



Genome-wide identification and phylogenetic analysis of the AP2/ERF gene superfamily in sweet orange (*Citrus sinensis*)

T.M. Ito, P.B. Polido, M.C. Rampim, G. Kaschuk and S.G.H. Souza

Laboratório de Biologia Molecular, Universidade Paranaense,
Umuarama, PR, Brasil

Corresponding author: S.G.H. Souza
E-mail: silviahulse@unipar.br

Genet. Mol. Res. 13 (3): 7839-7851 (2014)

Received November 14, 2013

Accepted March 7, 2014

Published September 26, 2014

DOI <http://dx.doi.org/10.4238/2014.September.26.22>

ABSTRACT. Sweet orange (*Citrus sinensis*) plays an important role in the economy of more than 140 countries, but it is grown in areas with intermittent stressful soil and climatic conditions. The stress tolerance could be addressed by manipulating the ethylene response factor (ERF) transcription factors because they orchestrate plant responses to environmental stress. We performed an *in silico* study on the ERFs in the expressed sequence tag database of *C. sinensis* to identify potential genes that regulate plant responses to stress. We identified 108 putative genes encoding protein sequences of the AP2/ERF superfamily distributed within 10 groups of amino acid sequences. Ninety-one genes were assembled from the ERF family containing only one AP2/ERF domain, 13 genes were assembled from the AP2 family containing two AP2/ERF domains, and four other genes were assembled from the RAV family containing one AP2/ERF domain and a B3 domain. Some conserved domains of the ERF family genes were disrupted into a few segments by introns. This irregular distribution of genes in the AP2/ERF superfamily in different plant species could be a result of genomic losses or duplication events in a common ancestor. The *in silico*

gene expression revealed that 67% of AP2/ERF genes are expressed in tissues with usual plant development, and 14% were expressed in stressed tissues. Because the AP2/ERF superfamily is expressed in an orchestrated way, it is possible that the manipulation of only one gene may result in changes in the whole plant function, which could result in more tolerant crops.

Key words: Expressed sequence tags; *In silico* analysis; Gene expression; Amino acid sequences

INTRODUCTION

Sweet orange (*Citrus sinensis*) is a fruit crop that plays an important role in the economy of more than 140 countries, which accumulated 66.4 million tons of oranges in 2010 (FAO, 2011). Brazil is the major worldwide producer and, together with China and the USA, produces about 50% of the sweet oranges in the world (USDA, 2012). A drawback is that sweet oranges are produced in areas that are usually subjected to intermittent oscillations in soil and climatic conditions (Yang et al., 2011), including drought and high and low temperatures. Therefore, in addition to field recommendations that may mitigate the effects of environmental stresses, orange producers may search for genetic breeding to result in more tolerant crops.

Plants have evolved to overcome several environmental changes associated with drought, waterlogging, air temperatures, plagues, grazing, and soil-borne diseases by developing a complex network of cellular, molecular, and biochemical signaling (Sharma et al., 2010; Rashid et al., 2012). Signaling and biological control of tolerance to environmental stresses are governed by genetic regulation at the transcriptional level and are mainly induced by transcript factors (Jofuku et al., 1994; Sharma et al., 2010).

Ethylene response factors (ERFs) are among the most vital transcription factors because they play important roles in plant development and regulation of plant defenses (Champion et al., 2009). ERF is involved in the transcriptional regulation of several biological processes related to growth, development, and response to environmental stresses (Sharma et al., 2010). It participates in the pathways that respond to hormones and biotic and abiotic stresses, such as jasmonate (JA), abscisic acid (ABA), drought (Liu et al., 1998; Gollmack et al., 2011), salinity (Gollmack et al., 2011), low temperature (Carvalho et al., 2011), and diseases (Li et al., 2011). Proteins of the ERF family were found in several plant species including *Arabidopsis* (Sakuma et al., 2002), rice (Nakano et al., 2006; Sharoni et al., 2011; Rashid et al., 2012), cotton (Champion et al., 2009), tomato (Sharma et al., 2010), grape (Licausi et al., 2010), and apple (Zhuang et al., 2011). The understanding of the ERF family in *C. sinensis* is still incipient, and only a few studies are found in the literature (Yang et al., 2011).

ERF genes are assembled within the APETALA2 (AP2)/ERF superfamily that encompasses the ERF, AP2, and related to ABI3/VP1 (RAV) families (Riechmann et al., 2000). The AP2 family contains two repeated AP2 domains, and the RAV family contains the AP2 domain and the B3 domain (Jofuku et al., 1994; Nakano et al., 2006). ERF family proteins contain only one AP2 domain and are subdivided into the ERF and core-binding factor/dehydration responsive element-binding (CBF/DREB) protein subfamilies (Jofuku et al., 1994; Sakuma et al., 2002; Nakano et al., 2006).

Therefore, knowing the important role of the ERF family in overcoming abiotic stresses and considering that few studies have characterized this gene family in *C. sinensis*, we performed an *in silico* study of the ERF transcription factors in the expressed sequence

tag (EST) database of *C. sinensis*. We observed that genes of the AP2/ERF family are usually expressed in tissues with normal plant development and also in tissues undergoing some kind of stress. Our analysis may contribute to the understanding of genetic and physiological mechanisms that determine tolerance to environmental stress.

