

Genome-wide analysis of the AP2/ERF superfamily in peach (*Prunus persica*)

C.H. Zhang¹, L.F. Shangguan², R.J. Ma¹, X. Sun², R. Tao², L. Guo¹,
N.K. Korir² and M.L. Yu¹

¹Institute of Horticulture, Jiangsu Academy of Agricultural Sciences,
Nanjing, Jiangsu, China

²College of Horticulture, Nanjing Agricultural University, Nanjing,
Jiangsu, China

Corresponding author: M.L. Yu
E-mail: mly1008@yahoo.com.cn

Genet. Mol. Res. (2012) **Ahead of Print**

Received March 27, 2012

Accepted June 27, 2012

Published October 17, 2012

DOI <http://dx.doi.org/10.4238/2012.October.17.6>

ABSTRACT. We identified 131 *AP2/ERF* (*APETALA2/ethylene-responsive factor*) genes in material from peach using the gene sequences of AP2/ERF amino acids of *Arabidopsis thaliana* (Brassicaceae) as probes. Based on the number of AP2/ERF domains and individual gene characteristics, the AP2/ERF gene superfamily in peach can be classified broadly into three families, ERF (ethylene-responsive factor), RAV (related to ABI3/VP1), and AP2 (*APETALA2*), containing 104, 5, and 21 members, respectively, along with a solo gene (ppa005376m). The 104 genes in the ERF family were further divided into 11 groups based on the group classification made for *Arabidopsis*. The scaffold localizations of the AP2/ERF genes indicated that 129 *AP2/ERF* genes were all located on scaffolds 1 to 8, except for two genes, which were on scaffolds 17 and 10. Although the primary structure varied among AP2/ERF superfamily proteins, their tertiary structures were similar. Most ERF family genes have no introns, while members of the AP2 family have more introns than genes in the ERF and RAV families. All sequences of AP2 family genes were disrupted by introns into several segments of varying sizes. The expression of the *AP2/ERF* superfamily genes was highest in the mesocarp; it was far higher than in the other

seven tissues that we examined, implying that *AP2/ERF* superfamily genes play an important role in fruit growth and development in the peach. These results will be useful for selecting candidate genes from specific subgroups for functional analysis.

Key words: Peach; AP2/ERF superfamily; Phylogenetic analysis; Characteristics

INTRODUCTION

The AP2/ERF superfamily is one of the largest groups of transcription factors (TFs) in plants. The AP2/ERF superfamily is defined by the AP2/ERF domain that consists of about 60 to 70 amino acids and is involved in DNA binding. In general, the AP2/ERF superfamily is further divided into 3 main families on the basis of the number of AP2/ERF domains and sequence similarities. AP2 (APETALA2) family proteins contain 2 repeated AP2/ERF domains: 1) the ERF (ethylene-responsive factor) family proteins, containing a single AP2/ERF domain, and 2) the RAV (related to ABI3/VP1) family proteins containing 2 different DNA-binding domains, AP2/ERF and B3 (Sakuma et al., 2002; Nakano et al., 2006).

It has been demonstrated that AP2/ERF proteins are involved in a variety of regulatory mechanisms throughout the plant life cycle. These proteins play significant roles in the transcriptional regulation of a variety of biological processes related to growth and development, as well as various plant biotic and abiotic stress responses. The AP2 domain was first identified as a repeated motif within the *Arabidopsis thaliana* AP2 protein (Jofuku et al., 1994). Genes in the AP2 family have been reported to be involved in the regulation of a variety of developmental processes, from flowering to vegetative growth. The processes include flower and seed development (Jofuku et al., 1994), spikelet meristem determinacy (Chuck et al., 1998), leaf epidermal cell identity (Moose and Sisco, 1996), and embryogenesis (Guillaumot et al., 2008). The ERF domain was first identified as a conserved motif in 4 DNA-binding proteins from tobacco interaction with a GCC box (core sequence: AGCCGCC), which is a DNA sequence involved in the ethylene-responsive transcription of genes (Ohme-Takagi and Shinshi, 1995). Subsequently, proteins in the ERF family of various plants were identified and implicated in a series of biological events, such as seed germination (Wang et al., 2008) and various stress-related stimuli (Yamamoto et al., 1999; Agarwal et al., 2006; Agarwal et al., 2010). In the case of the RAV family, ERFs were first identified in *A. thaliana* (Kagaya et al., 1999). The roles of the RAV family in the ethylene and brassinosteroid response have also been documented (Alonso et al., 2003; Hu et al., 2004). Until now, a variety of AP2/ERF genes have been successfully identified and analyzed in various plant species, such as *Arabidopsis* (Nakano et al., 2006), rice (Nakano et al., 2006; Sharoni et al., 2011), grapevine (Velasco et al., 2007), and soybean (Zhang et al., 2008).

Peach (*Prunus persica* (L.) Batch) is a fruit with major worldwide commercial implications and is one of the most genetically well-characterized species of the Rosaceae family (Layne and Bassi, 2008). The distinct advantages of peach make it suitable as a model genome species for *Prunus*, as well as for other species in the Rosaceae. Since the release of the peach genome sequence in 2010, current estimates indicate that peach has 28,689 transcripts and 27,852 genes (Jung et al., 2008). However, the biological functions of only a few of these genes are well known. Currently, the use of molecular-assisted selection in conjunction with conventional breeding techniques is an accepted strategy for breeding new peach cultivars

with enhanced biotic and abiotic stress resistance, nutritional value, and other novel traits (Layne and Bassi, 2008). The genes of the AP2/ERF superfamily play significant roles in the transcriptional regulation of a variety of biological processes related to growth and development, as well as various plant biotic and abiotic stress responses. To our knowledge, a report on the comprehensive analysis of AP2/ERF superfamily genes in peach is not available.

On the basis of the highly conserved sequences in the AP2/ERF domain of AP2/ERF superfamily, a computational analysis to identify AP2/ERF genes in peach appears feasible. In this study, the identification and characterization of AP2/ERF superfamily genes on the basis of the current availability of a large public Genome Database for Rosaceae (GDR) and the Arabidopsis Information Resource (TAIR) database was attempted. According to phylogenetic and protein motif structural analyses, the families of the AP2/ERF subfamily and ERF family groups in peach were classified using the AP2/ERF superfamily classification of *Arabidopsis* as a reference. In addition, the expression percentages of AP2/ERF superfamily genes in different tissues or organs were analyzed. The data generated from this study will contribute to studies on the selection of appropriate candidate genes from the AP2/ERF superfamily in peach for further functional characterization and understanding of the precise regulatory checkpoints that operate during developmental and stress responses.

MATERIAL AND METHODS

Isolation of predicted AP2/ERF genes in peach

On the basis of the proteins identified in AP2/ERF superfamily genes of *A. thaliana* downloaded from the TAIR database (Huala et al., 2001), the hidden Markov model (HMM) profile of the AP2/ERF superfamily was extracted from the Pfam, as described by Wang et al. (2010). With the aid of the HMM profile, a search against the peach proteins database in GDR was performed using the identified proteins of the AP2/ERF superfamily of *A. thaliana* as query sequences. This resulted in 131 amino acid sequences of AP2/ERF proteins being obtained, after removing redundant sequences. To confirm these predicted AP2/ERF superfamily proteins, the protein sequences were then searched for AP2/ERF domains by using InterProScan (<http://www.ebi.ac.uk/Tools/InterProScan/>) and CD search (Marchler-Bauer and Bryant, 2004; <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) web servers. The numbers, names, and locations of the conserved domains contained in each AP2/ERF protein sequence of peach were also recorded. Finally, CDS sequences corresponding to the predicted AP2/ERF superfamily proteins were extracted from total CDS sequences of peach downloaded from the GDR database (Jung et al., 2008).

Motif display and phylogenetic analysis of predicted AP2/ERF proteins in peach

The online tool of MEME (version 4.8.1) was used to search the conserved motifs shared by AP2/ERF proteins (<http://meme.nbcr.net/meme/cgi-bin/meme.cgi>), through uploading the file of amino acid sequences from the AP2/ERF superfamily in peach (Bailey et al., 2006). Parameters were set as described by Wang et al. (2010): 0 or 1 occurrence of a single motif per sequence, motif width ranges of 10 to 300 amino acids, and 5 as the maximum number of motifs that must be found. All other parameters were set at default. The amino acid sequences of the AP2/ERF superfamily in *Arabidopsis* were downloaded from the TAIR database (<http://www.arabidopsis.org/browse/genefamily/index.jsp>), and a multiple alignment analysis between peach

and *Arabidopsis* AP2/ERF proteins (amino acid sequences) was conducted using the Clustal W program built in the MEGA 4.1 software (Tamura et al., 2007). A phylogenetic tree of AP2/ERF superfamily proteins in *Arabidopsis* and peach was constructed using the Clustal W tool, in conjunction with the MEGA 4.1 software, by using neighbor joining method and a bootstrap of 1000 replicates. Using a combination of phylogenetic trees and conserved domain analysis with the characteristics and structures of genes, the AP2/ERF superfamily in peach was classified into several families. The ERF family of peach was further classified into smaller groups on the basis of the ERF group classification of *Arabidopsis* (Nakano et al., 2006).

Gene characteristics and structure analysis of predicted AP2/ERF genes in peach

The genomic sequences (peach v1.0), ID number, gene distribution on scaffold, and genome location of peach AP2/ERF genes were downloaded and obtained from Phytozome database (<http://www.phytozome.net/peach.php>). The open reading frame (ORF) length of peach AP2/ERF genes was analyzed using the ORF Finder in NCBI. A structural figure of peach AP2/ERF genes, including exon and intron numbers and locations as well as conserved domain locations, was constructed and displayed using the Gene Structure Display Server (GSDS) web-based bioinformatics tool (<http://gsds.cbi.pku.edu.cn/>).

