



Genome-wide analysis of the auxin response factor gene family in cucumber

S.Q. Liu² and L.F. Hu¹

¹Key Laboratory of Crop Physiology, Ecology and Genetic Breeding, Ministry of Education, Jiangxi Agricultural University, Nanchang, China

²School of Sciences, Jiangxi Agricultural University, Nanchang, China

Corresponding author: L.F. Hu

E-mail: lfhu_hn337@163.com

Genet. Mol. Res. 12 (4): 4317-4331 (2013)

Received July 31, 2012

Accepted January 31, 2013

Published April 2, 2013

DOI <http://dx.doi.org/10.4238/2013.April.2.1>

ABSTRACT. Auxin response factors (ARFs) participate in the transcriptional regulation of a series of biological processes related to growth and development. The *ARF* genes comprise a large multigene family in plants. Recently, a draft of the full cucumber (*Cucumis sativus*) genome assembly has been released; however, none of the *ARF* genes have been characterized. We made a comprehensive analysis of *ARF* genes in this species. Fifteen *ARF* genes were identified and could be divided into three classes. Intron presence and position were conserved to some extent within one phylogenetic group. Based on genome distribution analysis, tandem duplication appears to have contributed to *ARF* gene expansion. Specific motifs were selectively distributed among specific clades. Putative cis-elements involved in auxin response, light signaling responses, phytohormones, and defense responses were identified in promoter regions of the *ARF* genes. Almost all of the *ARF* genes exhibited a constitutive expression pattern, implying that there are functional redundancies among the ARF proteins.

Key words: *Cucumis sativus* L.; Auxin response factors; Phylogenetic analysis; Transcription factor; Genome sequence

INTRODUCTION

Auxins play a pivotal role in most major aspects of plant growth and development processes, including apical dominance, tropic response, vascular differentiation, embryo patterning, and root and shoot architecture (Quint and Gray, 2006). At the molecular level, most of these processes are regulated by auxin-response genes such as *Aux/IAA*, *Gretchen Hagen3 (GH3)* and *Small Auxin Up RNA (SAUR)*, which contain a conserved cis-responsive element TGTCTC called an auxin-responsive element (AuxRE) in their promoters (Guilfoyle and Hagen, 2007). Transcription factors that bind specifically to AuxRE and mediate auxin response, are called auxin response factors (ARFs). A typical ARF protein consists of three conserved domains, that is, the plant-specific B3-like DNA-binding domain (DBD) at the NH₂-terminus, the activation domain (AD) or repression domain (RD) in the middle region, and a conserved C-terminal dimerization domain (CTD). Usually, DBD cannot combine with AuxREs by itself, where other amino acids are needed. Moreover, the binding of ARF proteins with AuxREs through DBD is independent of auxin utilization. The amino acid composition of the middle region is critical in determining whether an ARF protein will act as an activator or repressor. AD is characterized by glutamine (Q) enrichment, while RD is rich in serine (S), proline (P), threonine (T), and glycine (G) residues (Ulmasov et al., 1999; Quint and Gray, 2006). CTD is very similar to domains III and IV of *Aux/IAA* proteins, which is responsible for the homo-dimerization of ARFs themselves and also the hetero-dimerization of ARFs and *Aux/IAAs* proteins. The ARF proteins are a large gene family, with 23, 25, 39, and 31 members in *Arabidopsis*, rice, *Populus trichocarpa*, and maize, respectively (Kalluri et al., 2007; Wang et al., 2007; Xing et al., 2011). Classical genetic strategies have led to the identification of *ARF* gene roles in plant growth and development. For instance, *arf* mutations result in the suppression of the hookless phenotype and hypocotyl bending (*AtARF2*) (Sessions et al., 1997; Ellis et al., 2005; Schruoff et al., 2006), defects in flower formation and patterning (*AtARF3*) (Nemhauser et al., 2000; Nishimura et al., 2005), abnormal vascular development and embryo axis formation (*AtARF5*) (Hardtke and Berleth, 1998; Vidaurre et al., 2007), impaired hypocotyl response to blue light, growth and auxin sensitivity (*AtARF7*) (Harper et al., 2000; Tatematsu et al., 2004), delayed fruit development and seed formation (*AtARF8*) (Goetz et al., 2006), and resistance to auxin and ethylene treatments (*AtARF19*) (Li et al., 2006). Moreover, phenotypes of double mutants for some sister pairs of ARF, e.g., *AtARF1/AtARF2*, *AtARF3/AtARF4*, *AtARF5/AtARF7*, *AtARF6/AtARF8*, and *AtARF10/AtARF16*, are more severe than the single mutant (Ellis et al., 2005; Nagpal et al., 2005; Wang et al., 2005). For example, the single *arf10* or *arf16* mutant has shown no obvious difference compared with wild type, but obvious defects in root tip and gravitropism development have been seen in the *arf10/arf16* double mutant (Wang et al., 2005). Functional redundancies of these ARF proteins may contribute to the phenomenon.