MATERIAL AND METHODS

Identification and phylogenetic analysis of AP2/ERF domains in sweet orange

Protein sequences of the AP2/ERF family genes in *Arabidopsis thaliana* were identified in The *Arabidopsis* Information Resource database (<http://www.arabidopsis.org/>) and used to research sweet orange (*C. sinensis*) orthologous genes in the Citrus Genome Database (<http://www.citrusgenomedb.org/species/sinensis>) using the protein basic local alignment search tool (BLASTP) (Altschul et al., 1997). Retrieved sequences were searched against all available sequences in the GenBank database (National Center for Biotechnology Information, NCBI; <http://www.ncbi.nlm.nih.gov>) using BLASTP and translated BLAST and were arranged into amino acid sequences with the open reading frame (ORF) finder online tool (NCBI; <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). Protein sequences with incomplete AP2/ERF domains or incorrect ORFs were excluded from further analyses.

Protein sequences containing the AP2/ERF domains were aligned using ClustalW 2.0, and redundant entries were excluded. The resulting sequences were arranged into an unrooted neighbor-joining tree with MEGA 5.10 (Tamura et al., 2011) using 1000 bootstrap replicates and the pair-wise deletion option.

Conserved motif display of AP2/ERF domains in sweet orange

Conserved motifs in sweet orange AP2/ERF protein sequences were further investigated using ClustalW 2.0 and MEME Suite 4.81 (<http://meme.nbcr.net/meme>) with the following parameters: optimum width of 6-80 amino acids, any number of motifs of repetitions of a motif, and maximum number of motifs set at 25. The resulting motifs and sequence conservation of recognizable functional domains were verified and validated using protein analysis and gene function tools from databases (NCBI database, European Bioinformatics Institute-European Molecular Biology Laboratory, and Expert Protein Analysis System from the Swiss Institute of Bioinformatics).

Intron/exon distribution of AP2/ERF domains in sweet orange

The distribution of intron positions in the AP2/ERF genes was uncovered by aligning the full-length cDNA transcripts with the complete genomic DNA sequences using the Gene Structure Display Server (<http://gsds.cbi.pku.edu.cn>). For a single full-length cDNA aligned against a continuous stretch of genomic sequence, exons are proximal blocks of homologous sequence between full-length cDNA and genomic sequences, whereas introns are gaps between exons consisting entirely of genomic sequence. The general distribution of introns within each coding DNA sequence was analyzed according to the distribution of exon sizes.

Investigation of predicted AP2/ERF gene expression profiles in citrus

ESTs were used to detect AP2/ERF gene expression patterns. First, we used the pre-

dicted nucleotide sequences of the AP2/ERF genes from sweet orange as query sequences to complete a nucleotide BLAST in the NCBI EST database (<http://www.ncbi.nlm.nih.gov/nuclest>). The redundant sequences were deleted. The expression data in different tissues were retrieved by analyzing the EST source information of the obtained 408 sequences.

RESULTS

Identification and phylogenetic analysis of AP2/ERF domains in sweet orange

We identified 108 putative genes encoding protein sequences of the AP2/ERF superfamily in *C. sinensis* (Table 1; [Tables S1](#) and [S2](#)). A relatively low number of genes of this plant species contain the AP2/ERF domain compared with the number of genes in other plant species (*Arabidopsis thaliana* = 147, *Vitis vinifera* = 149, *Oryza sativa* = 180, *Prunus persica* = 131, and *Malus domestica* = 259). The majority of the AP2/ERF domain-containing genes (i.e., 91 genes) in *C. sinensis* were assigned to the ERF family because they contained only one AP2/ERF domain, 13 genes were assigned to the AP2 family because they contained two AP2/ERF domains, and four genes were assigned to the RAV family because they contained one AP2/ERF domain and a B3 domain. On the basis of amino acid sequences, 34 genes (groups I-IV) were placed within the DREB subfamily, and 49 genes (groups V-X) were placed within the ERF subfamily.

Table 1. Summary of the AP2/ERF superfamily of *Citrus sinensis* and model species *Arabidopsis thaliana*, *Vitis vinifera*, *Oryza sativa*, *Prunus persica*, and *Malus x domestica*.

Species			<i>Citrus sinensis</i>	<i>Arabidopsis thaliana</i> ^a	<i>Vitis vinifera</i> ^b	<i>Oryza sativa</i> ^a	<i>Prunus persica</i> ^c	<i>Malus x domestica</i> ^d	
Family	Subgroup	Group	(N)	(N)	(N)	(N)	(N)	(N)	
ERF	DREB	I	6	10	5	9	6	10	
		II	6	15	8	16	9	13	
		III	18	23	22	27	23	22	
		IV	4	9	5	6	7	23	
	ERF	V	8	5	11	8	11	19	
		VI	3	8	5	6	3	6	
		VI-L	2	4	2	3	4	4	
		VII	4	5	3	15	6	8	
		VIII	10	15	11	15	10	31	
		IX	22	17	40	18	19	45	
		X	8	8	10	12	6	12	
		Xb-L	0	3	0	10	0	2	
				91	122	122	145	104	195
		AP2		13	18	20	29	21	51
RAV		4	6	6	5	5	6		
Soloist		0	1	1	1	1	7		
Total		108	147	149	180	131	259		

^aNakano et al. (2006); ^bLicausi et al. (2010); ^cZhang et al. (2012); ^dGirardi et al. (2013).

