developmental mechanisms (Riechmann and Meyerowitz, 1998). During the last decade, a considerable amount of research has indicated that overexpression of *ERF* family genes enhances the resistance of plants to environmental stress. For example, overexpression of *ERF* genes in chillies (Shin et al., 2002), *Arabidopsis* (Berrocal-Lobo et al., 2002), and tomato (Gu et al., 2002) improves plant disease resistance to viral and bacterial pathogens. The gene structure, phylogeny, and conserved motifs of the *ERF* gene family have

been extensively studied in several species, including a genome-wide analysis in the model plants *Arabidopsis* and rice (Nakano et al., 2006). However, few *ERF* genes have previously been studied in sorghum. With the completion of sorghum genome sequencing projects, phylogenetic analysis of *SbERF* genes would be helpful to investigate general function of *ERF* genes and the evolutionary process of the AP2/ERF domain in plants.

In many plants, there are often many more genes in the *ERF* superfamily than other families. In this study, we identified 126 distinct genes encoding AP2/ERF proteins in sorghum, which could be classified into 4 subfamilies based on the similarity of the AP2/ERF-binding domain: *AP2* (16 genes), *RAV* (4 genes), *ERF* (105 genes), and a soloist gene (*Sb09g002080.1*). Of the 105 *SbERF* genes, we identified 36 candidate genes in which the residues Glu-11, Ile-12, Arg-13, Arg-21, Trp-23, Leu-24, Gly-25, Ala-33, Ala-34, and Ala-36 are highly conserved (Figure S1), suggesting that these residues are essential for the function of the AP2/ERF domain. Based on the analysis of Sakuma et al. (2002) the 105 *SbERF* genes could be divided into *CBF/DREB* and *ERF* genes. Using a sorghum genome sequence database, we observed that *DREB* subfamily genes share the *DRE* element (TACCGACAT) of which C-4, G-5, and C-7 are essential for CBF/DREB-type proteins (Sun et al., 2008; Wang et al., 2010). Additionally, Val-7 and Glu-12 are highly conserved in CBF/DREB proteins (Figure 2). On the other hand, the *ERF* subfamily contains the core GCC box sequence (AGCCGCC), of which the G-2, G-5, C-7 are absolutely necessary (Gutterson and Reuber, 2004). Our analysis demonstrated that Ala-7 and Asp-12 are important for ERF proteins. The ERF subfamily can specifically bind the GCC box present in the promoter of many related protein genes and interfere with their expression, consequently the ERF subfamily plays a vital role in disease resistance.

On the basis of the homogeneous characteristics and evolutionary or structural relationships, both CBF/DREB (group A) and ERF proteins (group B) could be further divided into 6 distinct subgroups A-1 to A-6 and B-1 to B-6, respectively. We selected some sorghum ERF family groups for further analysis and identified the characteristics of each. The A-3, A-5 and B-2 groups contain an LWSY motif at the C-terminus, consistent with other ERF proteins, which respond to cold treatment (Jin and Liu, 2008), suggesting that the LWSY motif may be involved in plant cold tolerance. Comparative analysis of amino acid residues has indicated that the most conserved motifs in the AP2/ERF superfamily are present in other species, including *Arabidopsis*, rice, maize, and other plants. For example, the EAR motif in the C-terminal region of the VIIIa and IIa ERF protein subgroups (Nakano et al., 2006) is also present in SbERF proteins. This signature sequence is reported to function in gene repression transcriptional regulatory cascades (Ohta et al., 2001). Apart from the commonly conserved motifs, which are found in *Arabidopsis* and rice, other ERF family motifs, which play an important role in the regulation of biological processes, are present in sorghum. The 105 *SbERF* genes were found to have formed 33 sister pairs as a result of interaction of replication events to expand *SbERF* number. Through comparative analysis of the diversification and conservation of the *ERF* family in sorghum and maize, it was demonstrated that *SbERF* have a high homology with *ERF* in other monocot plants, with 76% *SbERF-ZmERF* sister pairs.