Characteristics of predicted AP2/ERF proteins in peach

The basic physical and chemical characteristics (primary structure) of peach AP2/ERF proteins were calculated using the online ProtParam tool (<http://www.expasy.org/tools/protparam.html>), including the number of amino acids, molecular weight, theoretical isoelectric point (pI), aliphatic index, and gravity. Analysis of the tertiary structure was completed on the online server FUGUE (version 2.0) (<http://tardis.nibio.go.jp/fugue/prfsearch.html>), which automatically constructs amino acid sequences form tertiary protein structures through homology-modeling (Shi et al., 2001). Swiss-PdbViewer (version 4.0.1) was used to analyze and display structure pictures (Arnold et al., 2006).

Investigation of predicted AP2/ERF gene expression profiles in peach

Expressed sequence tags (ESTs) were used to detect AP2/ERF gene expression patterns. First, we used the predicted nucleotide sequences of AP2/ERF from peach as query sequences to complete a nucleotide blast in the NCBI EST database (<http://www.ncbi.nlm.nih.gov/blast.cgi>). As a result, 12,740 sequences were obtained. Further mining was completed to screen the sequences that had a score value of >90 and >100% with maximum identity (Shangguan et al., 2011). Subsequently, the redundant sequences were deleted, and 322 final sequences were obtained. The expression data in different tissues were retrieved from analyzing the EST source information of the obtained 322 sequences.

RESULTS

Identification of predicted AP2/ERF genes in peach

To identify the AP2/ERF superfamily genes in peach, BLAST searches were performed

Analysis of AP2/ERF superfamily in peach

in the GDR database by using AP2/ERF proteins of *Arabidopsis* as query sequences. In total, 131 genes were identified as potential encoding AP2/ERF domains in peach. InterProScan results demonstrated that 14 genes were predicted to encode proteins containing double-repeated AP2/ERF domains. Five genes were predicted to encode a single AP2/ERF domain, together with one B3 domain. On the basis of the definition of the 3 families of the AP2/ERF superfamily, the former and the latter genes were assigned to the AP2 and the RAV families, respectively. In addition, 112 genes were predicted to encode proteins containing a single AP2/ERF domain. Overall, 104 of these 112 genes were assigned to the ERF family. The remaining 7 genes (ppa021782m, ppa014566m, ppa006202m, ppb023788m, ppa020991m, ppa018704m, and ppa018845m) encoded a single AP2/ERF domain and occurred in the same cluster as the 14 members of the AP2 family in the phylogenetic tree (Figure 1). Taking the AP2 family classification of *Arabidopsis* as a reference, these 7 proteins were assigned to the AP2 family. The remaining protein (ppa005376m) contained a single AP2 domain and a PAT1 domain, and its homology appeared to be quite low in comparison with that of the proteins of the AP2 or ERF family. Therefore, this gene was designated as a “soloist”. Overall, the 131 genes of the AP2/ERF superfamily in peach could be divided into 3 families; namely, ERF (104 genes), AP2 (21 genes) and RAV (5 genes). In addition, there was a soloist gene (ppa005376m), as shown in Table 1.

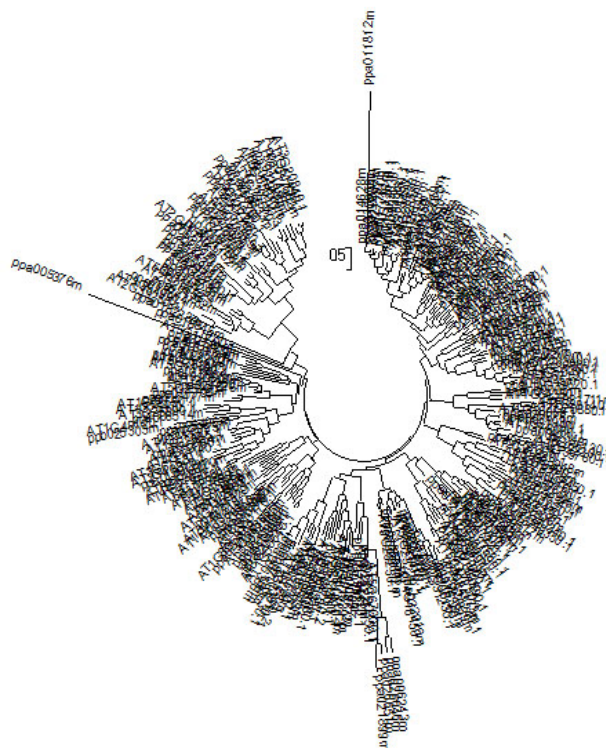


Figure 1. Phylogenetic tree of the AP2/ERF (APETALA2/ethylene-responsive factor) superfamily protein in peach and *Arabidopsis*. The tree was created by the bootstrap option of the CLUSTAL W multiple alignment packages and the neighbour-joining method using the 138 *Arabidopsis* AP2/ERF amino acid sequences and 131 peach AP2/ERF amino acid sequences.

Table 1. The number comparison of members of AP2/ERF families and groups in different plant species.

Plant species		<i>Prunus persica</i>		<i>Arabidopsis thaliana</i>		<i>Vitis vinifera</i>		<i>Oryza sativa</i>		<i>Cucumis sativus</i>	
Family	Group	N	%	N	%	N	%	N	%	N	%
ERF	I	6		10		5		9		5	
	II	9		15		8		16		10	
	III	23		23		22		27		20	
	IV	7		9		5		6		7	
	V	11		5		11		8		15	
	VI	3		8		5		6		8	
	VI-L	4		4		2		3		0	
	VII	6		5		3		15		3	
	VIII	10		15		11		15		11	
	IX	19		17		40		18		16	
	X	6		8		10		12		8	
	Xb-L	0		3		0		10		0	
		104	79.39	122	82.99	122	81.88	145	80.56	103	78.63
AP2		21	16.03	18	12.24	20	13.42	29	16.11	20	15.27
RAV		5	3.82	6	4.08	6	4.03	5	2.78	4	3.05
Soloist		1	0.76	1	0.68	1	0.67	1	0.56	4	3.05
Total		131		147		149		180		131	

Phylogenetic analysis and motif display of predicted AP2/ERF proteins in peach

To confirm the obtained classifications and to further categorize the ERF family, as well as to analyze the phylogenetic relationships, a phylogenetic tree based on the alignment of the AP2/ERF amino acid sequences between 131 AP2/ERF proteins in peach and AP2/ERF proteins in *Arabidopsis* was constructed (Figure 1). Following the group classification of ERF family genes in *Arabidopsis*, 104 ERF proteins in peach were further identified as belonging to 11 groups (categorized as groups I to X, and VI-L; Table 1), which correspond to group I to X and VI-L, as classified by Nakano et al. (2006) for *Arabidopsis*. ERF proteins belonging to the Xb-like group in *Arabidopsis* were not found in the peach genome. The potential members of this group in peach were instead classified as belonging to the RAV family, since the members of this group all contain a AP2 and a B3 domain. In particular, we found that the number of ERF genes belonging to group III and group IX in peach is relatively large. Even though ppa018223m was clustered together with AP2 family proteins, as shown in the phylogenetic tree (Figure 1), both InterProScan analysis and the CD search results indicate that ppa018223m has a single AP2 and a B3 domain. Consequently, the ppa018223m in this study was finally classified into the peach RAV family. The 104 genes in the ERF family were each named in order of generic name, family name, group name, and the order within each group (shown in a subsequent section). As for the AP2 and RAV families, the group name was not available. This numbering method provides a unique identifier for each AP2/ERF gene and allows these genes to be easily distinguished. Given the above classification, the 131 genes of the AP2/ERF family were subjected to further analyses.

The MEME web server was used to analyze motif distribution for peach AP2/ERF proteins (Figure 2 and Table 2). Among the total 131 amino acid sequences of peach AP2/ERF proteins, the shortest and longest sequences were 110 and 779 amino acids, respectively, with the total length of 131 sequences being 38089 amino acids. Motif 1 existed in all the AP2/ERF proteins, except ppb023788m and ppa019435m. Motif 4 had the longest sequence and reached 113 amino acids, followed by Motifs 1 and 3. Motifs 2 and 5 were the shortest, with 11 amino acids. Motif 2 was absent in 4 peach AP2/ERF proteins (ppa006202m, ppb023788m, ppa021782m, and ppa026483m), while Motif 3 was not observed in 7 peach AP2/ERF pro-

Analysis of AP2/ERF superfamily in peach

teins (ppa006202m, ppa020605m, ppa008514m, ppb025303m, ppa026651m, ppa010233m, and ppa011812m). Forty-seven peach ERF proteins belonged to groups I to IV, VI-L, and VIII to X, while 1 AP2 protein (ppa014566m) had no Motif 4. Motif 5 was not found in 20 peach AP2/ERF proteins, which mostly belonged to the AP2 and RAV families. It should be noted that 80% of the genes in the RAV family of peach do not have Motif 5.