These genes were further analyzed by aligning the deduced amino acid sequences of the AP2/ERF domains in *C. sinensis* and *Arabidopsis*. This alignment revealed that amino acid residues Gly-4, Arg-6, Arg-8, Glu-16, Trp-30, Leu-31, Gly-32, Ala-40, and Asn-64 were fully conserved in *C. sinensis*, and the amino acid residues Gly-11, Arg-28, Ala-41, Ala-43, Asp-45, and Gly-53 were conserved in at least 95% of the sequences (Figure 1; [Figure S1](#)). In *Arabidopsis*, Sakuma et al. (2002) observed that two main amino acid residues differentiate the ERF and DREB subfamilies: the ERF subfamily contains alanine and aspartate, and

the CBF/DREB subfamily contains valine and glutamic acid in the 14th and 19th positions, respectively. We observed that the majority of genes in the ERF subfamily contained a fully conserved alanine residue in the 14th position and aspartate, glutamic acid, histidine, or asparagine in the 19th position. Regarding the DREB subfamily, we observed that all genes contained a fully conserved valine in the 14th position and glutamic acid or leucine in the 19th position (Figure 1; [Figure S1](#)).

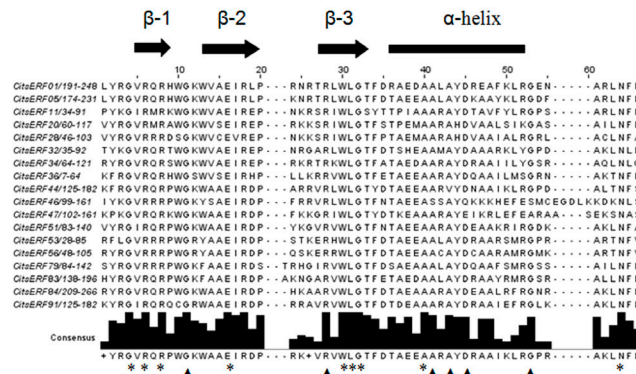


Figure 1. Alignment of the AP2/ERF domains from *Citrus sinensis* ethylene response factor proteins. Asterisks and arrowheads indicate identical and conserved amino acid residues, respectively. The black bar and arrows represent predicted α -helix and β -sheet regions, respectively.

The 91 putative genes belonging to the ERF family (DREB and ERF subfamilies, groups I-X) in citrus encoded 91 proteins and were named *CitsERF*. A clustering analysis using the neighbor-joining method and 58-59 amino acid sequences placed these 91 *CitsERFs* along a phylogenetic tree (Figure 2). The ERF proteins of *C. sinensis* were classified into 10 groups (I-X) according to their similarity to *Arabidopsis* AP2/ERF domain sequences.

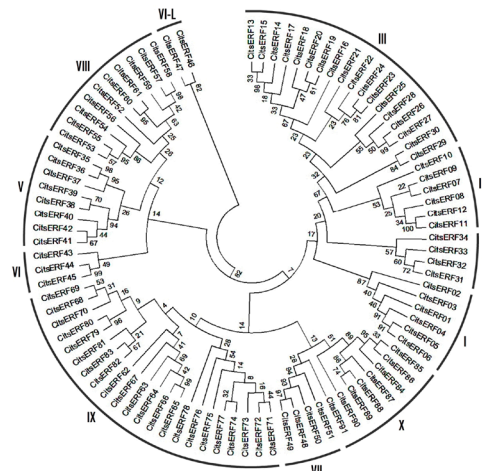


Figure 2. Phylogenetic analysis of 91 sweet orange ethylene response factor (ERF) proteins. An unrooted neighbor-joining phylogenetic tree was constructed using MEGA 5.10 for the multiple sequence alignment of 91 sweet orange ERF protein sequences. Groups are marked I to X (DREB subfamily: I, II, III, and IV; ERF subfamily: V, VI, VI-L, VII, VIII, IX, and X).

Conserved motifs of ERF family genes in sweet orange

We employed multiple sequence analyses to find conserved motifs outside the AP2 domain in ERF family genes of citrus and related them to their biological functions. The analyses indicated that the majority of protein sequences associated with the ERF family of citrus contained at least one motif in either the C- or N-terminal region. The likely locations of the motifs are shown in Figure 3.