SbERF genes map irregularly to the 10 sorghum chromosomes, and several are physically clustered together. A gene cluster is a chromosome region containing 2 or more genes, which encode the same or similar products, and gene clusters are useful for tracing recent evolutionary history (Engel et al., 1970; Caetano-Anollés, 2001). In this study, we found that 37 of the *SbERF* genes are located within 13 clusters, with chromosomes 1, 2, 3, 4, 6, and 7 contain-

ing 1, 3, 2, 2, 2, and 3 gene clusters, respectively, the largest of which contained 6 *ERF* genes. No clusters occur on chromosome 5, 8, 9, or 10. Additionally, we examined the expression profiles of *SbERF* genes in 11 sorghum tissues and organs using an EST database. However, 15 genes were found not to be expressed. It is possible that these genes might be expressed in different growth environments, which may interfere with the functional analysis of these genes.

The complete *ERF* classification, comparative species analysis, and predicted expression patterns of *SbERF* genes reported in this study will be helpful for future investigation of the biological functions of individual *ERF* genes and the evolution of *ERF* genes in different species. However, the analysis of genomic databases provides the foundation for the study of gene function; therefore, these results require experimental verification. Different growth environments including abiotic, drought, cold, and high salinity stress may significantly affect functional prediction and it is possible that the function of some genes may have been lost in sorghum after divergence.

ACKNOWLEDGMENTS

Research supported by grants from the Key Projects in the National Science & Technology Pillar Program (#2009BADA6B00) and the Ministry of Education Doctoral New Teachers Fund (#20103418120001). We also thank members of the Key Laboratory of Crop Biology of Anhui Province for their assistance in this study.

Supplementary material

REFERENCES

Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M, et al. (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, Solano R, Wisman E, et al. (2003). Five components of the ethylene-response pathway identified in a screen for weak ethylene-insensitive mutants in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 100: 2992-2997.

Bailey TL, Williams N, Misleh C and Li WW (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34: W369-W373.

Baker SS, Wilhelm KS and Thomashow MF (1994). The 5'-region of *Arabidopsis thaliana* cor15a has cis-acting elements that confer cold-, drought- and ABA-regulated gene expression. *Plant Mol. Biol.* 24: 701-713.

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

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

Boutilier K, Offringa R, Sharma VK, Kieft H, et al. (2002). Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* 14: 1737-1749.

Buttner M and 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. U. S. A.* 94: 5961-5966.

Caetano-Anollés G (2001). Novel strategies to study the role of mutation and nucleic acid structure in evolution. *Plant Cell Tiss. Org.* 67: 115-132.