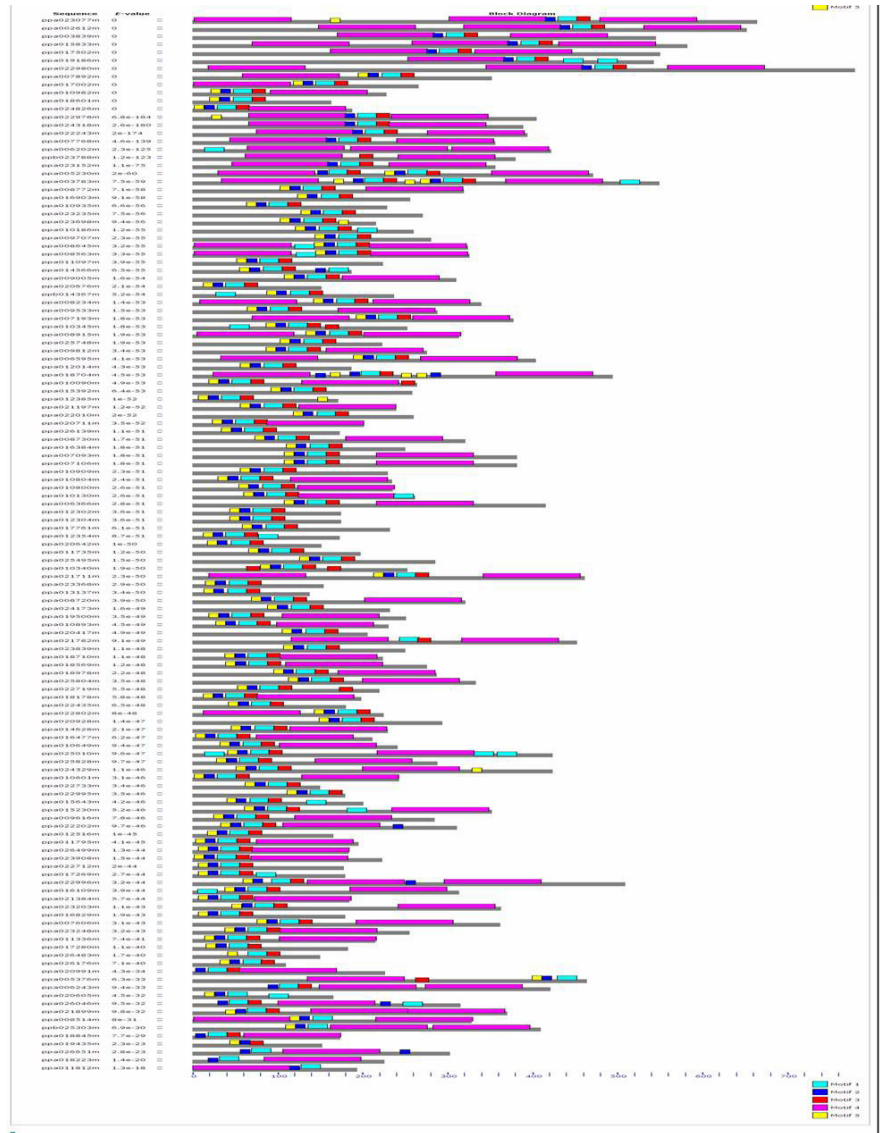


Figure 2. Motif distribution of AP2/ERF superfamily proteins in peach. Motifs of AP2/ERF superfamily proteins were investigated on the MEME web server. The green, blue, red, purple, and yellow boxes represent Motif 1, Motif 2, Motif 3, Motif 4, and Motif 5, respectively. The length of gray line shows the length of a sequence relative to all the other sequences; the position of a block shows where a motif has matched the sequence. The height of a block gives an indication of the significance of the match with taller blocks being more significant.

Table 2. The width and sequence of motif of AP2/ERF superfamily proteins in peach.

Motif name	Width (aa)	Sequence
Motif 1	22	RIWLGTFDTAEAAAYDVAAI
Motif 2	11	WGKWVAEIRDP
Motif 3	15	KMRGPKAKLNPHLI
Motif 4	113	QRTSIYRGVTRHRWTGRYEHLWDNSCWKEGQTRKGRQGGYDKEEKAARAYDLAALK YWGTTTHTNFPISNYEKEIEMKNMTRQEYVAHLRRKSSGFSRGASIYRGVTRHHQ
Motif 5	11	RHCHYRGVRQR

aa = amino acids.

Characteristics of predicted AP2/ERF genes in peach

To further investigate the relationship between the genetic divergence within the AP2/ERF superfamily and gene duplication in peach, the scaffold location of each AP2/ERF gene was determined from the peach genomic sequences. In total, 129 AP2/ERF genes were found unevenly distributed on scaffolds 1 to 8 of the peach genome, while scaffolds 17 and 10 contained just 1 ERF gene each (ppa006595m and ppa020676m, respectively). The numbers of AP2/ERF genes distributed on scaffolds 1 to 8 were 30, 13, 17, 6, 21, 17, 18, and 7, respectively (Table 3). Scaffold 1 appeared to contain 25 ERF, 1 RAV, and 4 AP2 genes. Scaffold 2 contained 11 ERF and 2 AP2 genes, while no RAV family gene was observed. Scaffold 3 contained 11 ERF, 1 RAV, and 5 AP2 genes. Interestingly, the peach ERF family contains 3 groups of genes that have undergone duplication events. These genes are Ppa012302m and Ppa012304m (Group 1), Ppa008645m and Ppa008563m (Group 2), Ppa007093m, Ppa007106m and Ppa006366m (Group 3), located on scaffolds 1, 1, and 3, respectively. In comparison, we found that the nucleotide sequences for Ppa012302m and Ppa012304m genes were identical. Ppa008563m and Ppa008645m almost had the same nucleotide sequences, except that the former had 6 more tandem nucleotides than the latter. The DNA sequences for Ppa007093m and Ppa007106m were identical, while Ppa006366m had similar nucleotide sequences, except for some differences in the tail sequence. No duplication of genes was found in the AP2 and RAV families of peach.

Schematic structures of AP2/ERF genes were shown by GSDS utility in Figure 3 and Table 3, whereby 75 genes in the ERF family, 4 genes in RAV family, and the soloist gene were found to have 1 exon and no introns. Overall, 22 genes in the ERF family and 1 gene in the RAV family contained a single intron and 2 exons in their ORF regions. The remaining 7 genes from the ERF family were unevenly distributed in groups II, III, VI-L, VII, and X, and had 2 to 4 introns. The members of the AP2 family have more exon and intron numbers than those in the ERF and RAV families, with 2 to 10 exons/introns. Furthermore, all coding sequences (conserved domains) of AP2 family genes were disrupted into several segments by introns. The number and position of the introns were relatively conserved in each family; thus, the classification of the AP2/ERF superfamily genes of peach in this study requires further validation.

Analysis of AP2/ERF superfamily in peach

Table 3. The characteristics of *AP2/ERF* superfamily genes in peach.

Family name	Group name	Gene name	Genome location		Scaffold distribution	ORF (bp)	Extron No.	Intron No.
			Gene ID	Location				
ERF	Group I	PpERFI-1	ppa010345m	692998-693990	1	762	2	1
		PpERFI-2	ppa009005m	11915546-11917173	5	933	1	0
		PpERFI-3	ppa021711m	12218348-12219733	3	1386	1	0
		PpERFI-4	ppa020676m	593610-594065	10	456	1	0
		PpERFI-5	ppa007193m	41299998-41301438	1	1137	1	0
	Group II	PpERFI-6	ppa008234m	17044062-17045391	6	1023	1	0
		PpERFII-1	ppa011336m	18867746-18868393	7	648	1	0
		PpERFII-2	ppa017280m	213972-214523	1	552	1	0
		PpERFII-3	ppa010649m	40489767-40490492	1	726	1	0
		PpERFII-4	ppa020605m	45190345-45190845	1	501	1	0
		PpERFII-5	ppa012302m	36018710-36020437	1	528	1	0
		PpERFII-6	ppa019435m	16231660-16232844	6	462	3	2
		PpERFII-7	ppa012304m	36018710-36020437	1	528	1	0
		PpERFII-8	ppa026483m	37774573-37775081	1	453	2	1
		PpERFII-9	ppa012516m	21547459-21547959	6	501	1	0
	Group III	PpERFIII-1	ppa018178m	45256849-45257448	1	600	1	0
		PpERFIII-2	ppa012354m	45233476-45234395	1	522	1	0
		PpERFIII-3	ppa026139m	19377380-19377898	2	519	1	0
		PpERFIII-4	ppa010130m	20261165-20262021	7	789	2	1
		PpERFIII-5	ppa009533m	36913157-36914539	1	867	2	1
		PpERFIII-6	ppa011735m	3632495-3633429	3	597	2	1
		PpERFIII-7	ppa010340m	7934425-7935186	5	762	1	0
		PpERFIII-8	ppa009812m	15192643-15193473	2	831	1	0
		PpERFIII-9	ppa021384m	10107139-10107696	5	558	1	0
		PpERFIII-10	ppa022712m	10104607-10105143	5	537	1	0
		PpERFIII-11	ppa017269m	10081392-10081934	5	543	1	0
		PpERFIII-12	ppa016829m	10067523-10068065	5	543	1	0
		PpERFIII-13	ppa026499m	10120279-10120833	5	555	1	0
		PpERFIII-14	ppa022719m	24808732-24809394	2	393	1	0
		PpERFIII-15	ppa020711m	14542381-14542989	7	609	1	0
	Group IV	PpERFIII-16	ppa015643m	14516336-14516941	7	606	1	0
		PpERFIII-17	ppa010935m	24801153-24802102	2	690	1	0
		PpERFIII-18	ppa017761m	10060832-10061530	5	699	1	0
		PpERFIII-19	ppa021197m	10085521-10086243	5	723	1	0
		PpERFIII-20	ppa010800m	10058909-10059836	5	711	1	0
		PpERFIII-21	ppa010909m	10054488-10055353	5	693	1	0
		PpERFIII-22	ppa011812m	18164940-18167739	8	585	5	4
		PpERFIII-23	ppa014628m	10051648-10052340	5	441	1	0
		PpERFIV-1	ppa015230m	23081760-23082818	2	1041	1	0
		PpERFIV-2	ppa023248m	32916656-32917423	1	768	1	0
	Group V	PpERFIV-3	ppa026176m	28247606-28247935	6	330	1	0
		PpERFIV-4	ppa009616m	10101239-10102130	7	858	1	0
		PpERFIV-5	ppa022996m	24559677-24561206	6	1380	1	0
		PpERFIV-6	ppa022733m	16997324-16997773	3	450	1	0
		PpERFIV-7	ppa007606m	23291301-23293301	2	1089	2	1
		PpERFV-1	ppa016477m	6355361-6356182	3	639	2	1
		PpERFV-2	ppa011795m	33868268-33869325	1	588	2	1
		PpERFV-3	ppa024826m	5379408-5380485	3	567	2	1
		PpERFV-4	ppa010601m	21251826-21252701	7	732	2	1
		PpERFV-5	ppa023908m	39157972-39158845	1	672	2	1
	Group VI	PpERFV-6	ppa025010m	19468943-19470217	3	1275	1	0
		PpERFV-7	ppa022202m	5858790-5860972	1	936	1	0
		PpERFV-8	ppa019500m	12907640-12908578	5	756	2	1
		PpERFV-9	ppa018710m	16882053-16882727	5	675	1	0
		PpERFV-10	ppa016109m	4278712-4279656	7	945	1	0
		PpERFV-11	ppa025828m	864935-865801	6	867	1	0
		PpERFVI-1	ppa008720m	25573878-25574850	2	966	1	0
		PpERFVI-2	ppa018978m	18964811-18965674	2	864	1	0
		PpERFVI-3	ppa025804m	11739213-11740214	5	1002	1	0

Continued on next page

Table 3. Continued.