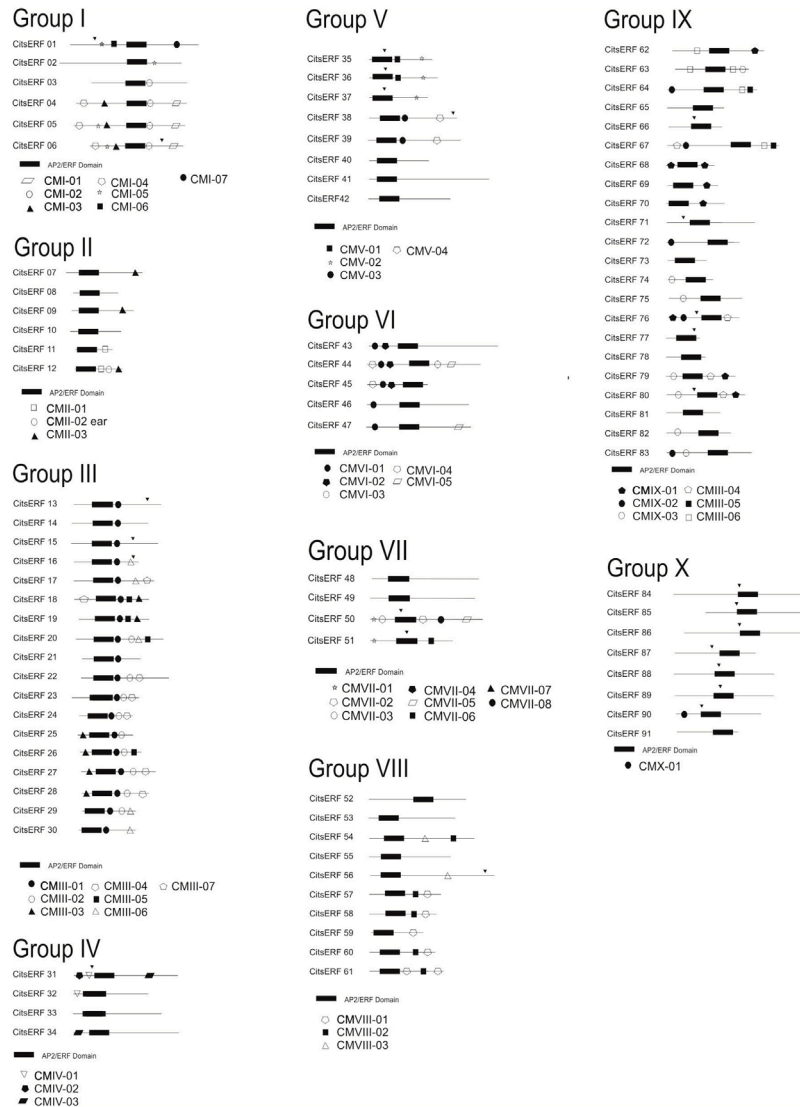


Figure 3. Distribution of conserved motifs in the ethylene response factor (ERF) family in *Citrus sinensis*. Conserved motifs are marked with different signs, as indicated below the motifs. Each black box represents an AP2/ERF domain. The positions of introns are marked with arrowheads.

In the DREB subfamily, group I (*CitsERF01* to *CitsERF06*) contained seven conserved motifs, from which, *CitsERF03* and *CitsERF06* had the CMI-02 motif in the C-terminal region. In group II (*CitsERF07* to *CitsERF12*), *CitsERF07*, *CitsERF09*, and *CitsERF12* contained the CMII-03 motif in the C-terminal region, and *CitsERF12* also contained a CMII-02 ERF-associated amphiphilic repression (EAR) and CMII-03 motifs in the C-terminal regional. This motif is illustrated in Figure 4A. Group III is a larger group, comprising 18 ERF genes (*CitsERF13* to *CitsERF30*). All genes of this group contained the CMIII-01 motif in the C-terminal region; *CitsERF22*, *CitsERF23*, *CitsERF24*, *CitsERF27*, and *CitsERF28* contained the CMIII-4-LWS motif in the C-terminal region; and *CitsERF20* and *CitsERF22* to *CitsERF29* contained the CMIII-02 motif. In group IV (*CitsERF31* to *CitsERF34*), only *CitsERF31* contained a motif, the CMIV-02 motif.

In the ERF subfamily, group V (*CitsERF35* to *CitsERF42*) contained the CMV-02 motif that was associated with *CitsERF35* to *CitsERF37*, the CMV-01 motif that was associated with *CitsERF35* and *CitsERF36*, and the CMV-03 and CMV-04 motifs that were associated with *CitsERF38* and *CitsERF39*; all motifs were in the C-terminal region. All genes of group VI (*CitsERF43* to *CitsERF47*) contained the CMVI-01 motif in the N-terminal region. *CitsERF43*, *CitsERF44*, and *CitsERF45* contained the CMVI-02 motif in the N-terminal region; *CitsERF44* and *CitsERF47* contained the CMVI-05 motif in the N-terminal region; and *CitsERF44* was the only gene to have the CMVI-03 motif in the N-terminal region. In group VII (*CitsERF48* to *CitsERF51*), *CitsERF50* and *CitsERF51* contained the CMVII-01 motif in the N-terminal region and the CMVII-02 motif in the C-terminal region. In addition, *CitsERF50* contained three more motifs in the C-terminal region, CMVII-03, CMVII-05, and CMVII-08. In group VIII (*CitsERF52* to *CitsERF61*), *CitsERF57* to *CitsERF61* had the CMVIII-1/EAR motif, which is related to amphiphilic repression, in the N-terminal region. The CMVIII-1/EAR motif is illustrated in Figure 4B. *CitsERF54*, *CitsERF57*, *CitsERF58*, *CitsERF60*, and *CitsERF61* contained the CMVIII-02 motif, and *CitsERF54* and *CitsERF56* had the CMVIII-03 motif in the C-terminal region. In group IX (*CitsERF62* to *CitsERF83*), six motifs were found: CMIX-01 to CMIX-06; all of the motifs were in the C-terminal region. Within this group, *CitsERF62*, *CitsERF68*, *CitsERF69*, *CitsERF70*, *CitsERF79*, and *CitsERF80* were associated with the CMIX-01 motif; *CitsERF64*, *CitsERF67*, *CitsERF72*, *CitsERF76*, and *CitsERF83* were associated with the CMIX-02 motif; *CitsERF63*, *CitsERF74*, *CitsERF75*, *CitsERF79*, *CitsERF80*, *CitsERF82*, and *CitsERF83* were associated with the CMIX-03 motif; *CitsERF76*, *CitsERF79*, and *CitsERF80* were associated with the CMIX-04 motif; *CitsERF64* and *CitsERF67* were associated with the CMIX-05 motif; and *CitsERF62*, *CitsERF63*, *CitsERF64*, and *CitsERF67* were associated with the CMIX-06 motif. In group X (*CitsERF84* to *CitsERF91*), only *CitsERF90* contained a motif, the CMX-01 motif, in the N-terminal region.