Chuck G, Meeley RB and 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, et al. (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, et al. (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, et al. (1996). AINTEGUMENTA, an *APETALA2*-like gene of *Arabidopsis* with pleiotropic roles in ovule development and floral organ growth. *Plant Cell* 8: 155-168.
Engel W, Hof JO and Wolf U (1970). Gene duplication by polyploid evolution: the isoenzyme of the sorbitol dehydrogenase in herring- and salmon-like fishes (*Isospondyli*). *Humangenetik* 9: 157-163.
Gu YQ, Wildermuth MC, Chakravarthy S, Loh YT, et al. (2002). Tomato transcription factors pti4, pti5, and pti6 activate defense responses when expressed in *Arabidopsis*. *Plant Cell* 14: 817-831.
Gutterson N and Reuber TL (2004). Regulation of disease resistance pathways by AP2/*ERF* transcription factors. *Curr. Opin. Plant Biol.* 7: 465-471.
Hao D, Yamasaki K, Sarai A and Ohme-Takagi M (2002). Determinants in the sequence specific binding of two plant transcription factors, CBF1 and NtERF2, to the DRE and GCC motifs. *Biochemistry* 41: 4202-4208.
Hu X, Cheng X, Jiang H, Zhu S, et al. (2010). Genome-wide analysis of cyclins in maize (*Zea mays*). *Genet. Mol. Res.* 9: 1490-1503.
Hu YX, Wang YX, Liu XF and 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.
Jin LG and Liu JY (2008). Molecular cloning, expression profile and promoter analysis of a novel ethylene responsive transcription factor gene GhERF4 from cotton (*Gossypium hirstum*). *Plant Physiol. Biochem.* 46: 46-53.
Kagaya Y, Ohmiya K and 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.
Moose SP and Sisco PH (1996). Glossy15, an *APETALA2*-like gene from maize that regulates leaf epidermal cell identity. *Genes Dev.* 10: 3018-3027.
Nakano T, Suzuki K, Fujimura T and Shinshi H (2006). Genome-wide analysis of the *ERF* gene family in *Arabidopsis* and rice. *Plant Physiol.* 140: 411-432.
Ohta M, Matsui K, Hiratsu K, Shinshi H, et al. (2001). Repression domains of class II *ERF* transcriptional repressors share an essential motif for active repression. *Plant Cell* 13: 1959-1968.
Okamuro JK, Caster B, Villarroel R, Van MM, et al. (1997). The AP2 domain of *APETALA2* defines a large new family of DNA binding proteins in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 94: 7076-7081.
Riechmann JL and Meyerowitz EM (1998). The AP2/EREBP family of plant transcription factors. *Biol. Chem.* 379: 633-646.
Sakuma Y, Liu Q, Dubouzet JG, Abe H, et al. (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.
Shin R, Park JM, An JM and Paek KH (2002). Ectopic expression of Tsi1 in transgenic hot pepper plants enhances host resistance to viral, bacterial, and oomycete pathogens. *Mol. Plant Microbe Interact.* 15: 983-989.
Shinshi H, Usami S and Ohme-Takagi M (1995). Identification of an ethylene-responsive region in the promoter of a tobacco class I chitinase gene. *Plant. Mol. Biol.* 27: 923-932.
Stockinger EJ, Gilmour SJ and 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. U. S. A.* 94: 1035-1040.
Stolf-Moreira R, Medri ME, Neumaier N, Lemos NG, et al. (2010). Cloning and quantitative expression analysis of drought-induced genes in soybean. *Genet. Mol. Res.* 9: 858-867.
Sun S, Yu JP, Chen F, Zhao TJ, et al. (2008). TINY, a dehydration-responsive element (DRE)-binding protein-like transcription factor connecting the DRE- and ethylene-responsive element-mediated signaling pathways in *Arabidopsis*. *J. Biol. Chem.* 283: 6261-6271.
Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.
Thomashow MF (1999). PLANT COLD ACCLIMATION: Freezing Tolerance Genes and Regulatory Mechanisms. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50: 571-599.
Thompson JD, Higgins DG and 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.
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, et al. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876-4882.
van der Graaff E, Dulk-Ras AD, Hooykaas PJ and Keller B (2000). Activation tagging of the LEAFY PETIOLE gene affects leaf petiole development in *Arabidopsis thaliana*. *Development* 127: 4971-4980.
Wang QJ, Xu KY, Tong ZG, Wang SH, et al. (2010). Characterization of a new dehydration responsive element binding factor in central arctic cowberry. *Plant Cell Tiss. Org.* 101: 211-219.

Yamaguchi-Shinozaki K and Shinozaki K (1994). A novel cis-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low-temperature, or high-salt stress. *Plant Cell* 6: 251-264.

Zhang G, Chen M, Chen X, Xu Z, et al. (2008). Phylogeny, gene structures, and expression patterns of the *ERF* gene family in soybean (*Glycine max* L.). *J. Exp. Bot.* 59: 4095-4107.

Zhuang J, Cai B, Peng RH, Zhu B, et al. (2008). Genome-wide analysis of the AP2/ERF gene family in *Populus trichocarpa*. *Biochem. Biophys. Res. Commun.* 371: 468-474.