Family name	Group name	Gene name	Genome location		Scaffold distribution	ORF (bp)	Extron No.	Intron No.
			Gene ID	Location				
	Group VI-L	PpERF VI-L-1	ppa022995m	9609976-9610556	7	540	2	1
		PpERF VI-L-2	ppa025495m	2225762-2226622	6	861	1	0
		PpERF VI-L-3	ppb025303m	17202489-17205280	5	1233	3	2
		PpERF VI-L-4	ppa008514m	29188725-29190132	1	987	2	1
	Group VII	PpERFVII-1	ppa008645m	9682323-9684425	1	975	2	1
		PpERFVII-2	ppa008563m	9682323-9684425	1	981	2	1
		PpERFVII-3	ppa008730m	21465216-21466550	8	966	2	1
		PpERFVII-4	ppa007093m	1542833-1545963	3	1149	2	1
		PpERFVII-5	ppa007106m	1542833-1545963	3	1149	2	1
		PpERFVII-6	ppa006366m	1542833-1545963	3	1248	3	2
	Group VIII	PpERFVIII-1	ppa010893m	13946865-13947794	4	696	1	0
		PpERFVIII-2	ppa010804m	16137175-16137884	3	708	1	0
		PpERFVIII-3	ppa022435m	19825610-19826155	8	546	1	0
		PpERFVIII-4	ppa012385m	2449518-2450320	4	519	1	0
		PpERFVIII-5	ppa020928m	6245994-6246878	3	885	1	0
		PpERFVIII-6	ppa010982m	10424027-10425076	4	687	1	0
		PpERFVIII-7	ppa010090m	2460366-2461342	4	795	1	0
		PpERFVIII-8	ppa024329m	24934935-24936209	1	1275	1	0
		PpERFVIII-9	ppa018569m	21650784-21651614	1	831	1	0
		PpERFVIII-10	ppa023203m	10492271-10493362	1	1092	1	0
	Group IX	PpERFIX-1	ppb014367m	11950245-11950958	7	714	1	0
		PpERFIX-2	ppa020417m	11935491-11936111	7	621	1	0
		PpERFIX-3	ppa016903m	11927754-11928662	7	771	2	1
		PpERFIX-4	ppa023698m	11941426-11942076	7	651	1	0
		PpERFIX-5	ppa025748m	11952264-11952935	7	672	1	0
		PpERFIX-6	ppa010186m	23987005-23987961	2	783	1	0
		PpERFIX-7	ppa009707m	6745332-6746183	5	846	1	0
		PpERFIX-8	ppa012014m	2667444-2668380	4	564	1	0
		PpERFIX-9	ppa022802m	23997736-23998410	2	675	1	0
		PpERFIX-10	ppa008915m	23994070-23995273	2	945	1	0
		PpERFIX-11	ppa007892m	6758124-6759321	5	1059	1	0
		PpERFIX-12	ppa024173m	2633064-2633762	1	699	1	0
		PpERFIX-13	ppa023839m	19495498-19496250	8	753	1	0
		PpERFIX-14	ppa016384m	27908013-27908984	6	753	1	0
		PpERFIX-15	ppa013137m	2616966-2617573	1	417	1	0
		PpERFIX-16	ppa023368m	19504277-19504741	8	465	1	0
		PpERFIX-17	ppa022010m	2669510-2672062	4	783	2	1
		PpERFIX-18	ppa018601m	2621868-2622359	1	492	1	0
		PpERFIX-19	ppa020642m	19503037-19503495	8	459	1	0
	Group X	PpERFX-1	ppa023235m	14266298-14267113	8	816	1	0
		PpERFX-2	ppa017002m	15615815-15616615	7	801	1	0
		PpERFX-3	ppa011097m	21815423-21817543	1	675	3	2
		PpERFX-4	ppa008772m	13146230-13148748	5	960	4	3
		PpERFX-5	ppa015392m	13140697-13142141	5	645	3	2
		PpERFX-6	ppa006595m	322660-324185	17	1215	2	1
RAV		PpRAV-1	ppa006243m	28947286-28948651	1	1266	1	0
		PpRAV-2	ppa026046m	18052310-18053257	3	945	1	0
		PpRAV-3	ppa026651m	4119490-4120669	6	912	2	1
		PpRAV-4	ppa021899m	22098619-22099731	7	1113	1	0
		PpRAV-5	ppa018223m	6721683-6722360	5	678	1	0
AP2	Double AP2 domain	PpAP2-1	ppa022243m	26638137-26640085	1	1185	8	7
		PpAP2-2	ppa023152m	7647362-7649481	1	1071	7	6
		PpAP2-3	ppa003839m	5868139-5871493	1	1638	8	7
		PpAP2-4	ppa023077m	35308331-35311476	1	1995	8	7
		PpAP2-5	ppa002612m	20514710-20517978	6	1959	8	7
		PpAP2-6	ppa015833m	2130011-2133106	3	1749	8	7
		PpAP2-7	ppa019186m	19187959-19192091	6	1632	9	8
		PpAP2-8	ppa017502m	3335008-3337757	6	1653	8	7
		PpAP2-9	ppa022980m	28712875-28716799	6	2574	7	6

Continued on next page

Analysis of AP2/ERF superfamily in peach

Table 3. Continued.

Family name	Group name	Gene name	Genome location		Scaffold distribution	ORF (bp)	Extron No.	Intron No.
			Gene ID	Location				
Single AP2 domain		PpAP2-10	ppa024318m	19878339-19880404	3	1170	8	7
		PpAP2-11	ppa022978m	19864657-19866637	3	1218	7	6
		PpAP2-12	ppa003783m	21564259-21567398	6	1650	10	9
		PpAP2-13	ppa005230m	22210661-22213780	6	1416	10	9
		PpAP2-14	ppa021782m	6220679-6224096	6	1359	8	7
		PpAP2-15	ppa007768m	7344817-7347764	3	1071	7	6
		PpAP2-16	ppa014566m	10332526-10334444	7	561	6	5
		PpAP2-17	ppa006202m	9381376-9385282	7	1269	8	7
		PpAP2-18	ppb023788m	5864627-5869352	2	1143	6	5
		PpAP2-19	ppa020991m	19871101-19871935	3	441	2	1
		PpAP2-20	ppa018704m	21286241-21288915	2	1485	9	8
		PpAP2-21	Ppa018845m	10342970-10345573	7	459	3	2
Soloist		aP2+pat1	Ppa005376m	587719-590683	6	1395	1	0

ORF = open reading frame.

Characteristics of predicted AP2/ERF proteins in peach

The primary structure of peach AP2/ERF proteins was calculated from the online ProtParam tool, as shown in Table 4. The negative grand average of hydropathicity (GRAVY) index indicated that 13 proteins in the ERF family belong to medium proteins, while the remaining members of the ERF family and all the members of RAV and AP2 families are hydrophilic proteins. The ProtParam results revealed that the number of amino acids was positively correlated with the molecular weight for AP2/ERF family proteins. Among 131 AP2/ERF proteins, 50 were alkaline amino acids and 81 were acidic amino acids on the basis of the value of theoretical pI. Each group contained both alkaline and acidic amino acids, except group VI, in which all were acidic amino acids for the predicted proteins in the ERF family.

The tertiary structures of representative predicted AP2/ERF superfamily proteins in peach were built through the SWISS-MODEL web-based tool (<http://swissmodel.expasy.org/>). Since the AP2/ERF superfamily of peach has a relatively large number of members, one tertiary structure was selected for each, to represent members of the RAV family, AP2 family, and each group of the ERF family, as well as the soloist (Figure 4). As shown in Figure 4, each tertiary structure of the peach AP2/ERF superfamily protein contains 1 α -helix, 3 β -sheets, and 5 random coils. Furthermore, the α -helix of each tertiary structure contains 16 amino acids displayed by Swiss-PdbViewer. However, the number of amino acids in each β -sheet or a random coil of each tertiary structure was different, thus creating slight variations in the length of β -sheets or random coils of each tertiary structure. In conclusion, the AP2/ERF superfamily proteins generally have similar tertiary structures, with slight differences in the length and amino acid composition of units that make up the tertiary structures.