Intron/exon structure of ERF family genes in citrus

We determined the size and distribution of exons in the ERF family genes in *C. sinensis* by aligning the full-length cDNA transcripts with the complete genomic DNA sequences. This type of alignment may provide information about the evolutionary relationships of the genome. The analysis indicated that the exon lengths ranged from 138.4 to 274.79 bp with the average size of 205.3 bp (standard deviation = 68.96). We observed that 26 of the 91 genes (28.6%) had

only one intron in each gene; these genes were distributed in the following groups: I, III, IV, V, VII, VIII, IX, and X (Figure 3). Some conserved domains of the ERF family genes were disrupted into a few segments by the introns. Furthermore, the number and position of the introns were relatively conserved in each group.

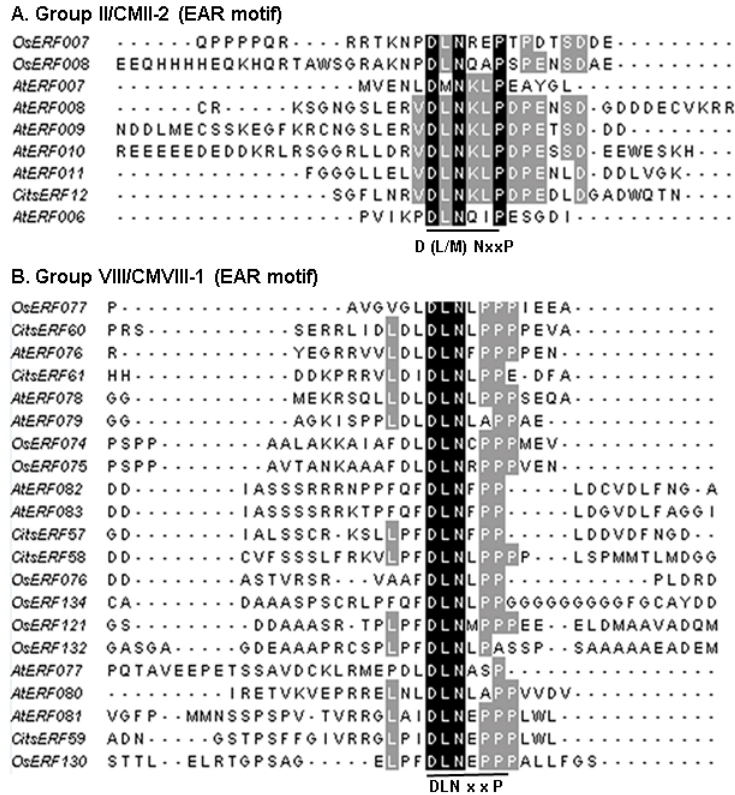


Figure 4. EAR motif-like sequences conserved in group II and VIII ERF proteins. **A.** Sequence alignment of C-terminal regions of group II proteins. **B.** Sequence alignment of C-terminal regions of group VIII proteins. Conserved motifs are underlined. Black and gray shadings indicate identical and conserved amino acid residues present in >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.

Expression profiles of predicted AP2/ERF genes in citrus

We identified 408 AP2/ERF ESTs from the citrus NCBI EST database for our *in silico* gene expression analysis. Results from our analysis are depicted in Figure 5. It is noteworthy that the AP2/ERF superfamily genes are expressed in tissues related to plant development (67% of the expression was in fruit, leaves, immature ovaries, flowers, seed, meristem, phloem, and roots), but it may also be stimulated by environmental stress events (14% of the expression was in *Xylella* infection, cold stress, pathogen attack, and insect attack). It also appears that the AP2/ERF genes are related to flavedo and albedo (7%), callus formation (7%), and other processes (5%).

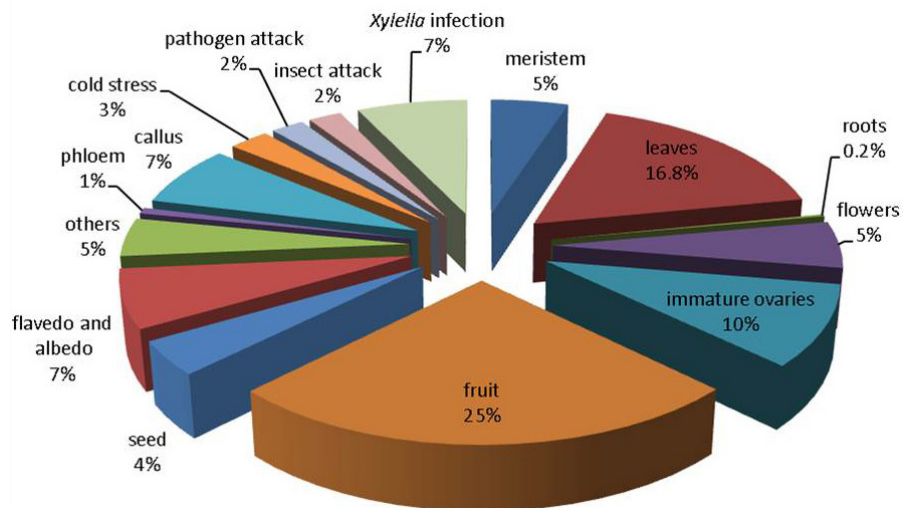


Figure 5. *In silico* expression of the ERF family genes in different *Citrus sinensis* tissues.