Cucumber (*Cucumis sativus* L.) is an economically and nutritionally important vegetable crop cultivated world-wide. The first assembly of the cucumber genome sequence has recently been published (Huang et al., 2009); however, to the best of our knowledge, no member of the cucumber ARF family has been characterized up to now. The aim of this study is the identification and phylogenetic and expression analysis of cucumber *ARF* genes, as well as a comparative analysis with *Arabidopsis*, rice and maize *ARF* members, which will help future studies in elucidating the precise roles of *CsARF* in cucumber growth and development.

MATERIAL AND METHODS

Database search for cucumber *ARF* genes

A hidden Markov model (HMM) was employed to identify the *ARF* genes from the cucumber genome. The *Arabidopsis* *ARF* sequences reported by Wang et al. (2007) were downloaded from the TIGR *Arabidopsis* annotated protein database (<http://www.tigr.org/tdb/e2k1/ath1/>) and their HMM profile was generated using the HMMER 3.0 software package (Durbin et al., 1998). This profile was used to search the complete proteome of cucumber available in CuGI (<http://cucumber.genomics.org.cn>), which was released by The Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences (IVF-CAAS). The redundant sequences at the same chromosome locus were removed from our data set. To further confirm the accuracy of these predicted genes, their amino acid sequences were then searched based on the conserved domains using the Simple Modular Architecture Research Tool (SMART) (Letunic et al., 2004).

Tree building

Multiple sequence alignments were performed on the *ARF* protein sequences using Clustal X (<http://www.clustal.org/>) with the default parameters (Larkin et al., 2007), and the alignments were then adjusted manually. A phylogenetic tree was constructed with aligned *ARF* protein sequences using MEGA4 (<http://www.megasoftware.net/mega4/mega.html>) (Tamura et al., 2007) employing the neighbor-joining (NJ) method with the following parameters: Poisson correction, pairwise deletion, and bootstrap (1000 replicates). The constructed tree file was visualized by TreeView 1.6.6 (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>) (Page, 1996).

Intron/exon structure, genome distribution and segmental duplication

The DNA and cDNA sequences corresponding to each predicted gene from the cucumber genome and annotation database CuGI were downloaded, and the intron distribution pattern and splicing phase were then analyzed using the web-based bioinformatics tool GSDS (<http://gsds.cbi.pku.edu.cn/>). To obtain the information of cucumber *ARF* gene location, we drew a map of the distribution of *CsARF* genes throughout the cucumber genome using the MapInspect tool (http://www.plantbreeding.wur.nl/UK/software_mapinspect.html). To detect the segment duplicated events, we analyzed the 100 kb DNA segments flanking each *CsARF* gene. Regions in different linkage groups that contained 6 or more homologous pairs with fewer than 25 nonhomologous gene interventions were defined as duplicated segments. A gene pair was considered tandem duplicated if it was separated by fewer than 5 intervening genes and shared $\geq 40\%$ amino acid sequence similarity. BioEdit 5.0.6 (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>) (Hall, 1999) was used to analyze the *CsARF* homologs for similarity on the phylogenetic tree.