Expression profiles of predicted AP2/ERF genes in peach

EST data provides valuable information for gene expression research. In total,

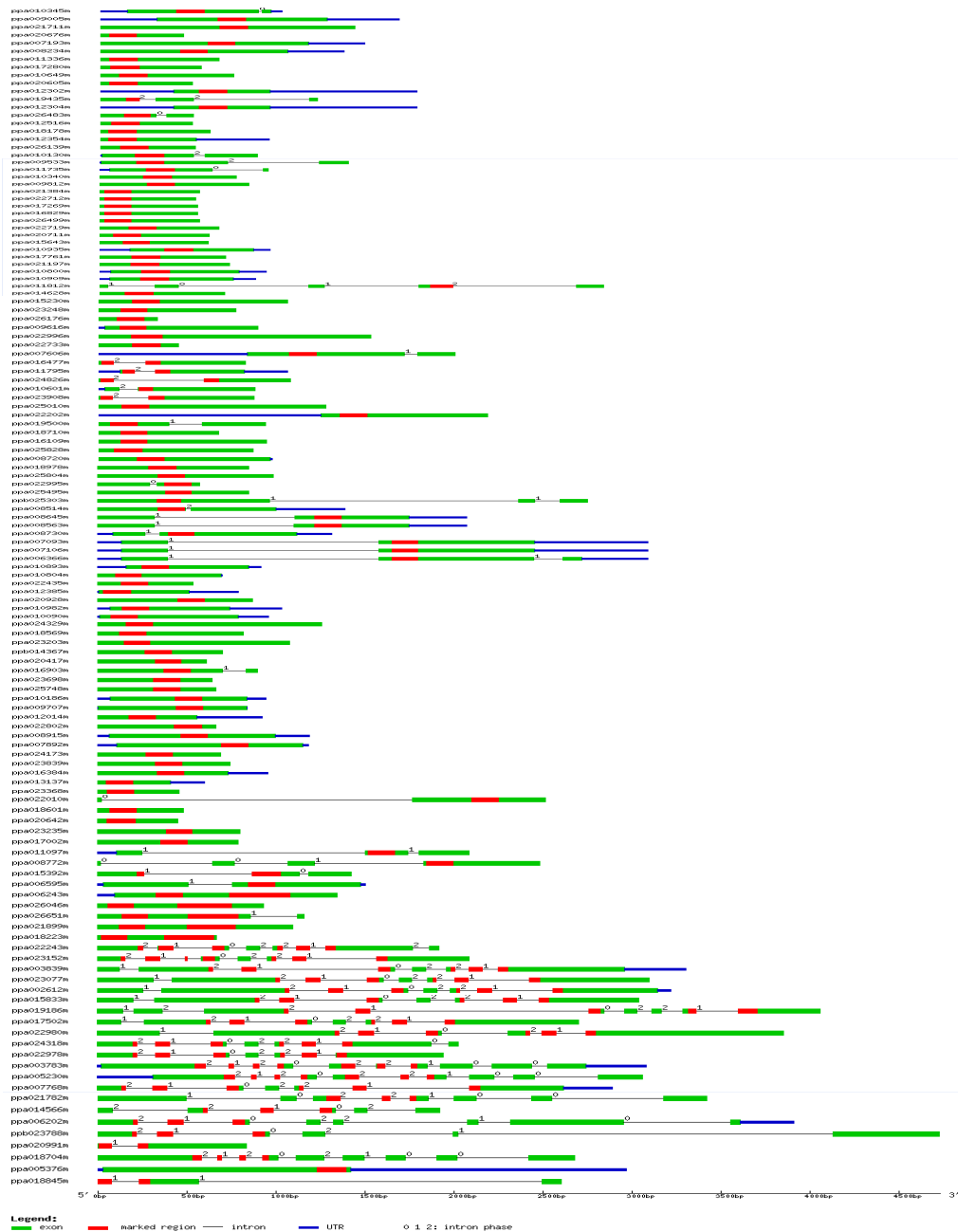


Figure 3. Gene structure of AP2/ERF superfamily in peach. Exons and introns are depicted by filled green boxes and single lines, respectively. Intron phases 0, 1 and 2 are indicated by numbers 0, 1 and 2 in the figure. Conserved sequences of AP2/ERF proteins are marked in red. UTRs are displayed by thick blue lines at the two ends.

Analysis of AP2/ERF superfamily in peach

Table 4. The physical-chemical analysis of AP2/ERF superfamily proteins in peach.

Family name	Group name	Gene name	Gene ID	No. of amino acid	Molecular weight (Da)	Theoretical pI	Aliphatic index	Gravy
ERF	Group I	PpERFI-1	ppa010345m	253	28576.2	9.83	63.00	-0.771
		PpERFI-2	ppa009005m	310	34377.6	8.34	64.55	-0.555
		PpERFI-3	ppa021711m	461	51713.0	5.97	45.60	-0.980
		PpERFI-4	ppa020676m	152	16590.8	9.91	60.46	-0.715
		PpERFI-5	ppa007193m	378	41236.8	6.37	58.94	-0.619
		PpERFI-6	ppa008234m	340	37683.0	7.71	65.47	-0.608
	Group II	PpERFII-1	ppa011336m	215	23980.1	5.16	46.79	-0.805
		PpERFII-2	ppa017280m	183	20044.5	7.74	65.63	-0.484
		PpERFII-3	ppa010649m	241	26308.4	5.61	56.39	-0.421
		PpERFII-4	ppa020605m	166	17890.7	5.62	59.94	-0.564
		PpERFII-5	ppa012302m	175	19844.9	9.39	51.31	-1.107
		PpERFII-6	ppa019435m	153	17272.1	9.36	48.43	-1.159
		PpERFII-7	ppa012304m	175	19844.9	9.39	51.31	-1.107
		PpERFII-8	ppa026483m	150	17240.2	9.42	60.47	-1.039
		PpERFII-9	ppa012516m	166	17849.0	8.72	59.46	-0.616
	Group III	PpERFIII-1	ppa018178m	199	22101.0	5.54	77.04	-0.426
		PpERFIII-2	ppa012354m	173	19078.2	5.44	57.69	-0.598
		PpERFIII-3	ppa026139m	173	19720.6	6.21	45.20	-1.086
		PpERFIII-4	ppa010130m	262	28536.4	4.94	57.82	-0.657
		PpERFIII-5	ppa009533m	288	31806.4	5.72	63.68	-0.688
		PpERFIII-6	ppa011735m	198	21307.6	10.04	63.18	-0.765
		PpERFIII-7	ppa010340m	253	28183.4	4.91	54.11	-0.941
		PpERFIII-8	ppa009812m	276	30098.8	4.94	51.45	-0.880
		PpERFIII-9	ppa021384m	185	19626.9	4.89	66.16	-0.381
		PpERFIII-10	ppa022712m	178	19407.7	5.67	62.64	-0.581
		PpERFIII-11	ppa017269m	180	19546.9	5.50	64.17	-0.486
		PpERFIII-12	ppa016829m	180	19510.8	5.50	59.89	-0.526
		PpERFIII-13	ppa026499m	184	19705.9	4.64	62.17	-0.465
		PpERFIII-14	ppa022719m	220	23225.6	4.75	62.27	-0.592
		PpERFIII-15	ppa020711m	202	21536.6	4.91	69.70	-0.434
		PpERFIII-16	ppa015643m	201	22481.2	4.95	67.56	-0.511
		PpERFIII-17	ppa010935m	229	24976.9	5.14	64.02	-0.411
		PpERFIII-18	ppa017761m	232	26157.4	5.03	74.91	-0.492
		PpERFIII-19	ppa021197m	240	27217.8	7.69	63.83	-0.634
	Group IV	PpERFIII-20	ppa010800m	236	26352.7	5.17	69.11	-0.563
		PpERFIII-21	ppa010909m	230	25658.8	5.13	66.74	-0.497
		PpERFIII-22	ppa011812m	194	21923.9	9.92	64.85	-0.779
		PpERFIII-23	ppa014628m	230	25521.9	7.78	70.04	-0.488
		PpERFIV-1	ppa015230m	352	38119.8	6.55	51.02	-0.807
		PpERFIV-2	ppa023248m	255	28137.5	6.55	58.55	-0.616
		PpERFIV-3	ppa026176m	110	11966.5	10.34	47.18	-0.871
		PpERFIV-4	ppa009616m	285	32054.8	5.73	65.05	-0.728
		PpERFIV-5	ppa022996m	509	57591.9	4.78	55.36	-0.936
		PpERFIV-6	ppa022733m	150	16331.4	10.01	62.60	-0.664
		PpERFIV-7	ppa007606m	362	39578.9	4.62	50.97	-0.900
	Group V	PpERFV-1	ppa016477m	212	24310.1	5.94	65.38	-0.766
		PpERFV-2	ppa011795m	195	22413.4	6.66	63.64	-0.751
		PpERFV-3	ppa024826m	188	20855.2	5.97	69.57	-0.656
		PpERFV-4	ppa010601m	243	26570.5	7.02	66.34	-0.621
		PpERFV-5	ppa023908m	223	24759.6	9.10	62.15	-0.748
		PpERFV-6	ppa025010m	424	46904.3	4.58	60.19	-0.734
		PpERFV-7	ppa022202m	311	33725.1	9.14	53.92	-0.725
		PpERFV-8	ppa019500m	251	26755.3	5.95	63.15	-0.535
		PpERFV-9	ppa018710m	224	25367.4	8.71	66.25	-0.768
		PpERFV-10	ppa016109m	314	34636.3	5.44	59.43	-0.538
	Group VI	PpERFV-11	ppa025828m	288	31688.4	6.01	69.20	-0.579
		PpERFVI-1	ppa008720m	321	36086.0	4.44	66.51	-0.593
		PpERFVI-2	ppa018978m	287	32269.0	4.95	65.64	-0.593
	Group VI-L	PpERFVI-3	ppa025804m	333	37018.9	4.98	62.70	-0.559
		PpERF VI-L-1	ppa022995m	180	20180.9	10.30	82.33	-0.478
		PpERF VI-L-2	ppa025495m	286	31565.0	5.63	66.78	-0.541
		PpERF VI-L-3	ppb025303m	410	46466.9	5.38	71.80	-0.618
		PpERF VI-L-4	ppa008514m	328	36878.9	4.91	67.74	-0.750

Continued on next page

Table 4. The physical-chemical analysis of AP2/ERF superfamily proteins in peach.