DISCUSSION

The AP2/ERF superfamily transcription factors orchestrate the pathways of plant development and plant responses to biotic and abiotic stresses (Sakuma et al., 2002). The manipulation of the complex metabolic pathways governed by the AP2/ERF superfamily may represent an interesting approach in breeding strategies to generate crops that are tolerant to intermittent oscillations in soil and climatic conditions.

Earlier studies identified genes belonging to the AP2/ERF superfamily in other plant species (Table 1), usually using *A. thaliana* as a plant model for classification (Nakano et al., 2006). Considering *C. sinensis*, we identified 108 genes encoding transcription factors of the AP2/ERF superfamily, while much larger numbers of genes were identified in other plant species (Table 1). However, the low number of AP2/ERF genes in *C. sinensis* is not associated with the size of the genome itself. Although *C. sinensis* has a much larger genome (367-396 Mbp) (Arumuganathan and Earle, 1991) than *Arabidopsis* (145 Mbp), the structure and phylogenic arrangements are quite similar. Indeed, other studies have shown that the AP2/ERF superfamily in soybeans (with a genome size of 1115 Mbp) (Zhang et al., 2008) appeared to be quite similar to that found in rice (430 Mbp) (Rashid et al., 2012), grape (475 Mbp) (Licausi et al., 2010), and cucumber (367 Mbp) (Hu and Liu, 2011). In fact, any irregular distribution of genes in the AP2/ERF superfamily among different plant species may have evolved from genomic losses or duplication events of a common ancestor, contributing to the expansion or restriction of the number of these genes (Hu and Liu, 2011; Rashid et al., 2012; Zhang et al., 2012).

The alignment analysis revealed that ERF proteins contain conserved Ala-14 and Asp/His/Glu/Asn-19 residues, whereas DREB proteins contain conserved Val-14 and Glu/Leu-19 residues. These highly conserved amino acid residues play an important role in the ERF gene functions, and they are probably involved in different forms of contact with the DNA. It was already known that the AP2/ERF domain recognizes DNA through the conserved arginine and tryptophan residues located in β -sheets (Allen et al., 1998). Likewise, Ala-37 in the ERF

domain plays an important role in the stability of the ERF domain and in the binding of the DRE element or GCC-box to the DNA (Liu et al., 2006).

We confirmed the congruence of 10 groups of ERF family genes in *Arabidopsis* and *C. sinensis*, showing that the methodology of Nakano et al. (2006) is also suitable for *C. sinensis*. Similarities in the AP2/ERF superfamily of different plant species imply that they have common gene ancestors. For instance, there is evidence that most groups and subgroups of genes within the dicotyledons and between dicotyledons and monocotyledons are alike, suggesting that many genes existed much before the diversification of plants into dicotyledons and monocotyledons (Sharma et al., 2010; Zhang et al., 2012). In the same way, some groups and subgroups of genes were present in one plant species but not in the other. For example, groups XI-XIV occurred only in the ERF family in rice, but not in *Arabidopsis* and other dicotyledons (Zhang et al., 2008), suggesting that they evolved or were lost after diversification (Zhang et al., 2008). In this study, most of the genes with an AP2 domain were conserved within groups III and IX. Nakano et al. (2006) pointed out that these groups play an important role in the plant tolerance to abiotic and biotic stress in different plant species. Therefore, a relatively large number of genes within these groups should be attributed to several adaptive strategies to environmental changes over thousands of years.

Some motifs are located out of the binding region and play important functions for the transcription factors. These motifs represent functional domains and are often conserved between members of groups or subgroups of larger families of transcription factors that share similar functions (Nakano et al., 2006). By analyzing the distribution of motifs along the phylogenetic tree of ERF transcription factors in *C. sinensis*, we observed that the motifs of *C. sinensis* that were located out of the binding region of the AP2 domain share structural similarities with their homologs in *Arabidopsis*, rice, and soybean (Nakano et al., 2006; Zhang et al., 2008). Ohta et al. (2001) identified an ERF motif that was associated with amphiphilic repression and named it EAR. The EAR motif has a conserved sequence [(L/F)DLN(L/F)xP] in the C-terminal region of repressing ERF proteins and plays important roles in several biological functions, downregulating genes involved in different pathways of hormonal and stress signaling (Fujimoto et al., 2000; Ohta et al., 2001). In this study, the EAR motif was only found in the proteins of groups II and VIII, and was identified as CMII-2 EAR and CMVIII-1, respectively (Figures 3 and 4).

Two other members of group VIII in *Arabidopsis*, AtERF4 and AtERF7, were characterized (Ohta et al., 2001). Both AtERF4 and AtERF7 downregulate the gene *PDF1.2*, which contains a GCC-box in its promoter region. Previous studies showed that AtERF4 downregulated the expression of ethylene-, JA-, and ABA-responsive genes (Ohta et al., 2001). The DREB protein containing the EAR motif, named DEAR1, functioned as a repressor of the binding proteins related to the dehydration response element that mediates the responses to abiotic and biotic stresses (Tsutsui et al., 2009).

The CMIV-2 motif in the N-terminal region of group IV probably functions as a nuclear location signal (El Kayal et al., 2006) in a similar way as the nuclear transport signal that was previously described by Stockinger et al. (1997), which is involved in protein trafficking. Recently, the CMIV-2 motif was described as an essential factor in *Arabidopsis*, where it participates in the CBF1 binding to the DNA, being indispensable for its transcriptional activity (Canella et al. 2010). The CMIX-2 and CMIX-3 motifs of group IX are putative acid regions that may function as transcription activator domains (Fujimoto et al., 2000). A comparative analysis of the motifs in *C. sinensis* and *Arabidopsis* suggested that the protein functions are

conserved in the ERF family. In fact, most of the motifs are distributed in specific clades in the phylogenetic tree, showing structural and functional similarities between members of the same group (Figures 2 and 3).