Conserved motif prediction

We searched for conserved motifs among cucumber *ARF* proteins using the MEME

tool (<http://meme.sdsc.edu/meme/cgi-bin/meme.cgi>) (Bailey and Elkan, 1994) to find similar sequences shared by these members.

Promoter regions and expression analysis of cucumber *ARF* genes

The 1.5-kb DNA sequences upstream of the start codon (ATG) corresponding to each *ARF* gene from the cucumber genome database CuGI were downloaded, and the putative cis-elements were then analyzed using the online PLACEcare web tool (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>).

PCR primers were designed to avoid the conserved regions. Primer sequences are shown in detail in [Table S1](#). Seeds of the ‘Chinese long’ 9930 inbred line, commonly used in modern cucumber breeding (Huang et al., 2009), were germinated and grown in trays containing soil mixture (peat:sand:pumice, 1:1:1, v/v/v). Plants were adequately watered and grown at day/night temperatures of 24°/18°C with a 16-h photoperiod. Total RNA of root, stem, leaf, flower, and fruit of cucumbers at the stage of 20 main-stem nodes was isolated using the TRIzol reagent (Invitrogen, USA). RT-PCR was carried out according to manufacturer instructions (Tiangen Biotech Co. Ltd., Beijing China). The cucumber actin cDNA fragment (161 bp) was employed as the internal standard for each gene.

RESULTS

Identification of 15 *CsARF* genes

To identify the full complement of *ARF* genes in cucumber, we built an ARF domain hidden Markov model using the HMMER program based on the multiple sequence alignment results of *Arabidopsis* ARF protein domain sequences (Wang et al., 2007). The cucumber amino acid sequences in CuGI (<http://cucumber.genomics.org.cn>) were searched using the Hmsearch command with a probability E value threshold of 0.1, as recommended by the HMMER user guide, and a total of 15 candidate ARF-domain sequences were obtained. To further verify the reliability of these candidate sequences, we performed SMART analysis of these putative ARF-domain sequences and found that all 15 candidate genes had a typical ARF domain. These 15 cucumber *ARF* genes (*CsARF*; Table 1) were subjected to further analyses. The number designation was based on the position from the top to the bottom on the cucumber chromosomes 1 to 7.

Phylogenetic analysis of the *CsARF* genes

To determine the evolutionary relations of the CsARF proteins, an unrooted NJ phylogenetic tree using bootstrap analysis (1000 replicates) was constructed based on the multiple sequence alignments of 15 CsARF proteins. Similar to findings in rice (Wang et al., 2007) and *Solanum lycopersicum* (Wu et al., 2011), the 15 CsARF members could be classified into three classes (I-III), each containing 6, 5 and 4 members, respectively (Figure 1). Class III seemed more divergent compared with those grouped into classes I and II. It was worth noting that 10 CsARF proteins fell into related sister pairs, namely CsARF06/CsARF08, CsARF01/CsARF07, CsARF03/CsARF04, CsARF09/CsARF02, and CsARF10/CsARF11, with a very

Table 1. *Cs:ARF* genes identified in this study according to their *Cs:ARF* numbers determined by the position from the top to the bottom on the cucumber chromosomes 1 to 7.