Family name	Group name	Gene name	Gene ID	No. of amino acid	Molecular weight (Da)	Theoretical pI	Aliphatic index	Gravy
RAV	Group VII	PpERFVII-1	ppa008645m	324	35439.5	8.04	63.86	-0.785
		PpERFVII-2	ppa008563m	326	35567.6	8.04	63.77	-0.776
		PpERFVII-3	ppa008730m	321	35682.6	5.68	59.63	-0.845
		PpERFVII-4	ppa007093m	382	42495.9	4.81	55.73	-0.765
		PpERFVII-5	ppa007106m	382	42495.9	4.81	55.73	-0.765
		PpERFVII-6	ppa006366m	415	46253.3	5.07	58.36	-0.706
	Group VIII	PpERFVIII-1	ppa010893m	231	25022.7	7.62	56.32	-0.679
		PpERFVIII-2	ppa010804m	235	25471.3	6.58	57.79	-0.694
		PpERFVIII-3	ppa022435m	181	19951.0	9.99	65.36	-0.561
		PpERFVIII-4	ppa012385m	172	18303.6	9.20	73.20	-0.367
		PpERFVIII-5	ppa020928m	294	32342.6	9.60	66.33	-0.537
		PpERFVIII-6	ppa010982m	228	24245.0	9.62	61.01	-0.529
		PpERFVIII-7	ppa010090m	264	28447.6	9.47	66.93	-0.533
		PpERFVIII-8	ppa024329m	424	46296.9	7.73	44.01	-0.692
		PpERFVIII-9	ppa018569m	276	29938.4	4.94	32.21	-0.826
		PpERFVIII-10	ppa023203m	363	39938.6	5.60	49.28	-0.711
	Group IX	PpERFIX-1	ppb014367m	237	25948.7	5.47	70.51	-0.530
		PpERFIX-2	ppa020417m	206	22559.0	5.97	65.97	-0.664
		PpERFIX-3	ppa016903m	256	28258.5	5.75	63.63	-0.549
		PpERFIX-4	ppa023698m	216	23968.8	5.99	77.64	-0.586
		PpERFIX-5	ppa025748m	223	24950.7	5.41	72.20	-0.606
		PpERFIX-6	ppa010186m	260	29210.7	6.53	70.27	-0.668
		PpERFIX-7	ppa009707m	281	30157.0	9.09	76.65	-0.411
		PpERFIX-8	ppa012014m	187	20369.5	9.51	67.33	-0.600
		PpERFIX-9	ppa022802m	225	25176.0	6.71	58.98	-0.753
		PpERFIX-10	ppa008915m	314	35018.8	6.17	62.77	-0.765
		PpERFIX-11	ppa007892m	352	39322.1	5.90	68.52	-0.617
		PpERFIX-12	ppa024173m	232	25950.7	5.54	70.65	-0.626
		PpERFIX-13	ppa023839m	250	28411.7	5.04	65.20	-0.567
		PpERFIX-14	ppa016384m	250	28487.0	4.92	64.4	-0.744
		PpERFIX-15	ppa013137m	138	15204.7	6.84	56.67	-0.902
		PpERFIX-16	ppa023368m	154	17095.6	7.79	52.08	-0.798
RAV	Group X	PpERFX-8	ppa022010m	260	29153.1	5.95	75.85	-0.505
		PpERFX-9	ppa018601m	163	18008.6	6.85	49.82	-0.938
		PpERFX-10	ppa020642m	152	17339.9	5.97	54.01	-1.007
		PpERFX-1	ppa023235m	271	30318.1	8.16	54.56	-0.786
		PpERFX-2	ppa017002m	266	29255.3	7.67	59.51	-0.777
		PpERFX-3	ppa011097m	224	24963.4	5.88	52.77	-0.867
	Group X	PpERFX-4	ppa008772m	319	35646.3	7.09	55.74	-0.738
		PpERFX-5	ppa015392m	259	29137.0	9.08	60.27	-1.010
		PpERFX-6	ppa006595m	404	43314.9	7.07	63.34	-0.517
		PpRAV-1	ppa006243m	421	45813.0	8.89	68.31	-0.611
AP2	Double AP2 domain	PpRAV-2	ppa026046m	315	36448.4	8.08	74.22	-0.637
		PpRAV-3	ppa026651m	303	35224.5	6.41	66.93	-0.681
		PpRAV-4	ppa021899m	370	41765.7	6.27	62.65	-0.518
		PpRAV-5	ppa018223m	226	26671.1	8.89	58.72	-0.615
		PpAP2-1	ppa022243m	549	60645.5	6.39	48.74	-0.895
		PpAP2-2	ppa023152m	356	40250.7	8.10	8.1	-0.700
		PpAP2-3	ppa003839m	545	59357.7	8.34	52.2	-0.728
		PpAP2-4	ppa023077m	664	72809.7	6.62	60.33	-0.648
		PpAP2-5	ppa002612m	652	72711.4	6.39	55.89	-0.792
		PpAP2-6	ppa015833m	582	63794.7	6.37	50.34	-0.857
		PpAP2-7	ppa019186m	543	60008.0	5.75	57.42	-0.602
		PpAP2-8	ppa017502m	550	60595.3	6.22	61.44	-0.697
		PpAP2-9	ppa022980m	779	85422.6	6.30	51.68	-0.820
		PpAP2-10	ppa024318m	389	44121.9	8.82	57.48	-0.818
		PpAP2-11	ppa022978m	405	45663.5	8.54	57.90	-0.800
		PpAP2-12	ppa003783m	394	44452.6	9.47	63.91	-0.807
		PpAP2-13	ppa005230m	471	51879.7	6.90	65.52	-0.50
		PpAP2-14	ppa007768m	356	39806.7	8.65	57.36	-0.862

Continued on next page

Analysis of AP2/ERF superfamily in peach

Table 4. The physical-chemical analysis of AP2/ERF superfamily proteins in peach.

Family name	Group name	Gene name	Gene ID	No. of amino acid	Molecular weight (Da)	Theoretical pI	Aliphatic index	Gravy
Single AP2 domain		PpAP2-15	ppa021782m	452	49443.7	6.15	57.85	-0.658
		PpAP2-16	ppa014566m	187	21331.0	9.75	45.94	-1.042
		PpAP2-17	ppa006202m	422	46822.2	9.01	61.14	-0.770
		PpAP2-18	ppb023788m	380	42295.6	5.35	56.82	-0.831
		PpAP2-19	ppa020991m	227	25172.4	4.54	60.66	-0.648
		PpAP2-20	ppa018704m	494	53439.0	6.24	57.71	-0.612
		PpAP2-21	ppa018845m	174	19448.8	9.12	70.17	-0.685
Soloist		aP2+pat1	ppa005376m	464	51527.4	10.06	60.52	-0.981

pI = isoelectric point.

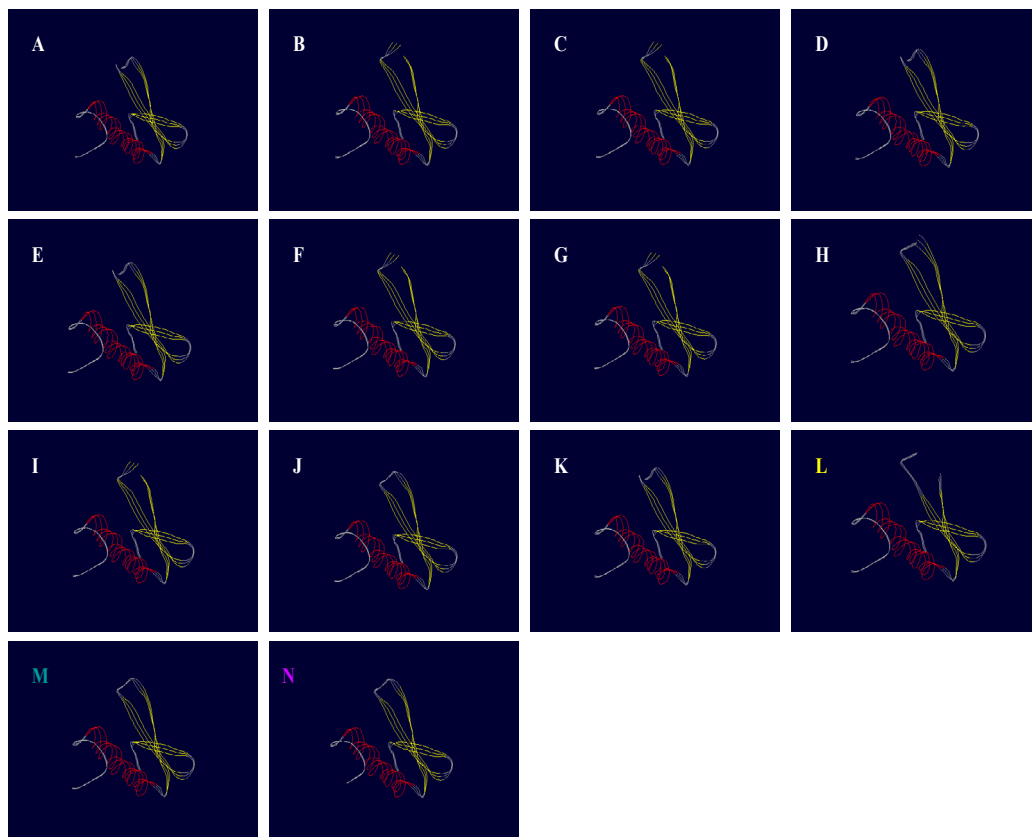


Figure 4. Tertiary structures of fourteen representative AP2/ERF superfamily proteins in peach. **A.** ppa010345m (ERF-Group I); **B.** ppa011336m (ERF-Group II); **C.** ppa018178m (ERF-Group III); **D.** ppa015230m (ERF-Group IV); **E.** ppa016477m (ERF-Group V); **F.** ppa008720m (ERF-Group VI); **G.** ppa022995m (ERF-Group VI-L); **H.** ppa008645m (ERF-Group VII); **I.** ppa010893m (ERF-Group VIII); **J.** ppa014367m (ERF-Group IX); **K.** ppa023235m (ERF-Group X); **L.** ppa006243m (RAV); **M.** ppa022243m (AP2); **N.** ppa005376m (Soloist); red = alpha helix; yellow = beta sheet extended strand; gray = random coil.