In a previous study, Sakuma et al. (2002) reported that most genes in the ERF family of *Arabidopsis* contain hardly any introns, and they identified only one intron in four genes. Later, Nakano et al. (2006) determined that this intron in the ERF family was conserved in groups V, VII, X, and Xb-L. The “introns-early” hypothesis advocates that the localization of introns may be conserved between different species if the introns already existed in a common ancestor of monocotyledons and dicotyledons (Rogozin et al., 2005). Therefore, as far as the “introns-early” hypothesis is concerned, there is significant similarity between the AP2/ERF genetic structures of *Arabidopsis* and *C. sinensis*. Of the 91 *CitsERF* genes that we identified, 25 genes contained an intron in the same location as its ortholog in *Arabidopsis* (Figure 3).

The gene expression patterns of the 91 genes of the ERF family in *C. sinensis* were investigated based on their EST annotations (Figure 5). The *in silico* analysis revealed that 25% of ESTs were expressed in fruit libraries. A number of physical and chemical parameters, including ethylene, stimulated the regulation of ERF transcription factors, suggesting that they are also involved in the crosstalk of fruit development and ripening (Guo and Ecker, 2004). Tournier et al. (2003) demonstrated that *SlERF2*, an ERF that binds a GCC-box, plays an important role during the ripening of tomato. Similar patterns were observed for ERFs expressed during the ripening of different fruits (Xiao et al., 2013). The overexpression of *LeERF1* shortens the fruit postharvest life (Li et al., 2007), while *LeERF2* modulates ethylene biosynthesis in tomato and tobacco. Therefore, evidence that ERFs are associated with flavor and fruit texture during ripening is increasing (Licausi et al., 2010).

Interestingly, Figure 5 shows that the AP2/ERF superfamily genes were expressed more often in healthy plant tissues (67% of the time) than in stressed tissues (14%). In fact, although the genetic sequences of *C. sinensis* are complete in the database, we cannot be sure that all stressful possibilities have been considered. However, despite the fact that the gene expression of the ERF family was associated with regular plant metabolism, there are important insights regarding its role in plant stress tolerance.

A relevant study showed that several ERF proteins induce plant tolerance to biotic and abiotic stresses (Guo et al., 2004). Our *in silico* analysis revealed that only 14% of ESTs were expressed in libraries of tissues submitted to some kind of stress. These genes contain an AP2/ERF domain that may bind to either a GCC-box or a DRE with a motif core of A/GCCGAC (Liu et al., 1998). This implies that the signaling pathways mediated by DREB and ERF are not exclusive of each other, but that they work in crosstalk between themselves (Zhang et al., 2008). The overexpression of the DREB genes in *Arabidopsis* activates the expression of some other genes related to drought, low temperatures, and salt tolerance (Liu et al., 1998; Sakuma et al., 2006). The overexpression of *AoDREB* of *Asparagus officinalis* L. in transgenic *Arabidopsis* elicited the expression of *rd29A* and *COR15A*, resulting in better plant tolerance to salt stress (Liu et al., 2010). Further studies should consider these genes as promising candidates for improving the mechanisms to regulate the ERF genes in *C. sinensis* in different stresses.

Genetic evolution has selected traits that allow plants to survive under several stressful soil and climatic conditions. This study provided some information about the AP2/ERF superfamily in *C. sinensis*, which is related to responses to environmental stresses and may become useful as tools to engineer stress-tolerant plants. Because the AP2/ERF superfamily is expressed in an orchestrated way, it is possible that manipulating only one gene may cause changes in the whole plant function, which could result in more tolerant crops.

ACKNOWLEDGMENTS

Research supported by UNIPAR (Project #24394/2013).