Serial No.	Gene name	Gene ID	Chromosomal location	Length of chromosomes (bp)	CuGI (5'-3')	Length (amino acids)	Group name
1	<i>Cs:ARF01</i>	Csa005276	1	25430643	51893-56226	841	I
2	<i>Cs:ARF02</i>	Csa021954	1	25430643	15603180-15598000	831	IIa
3	<i>Cs:ARF03</i>	Csa009210	2	22000590	6039111-6044343	862	IIa
4	<i>Cs:ARF04</i>	Csa012237	2	22000590	4179147-4186177	1102	IIa
5	<i>Cs:ARF05</i>	Csa017897	3	34445220	30715082-30720189	1017	IIa
6	<i>Cs:ARF06</i>	Csa019361	4	22434855	2776072-2773109	616	I
7	<i>Cs:ARF07</i>	Csa010564	5	28091742	11527358-11518543	677	I
8	<i>Cs:ARF08</i>	Csa012805	5	28091742	16166427-16160736	680	I
9	<i>Cs:ARF09</i>	Csa019265	5	28091742	17879709-17889691	647	IIa
10	<i>Cs:ARF10</i>	Csa011935	6	26547709	18407965-18410402	694	III
11	<i>Cs:ARF11</i>	Csa015176	6	26547709	8680383-8677529	716	III
12	<i>Cs:ARF12</i>	Csa020090	6	26547709	12244429-12250875	822	IIb
13	<i>Cs:ARF13</i>	Csa007296	7	17370973	10311115-10307714	693	I
14	<i>Cs:ARF14</i>	Csa020560	Scaffold0000166	207881	132914-135409	703	III
15	<i>Cs:ARF15</i>	Csa022361	Scaffold0000319	21867	1498-3796	549	III

strong bootstrap support (above 99%). To investigate the evolutionary relations of cucumber ARFs with those in other species, a phylogenetic tree was built from alignments of the complete protein sequences of 15 CsARFs, 25 rice OsARFs, 23 *Arabidopsis* AtARFs and 31 maize ZmARFs (Figure 2). According to the Wu et al. (2011) classification criterion, all these 94 ARFs were grouped into I-IV classes. Class I contained 31 members and could be divided into groups Ia (23 members) and Ib (8 members), where group Ia could be further classified into subgroup Ia-1(15 members) and Ia-2 (8 members). Class II had 42 members including 30 members in subgroup IIa and 12 members in IIb. Class III and IV contained 17 and 4 members, respectively. Interestingly, groups Ia, IIa, IIb, class III and class IV contained ARF members from all four species, while group Ib only had members from *Arabidopsis*. Moreover, in class IV, only monocotyledon (rice and maize) ARFs members were exhibited. In the combined phylogenetic tree, a total of 33 sisters including 13 OsARF/ZmARF pairs, 8 CsARF/AtARF pairs, 2 AtARF/AtARF pairs, 8 ZmARF/ZmARF pairs, and 2 CsARF/CsARF pairs were distributed. The maximum number of members of sister pairs was found in group IIa with 11 sister members. In addition to the above 1:1 orthologous relationship, the 1:n and n:1 orthologous relationships such as single ZmARF and multiple OsARF, single OsARF and multiple ZmARF, as well as single CsARF and multiple AtARF, were also demonstrated. The third n:n orthologous relationship was found in the class III cluster ZmARF8/ZmARF15/OsARF08/OsARF10 and ZmARF05/ZmARF31/OsARF13/OsARF20 in class IV.

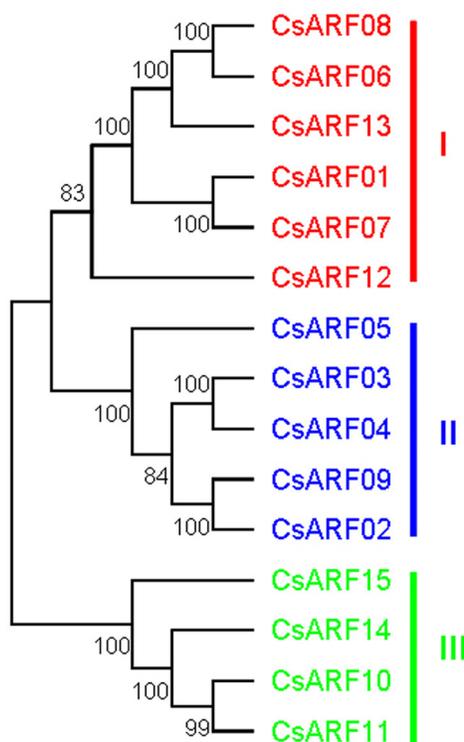


Figure 1. Neighbor-joining tree of cucumber auxin response factor (ARF) sequences. Branches with less than 50% bootstrap support are collapsed.