12,740 AP2/ERF ESTs were searched in the peach NCBI EST database. The expression statistics of predicted AP2/ERF family genes in different organs or tissues of peach were analyzed, as shown in Figure 5 where the percentage of predicted AP2/ERF family genes expressed in mesocarp and fruit skin hold the first and second places among all tissues or organs, reaching 87.6 and 5.0%, respectively. This was followed by 3.4, 1.2, 0.9, 0.9, and 0.6% for fruit mesocarp plus epidermis, shoot, leaf, whole fruit minus the stone, and the flower bud, respectively. The percentage of predicted AP2/ERF family genes expressed in branch bark (0.3%) was the lowest among all tissues or organs.

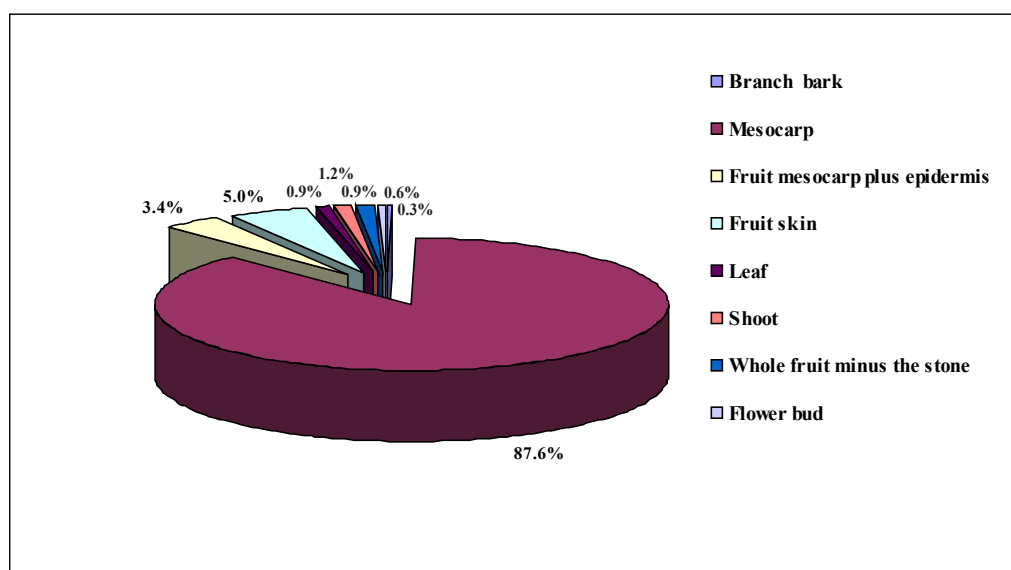


Figure 5. Expression of AP2/ERF superfamily genes in different peach tissues or organs *in silico*.

DISCUSSION

In this study, a comprehensive analysis for genes encoding AP2/ERF proteins in the peach genome was performed, resulting in the identification of 131 AP2/ERF superfamily genes. The availability of the complete genome sequences and the previous identification of ERF/AP2 superfamily genes from some plant species (Nakano et al., 2006; Licausi et al., 2010; Hu and Liu, 2011) enabled the comparison of individual families and groups, basing the classification of these genes on *Arabidopsis* (Nakano et al., 2006) as a criterion. Overall (Table 1), the AP2/ERF superfamily has a similar number of genes in peach and cucumber, which is less than that in *Arabidopsis* (147), grapevine (149), and rice (180). The number of these genes from the RAV family is highly conserved among species, with 6 members in *Arabidopsis* and grapevine, 5 in peach and rice, and 4 in cucumber. The AP2 family contains a similar number of genes in grapevine (20), cucumber (20), *Arabidopsis* (18), and peach (21), while it is 29 in rice (Table 1). The gene number in each family of peach does not differ much from that of other species, despite the fact that peach has a large genome of 220-230 Mbp (Jung et al.,

2008), which is larger than that of *A. thaliana* (145 Mbp) (Huala et al., 2001) and shorter than that of *Oryza sativa* (430 Mbp) (Goff et al., 2002), cucumber (367 Mbp) (Ren et al., 2009), and grapevine (475 Mb) (Velasco et al., 2007). This observation is clearly illustrated by the gene numbers in the RAV family, which remain relatively constant among most species. This indicates the possibility that the structure and phylogeny of the AP2/ERF superfamily genes is highly similar in plant species. The uneven distributions of the AP2/ERF gene family in various plants may have arisen from some gene duplication events.

Although the overall number of genes belonging to each family of the AP2/ERF superfamily was different among the plant species, the percentage gene number in each family to the total number of genes in the AP2/ERF superfamily among plant species was similar, regardless of whether the comparison was among woody and herbaceous species or among vegetable and fruit species (Table 1). Furthermore, we found that the percentage of each family member to the total numbers of AP2/ERF superfamily follows the rule where the percentage of gene numbers in the ERF family/total numbers of AP2/ERF superfamily for each plant species is about 4 times as high as that for the AP2 family. In comparison, the gene number for the AP2 family is about 4 times as high as that in RAV family, while that of RAV family is about 4 to 5 times more than soloist genes. We found that the soloist protein coded by a single-copy gene was present in almost all the AP2/ERF superfamilies identified in different plant species, with the exception of cucumber, which has 4 soloist genes. A divergent member (Ppa005376m) containing an AP2 and a PAT1 domain through the CD search (PAT1 domain was not shown in Figure 3) was also observed in this study. The conservation of the sequence, plus the fact that it is weakly related to other AP2/ERF genes, suggests that this gene diverged from the AP2 family early in the evolution of the plant species. However, how this gene retained its single-copy feature during the separation of the woody and herbaceous plants and the spread of the eucosid clade requires explanation. In addition, the genes numbers in group III and IX of the ERF family hold the first and second place compared to those in all other groups of the ERF family for each studied plant species, apart from grapevine (Table 1). Nakano et al. (2006) concluded that the members of group III and group IX might play vital roles in abiotic and/or biotic stress tolerance or resistance in several species; hence, the authors deduced that a relatively large number of genes in these groups might be the consequence of evolutionary adaptations to various environmental changes. Like in tomato, grapevine, and cucumber, the ERF family proteins in peach are classified into 11 groups. Since all 11 groups are present in monocot and dicot systems, it is possible that major diversification of this family predates monocot-dicot divergence. These similar traits indicate the AP2/ERF superfamily genes had a feature of high conservation during the evolution or separation process of various plant species.

It has been widely accepted that the intron/exon position pattern provides clues about evolutionary relationships (Hu and Liu, 2011). Sakuma et al. (2002) reported that most genes in the ERF family of *Arabidopsis* possess no introns, with only 4 genes having an intron. Our schematic structure analysis indicates that the absence of introns in most ERF family genes is also a feature in peach (Table 3 and Figure 3). We also found that the genes of the AP2 family in peach have the highest number of introns and exons compared to those in the ERF and RAV families. Furthermore, all the repeated double AP2 conserved domain sequences in the AP2 family were disrupted into several segments by introns. These findings in peach are inconsistent with those of a previous study on rice, where the number of introns was found to

vary from 5 to 11 in the AP2 family (Sharoni et al., 2011). Besides further validating the family classification of the AP2/ERF superfamily, this observation also indicates that AP2/ERF family genes have similar characteristics among plants (i.e., both dicots and monocots), which may contribute to their functional similarity within the same family.

On the basis of scaffold distribution and genome location, the phenomenon of gene duplication was also found in the peach genome. For example, in the ERF family, Ppa012302m and Ppa012304m, Ppa008645m and Ppa008563m, Ppa007093m, Ppa007106m and Ppa006366m were duplicated and were located on scaffolds 1, 1, and 3, respectively (Table 3). Similar to *Arabidopsis*, the duplication events of these genes may have a major influence on the amplification of members of a gene family in the peach genome. In addition, the pairwise evolution of duplicated genes also indicates that these genes might coordinately regulate certain biological processes common to this species, such as signal transduction and transcription. This is supported by previous findings, which demonstrate that the duplicated genes involved in signal transduction and transcription are preferentially retained (Blanc and Wolfe, 2004).

Comparative analysis of the phylogenetic relationships among the AP2/ERF genes of peach and *Arabidopsis* was performed in this study. The results reveal much about the diversification and conservation of the AP2/ERF superfamily in peach, where segmented or whole gene duplication, as well as a more ancient transposition and homing, might have contributed to the expansion of the AP2/ERF gene superfamily. During this expansion, many groups have evolved, resulting in a high level of functional divergence in the AP2/ERF gene superfamily. For example, the Xb-like group of ERF family genes is present in *Arabidopsis* but not in peach species (Table 1); instead the members of Xb-like group in *Arabidopsis* and RAV proteins of peach were clustered together (Figure 1). We therefore deduced that these members might have only evolved in peach after functional divergence. Since peach is a woody species, selection either during domestication from its wild ancestor or during subsequent agricultural improvement may also have been important for the evolution of the peach AP2/ERF superfamily.