[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.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.
- Arumuganathan K and Earle ED (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9: 208-218.
- Canella D, Gilmour SJ, Kuhn LA and Thomashow MF (2010). DNA binding by the Arabidopsis CBF1 transcription factor requires the PKKP/RAGRxKFXETRHP signature sequence. *Biochim. Biophys. Acta* 1799: 454-462.
- Carvalho MA, Pino MT, Jeknic Z, Zou C, et al. (2011). A comparison of the low temperature transcriptomes and CBF regulons of three plant species that differ in freezing tolerance: *Solanum commersonii*, *Solanum tuberosum*, and *Arabidopsis thaliana*. *J. Exp. Bot.* 62: 3807-3819.
- Champion A, Hebrard E, Parra B, Bournaud C, et al. (2009). Molecular diversity and gene expression of cotton ERF transcription factors reveal that group IXa members are responsive to jasmonate, ethylene and *Xanthomonas*. *Mol. Plant Pathol.* 10: 471-485.
- El Kayal W, Navarro M, Marque G, Keller G, et al. (2006). Expression profile of CBF-like transcriptional factor genes from *Eucalyptus* in response to cold. *J. Exp. Bot.* 57: 2455-2469.
- FAO (Food and Agriculture Organization of the United Nations) (2011). Citrus Fruit. Available at [http://www.fao.org/docrep/006/y5143e/y5143e12.htm]. Accessed November 14, 2012.
- Fujimoto SY, Ohta M, Usui A, Shinshi H, et al. (2000). *Arabidopsis* ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant Cell* 12: 393-404.
- Girardi CL, Rombaldi CV, Dal Cero J, Nobile PM, et al. (2013). Genome-wide analysis of the AP2/ERF superfamily in apple and transcriptional evidence of ERF involvement in scab pathogenesis. *Sci. Hortic.* 151: 112-121.
- Golldack D, Luking I and Yang O (2011). Plant tolerance to drought and salinity: stress regulating transcription factors and their functional significance in the cellular transcriptional network. *Plant Cell Rep.* 30: 1383-1391.
- Guo H and Ecker JR (2004). The ethylene signaling pathway: new insights. *Curr. Opin. Plant Biol.* 7: 40-49.
- Guo ZJ, Chen XJ, Wu XL, Ling JQ, et al. (2004). Overexpression of the AP2/EREBP transcription factor OPBP1 enhances disease resistance and salt tolerance in tobacco. *Plant Mol. Biol.* 55: 607-618.
- Hu L and Liu S (2011). Genome-wide identification and phylogenetic analysis of the ERF gene family in cucumbers. *Genet. Mol. Biol.* 34: 624-633.
- Jofuku KD, den Boer BG, Van Montagu M and Okamoto JK (1994). Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. *Plant Cell* 6: 1211-1225.
- Li CW, Su RC, Cheng CP, Sanjaya, et al. (2011). Tomato RAV transcription factor is a pivotal modulator involved in the AP2/EREBP-mediated defense pathway. *Plant Physiol.* 156: 213-227.
- Li YC, Zhu BZ, Xu WT, Zhu HL, et al. (2007). LeERF1 positively modulated ethylene triple response on etiolated seedling, plant development and fruit ripening and softening in tomato. *Plant Cell Rep.* 26: 1999-2008.
- Licausi F, Giorgi FM, Zenoni S, Osti F, et al. (2010). Genomic and transcriptomic analysis of the AP2/ERF superfamily in *Vitis vinifera*. *BMC Genom.* 11: 719.
- Liu Q, Kasuga M, Sakuma Y, Abe H, et al. (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.
- Liu Y, Chen H, Zhuang D, Jiang D, et al. (2010). Characterization of a DRE-binding transcription factor from asparagus (*Asparagus officinalis* L.) and its overexpression in *Arabidopsis* resulting in salt- and drought-resistant transgenic plants. *Int. J. Plant Sci.* 171: 12-23.
- Liu Y, Zhao TJ, Liu JM, Liu WQ, et al. (2006). The conserved Ala37 in the ERF/AP2 domain is essential for binding with the DRE element and the GCC box. *FEBS Lett.* 580: 1303-1308.
- 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.
- Rashid M, Guangyuan H, Guangxiao Y, Hussain J, et al. (2012). AP2/ERF transcription factor in rice: Genome-wide canvas and syntenic relationships between monocots and eudicots. *Evol. Bioinform. Online* 8: 321-355.
- Riechmann JL, Heard J, Martin G, Reuber L, et al. (2000). *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. *Science* 290: 2105-2110.
- Rogozin IB, Wolf YI, Babenko VN and Koonin EV (2005). Dollo Parsimony and the Reconstruction of Genome Evolution. In: Parsimony, Phylogeny, and Genomics (Albert VA, ed.). Oxford University Press, Oxford, 190-200.
- 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.
- Sakuma Y, Maruyama K, Qin F, Osakabe Y, et al. (2006). Dual function of an *Arabidopsis* transcription factor DREB2A in water-stress-responsive and heat-stress-responsive gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 103: 18822-18827.
- Sharma MK, Kumar R, Solanke AU, Sharma R, et al. (2010). Identification, phylogeny, and transcript profiling of ERF family genes during development and abiotic stress treatments in tomato. *Mol. Genet. Genom.* 284: 455-475.
- Sharoni AM, Nuruzzaman M, Satoh K, Shimizu T, et al. (2011). Gene structures, classification and expression models of the AP2/EREBP transcription factor family in rice. *Plant Cell Physiol.* 52: 344-360.
- 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.
- Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.
- Tournier B, Sanchez-Ballesta MT, Jones B, Pesquet E, et al. (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.
- Tsutsui T, Kato W, Asada Y, Sako K, et al. (2009). DEAR1, a transcriptional repressor of DREB protein that mediates plant defense and freezing stress responses in *Arabidopsis*. *J. Plant Res.* 122: 633-643.
- USDA (United States Department of Agriculture) (2012). Reports. Available at [<http://www.usdabrazil.org.br/home/reports.asp>]. Accessed November 14, 2012.
- Xiao YY, Chen JY, Kuang JF, Shan W, et al. (2013). Banana ethylene response factors are involved in fruit ripening through their interactions with ethylene biosynthesis genes. *J. Exp. Bot.* 64: 2499-2510.
- Yang L, Hu CH, Li N, Zhang JY, et al. (2011). Transformation of sweet orange [*Citrus sinensis* (L) Osbeck] with *PthA-nls* for acquiring resistance to citrus canker disease. *Plant Mol. Biol.* 75: 11-23.
- Zhang CH, Shangguan LF, Ma RJ, Sun X, et al. (2012). Genome-wide analysis of the AP2/ERF superfamily in peach (*Prunus persica*). *Genet. Mol. Res.* 11: 4789-4809.
- 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, Yao QH, Xiong AS and Zhang J (2011). Isolation, phylogeny and expression patterns of AP2-like genes in apple (*Malus x domestica* Borkh). *Plant Mol. Biol. Rep.* 29: 209-216.