MEME is widely used to analyze sequences uploaded for similarities among DNA or proteins sequences, producing a motif for each pattern that is discovered. The E-value of a sequence is the expected number of sequences in a random database of the same size that would match the motifs, as well as the sequence, and is equal to the combined P-value of the sequence times the number of sequences in the database (Bailey et al., 2006). The E-value of 7 proteins in the AP2 family (30%) appeared to be 0 (Figure 2), indicating that sequence similarity among these genes is very high. Overall, 16 (76.1%) out of 21 proteins of AP2 family have very similar E-values, which in some ways confirm that members within a given group may have common and recent evolutionary origins. In comparison, among the 21 genes in the peach AP2 family, the Motif 4 sequence is repeated 2 to 3 times in 17 genes, just once in 3 genes, and is absent in 1 gene (ppa014566m). All genes in the RAV family have Motif 5, with the exception of ppa021899m. Proteins within a subgroup that share these motifs are likely to share similar functions (Hu and Liu, 2011). Although the functions of most of these motifs have not been identified, some may play vital roles in transcriptional regulation.

Primary protein structure analysis, secondary protein structure prediction, and tertiary protein structure modeling are important research fields in life sciences. The differences in physical-chemical properties of side chains (amino acid sequences or primary structure) may generate the diversity of three-dimensional protein-folds observed in nature. Experiments performed decades ago demonstrated that the information specifying the three-dimensional

structure of a protein is contained in its amino acid sequence (Anfinsen et al., 1961; Anfinsen, 1973). Elements of secondary structure and super-secondary structure may then combine to form the full three-dimensional fold of a protein or its tertiary structure (Singh, 2001). Regular tertiary structures include the α -helix (H), β -sheets (S), and random coils. All peach proteins in the AP2/ERF superfamily contain 1 α -helix, 3 β -sheets, and 5 random coils (Figure 4). This observation is generally consistent with that in *Arabidopsis* (Allen et al., 1998), whereby the AP2/ERF domain is reported to contain an N-terminal, which is a three-stranded β -sheet that recognizes a target sequence as well as a C-terminal α -helix.

Although the protein sequences in the AP2/ERF superfamily were different, the protein tertiary structures were generally similar in this study. This similarity might be explained to a certain extent by the previous conclusion. For instance, protein structure is more conserved than protein sequence, and 2 sequences that share more than 30% sequence identity are likely to have similar structures (Singh, 2001).

Large amount of ESTs of peach deposited into the GenBank have been created by partially sequencing randomly isolated gene transcripts and have proved valuable in molecular biology (Clifton and Mitreva, 2009). Identification of AP2/ERF family genes and the analysis of gene expression patterns through large amount of ESTs are classic examples. In this study, to detect the expression patterns of the 131 genes in the AP2/ERF superfamily of peach, 8 tissue types were analyzed according to the annotation of ESTs (Figure 5). The proportion of expression of AP2/ERF superfamily genes was highest in the mesocarp than in the other 7 tissue types. The percentage of AP2/ERF superfamily genes expressed in fruit skin and fruit mesocarp plus epidermis held the second and third places among all tissues, reaching 5.0 and 3.4%, respectively. Therefore, we speculate that the AP2/ERF superfamily, especially ERF family genes, may play an important role mainly in fruit growth and development in peach. This is consistent with the results of studies on the fruit of several other plant species. In kiwi, the expression of transcripts of most AdERFs was higher during early fruit development, with the exception of AdERF3, which increased with maturity. Several AdERFs were apparently downregulated by ethylene. Yin et al. (2010) proposed that in kiwifruit, ERFs play a crucial role in the transcriptional regulation of ripening-related genes and in the regulation of kiwi fruit-ripening processes. In tomato, LeERF2 transcripts are most abundant in ripe fruit and its target genes are likely to play a role in the ripening process (Tournier et al., 2003). Overexpression of LeERF1 shortens fruit postharvest life (Li et al., 2007), while LeERF2 regulates ethylene production in tomato and tobacco by modulating ethylene biosynthesis genes. Guo and Ecker (2004) reported that members of the ERF family act as positive controllers of the ethylene response and are potentially involved in cross talk with other signals. Within the pathway of ethylene signal transduction, ERFs are plant-specific nucleus-localized proteins, serving as TFs that bind conserved motifs in promoter regions of target genes and thus providing a route for the activation of the ethylene signal at the level of target gene transcription.

In conclusion, this research has guided us a step further towards understanding the basic information of the AP2/ERF superfamily in peach. Phylogenetic and comparative analyses of AP2/ERF genes in peach may serve as a first step towards the comprehensive functional characterization of the AP2/ERF gene family by reverse genetic approaches in the future. Therefore, the results of the current study contribute toward identifying candidate genes for detailed characterization, as well as providing useful information about the breeding of new cultivars that may be able to adapt to less favorable environmental conditions.

ACKNOWLEDGMENTS

Research supported by the Postdoctoral Scientific Research Foundation of Jiangsu Province (#1002024B). We are grateful to Prof. Jinggui Fang for his critical review of the manuscript.

REFERENCES

- Agarwal P, Agarwal PK, Joshi AJ, Sopory SK, et al. (2010). Overexpression of PgDREB2A transcription factor enhances abiotic stress tolerance and activates downstream stress-responsive genes. *Mol. Biol. Rep.* 37: 1125-1135.
- Agarwal PK, Agarwal P, Reddy MK and Sopory SK (2006). Role of DREB transcription factors in abiotic and biotic stress tolerance in plants. *Plant Cell Rep.* 25: 1263-1274.
- 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, Leisse TJ, Kim CJ, et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301: 653-657.
- Anfinsen CB (1973). Principles that govern the folding of protein chains. *Science* 181: 223-230.
- Anfinsen CB, Haber E, Sela M and White FH Jr (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* 47: 1309-1314.
- Arnold K, Bordoli L, Kopp J and Schwede T (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195-201.
- 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.
- Blanc G and Wolfe KH (2004). Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* 16: 1679-1691.
- 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.
- Clifton SW and Mitreva M (2009). Strategies for undertaking expressed sequence tag (EST) projects. *Methods Mol. Biol.* 533: 13-32.
- Goff SA, Ricke D, Lan TH, Presting G, et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296: 92-100.
- Guillaumot D, Lelu-Walter MA, Germot A, Meytraud F, et al. (2008). Expression patterns of LmAP2L1 and LmAP2L2 encoding two-APETALA2 domain proteins during somatic embryogenesis and germination of hybrid larch (*Larix x marschlinii*). *J. Plant Physiol.* 165: 1003-1010.
- Guo H and Ecker JR (2004). The ethylene signaling pathway: new insights. *Curr. Opin. Plant Biol.* 7: 40-49.
- Hu LF and Liu SQ (2011). Genome-wide identification and phylogenetic analysis of the ERF gene family in cucumbers. *Genet. Mol. Biol.* 34: 624-633.
- 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.
- Huala E, Dickerman AW, Garcia-Hernandez M, Weems D, et al. (2001). The *Arabidopsis* Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.* 29: 102-105.
- 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.
- Jung S, Staton M, Lee T, Blenda A, et al. (2008). GDR (Genome Database for Rosaceae): integrated web-database for Rosaceae genomics and genetics data. *Nucleic Acids Res.* 36: D1034-D1040.
- 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.
- Layne DR and Bassi D (2008). The Peach: Botany, Production and Uses. CABI, London.
- Li Y, Zhu B, Xu W, Zhu H, 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 Genomics* 11: 719.
- Moose SP and Sisco PH (1996). Glossy15, an APETALA2-like gene from maize that regulates leaf epidermal cell identity.

Analysis of AP2/ERF superfamily in peach

- 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.
- Ohme-Takagi M and Shinshi H (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *Plant Cell* 7: 173-182.
- Ren Y, Zhang Z, Liu J, Staub JE, et al. (2009). An integrated genetic and cytogenetic map of the cucumber genome. *PLoS One* 4: e5795.
- 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.
- Shangguan LF, Wang C, Fang JG, Li XY, et al. (2011). Isolation of effective genes with digital northern platform of NCBI Grapevine (*Vitis vinifera* L.) EST sequences. *Sci. Agr. Sin.* 44: 2748-2759.
- 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.
- Shi J, Blundell TL and Mizuguchi K (2001). FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* 310: 243-257.
- Singh M (2001). Predicting Protein Secondary and Supersecondary Structure. CRC Press LLC, Florida.
- 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.
- 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.
- Velasco R, Zharkikh A, Troggio M, Cartwright DA, et al. (2007). A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS One* 2: e1326.
- Wang C, Wang H, Zhang J and Chen S (2008). A seed-specific AP2-domain transcription factor from soybean plays a certain role in regulation of seed germination. *Sci. China C Life Sci.* 51: 336-345.
- Wang Y, Deng D, Bian Y, Lv Y, et al. (2010). Genome-wide analysis of primary auxin-responsive Aux/IAA gene family in maize (*Zea mays* L.). *Mol. Biol. Rep.* 37: 3991-4001.
- Yamamoto S, Suzuki K and 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.
- Yin XR, Allan AC, Chen KS and Ferguson IB (2010). Kiwifruit *EIL* and *ERF* genes involved in regulating fruit ripening. *Plant Physiol.* 153: 1280-1292.
- Zhang GY, Chen M, Chen XP and Xu ZS (2008). Phylogeny, gene structures, and expression patterns of the ERF gene family in soybean (*Glycine max* L.). *J. Exp. Bot.* 59: 4095-4107.