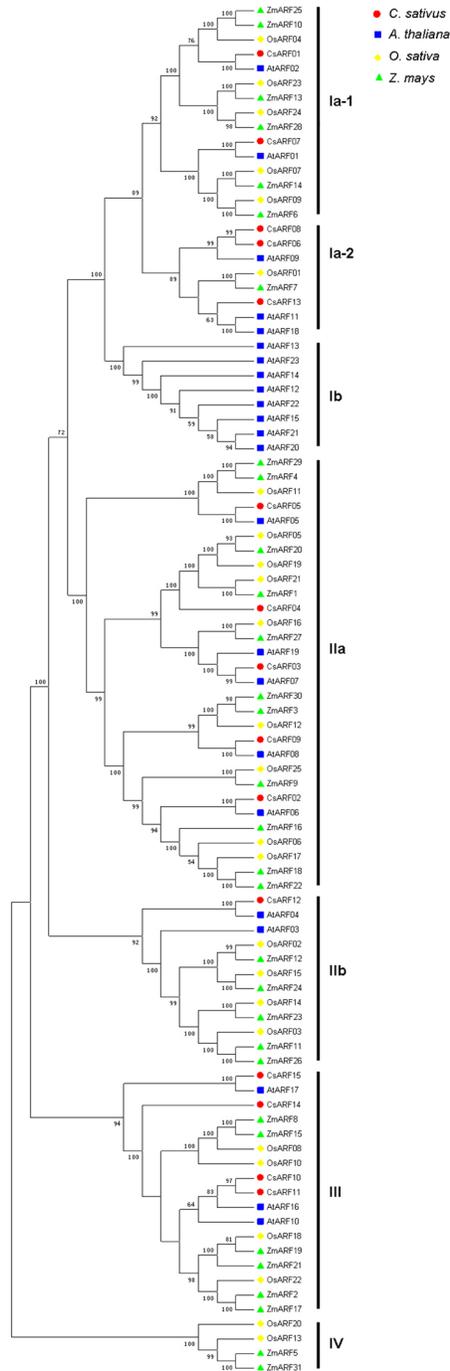


Figure 2. Neighbor-joining tree of cucumber, *Arabidopsis*, rice, and maize auxin response factor (ARF) sequences. Branches with less than 50% bootstrap support are collapsed.

Structure and evolution of the *CsARF* genes

Generally, the pattern of intron positions can provide some clues on evolutionary relationships. The previous study revealed that the rice and *Arabidopsis* class I and II ARF members contained multiple introns, while class III genes had fewer introns. To investigate the intron numbers and positions, a comparison of the full-length cDNA sequences with the corresponding genomic DNA sequences was performed. As expected, a very striking bimodal distribution of introns was observed in cucumber ARF family genes, similar to that in *Arabidopsis* and rice. For instance, among the 11 *CsARF* members in class I and II, 9 (82%) contained at least 10 introns with a maximum (17 introns) for *CsARF05* in class II, whereas for the 4 class III members, just one (*CsARF11* and *CsARF15*) or two introns (*CsARF10* and *CsARF14*) were observed (Figure 3). Further analysis showed that, among the 15 intron positions distributed in the 6 class I members, a total of 13 positions (13/15) were highly conserved (indicated with solid arrows). Analysis of the intron phases revealed that 6 positions had phase 0 (splicing occurring after the third nucleotide of the codon), 3 positions had phase 1 (splicing occurring after the first nucleotide of the codon), and the remaining 4 positions had phase 2 (splicing occurring after the second nucleotide of the codon). Compared with class I, despite that more than 13 conserved positions were also shown in class II, the un-conserved intron positions (indicated with open arrows) seemed more frequent, e.g., 3 unconserved intron positions were found in *CsARF05*, and 1 between intron 11 and 12 in the 4 class II members. In the 4 class III members, only two conserved positions (5 and 13) were found. On the whole, the presence and position were conserved to some extent within one phylogenetic group, indicating that the above classification of cucumber ARF family genes was reliable.

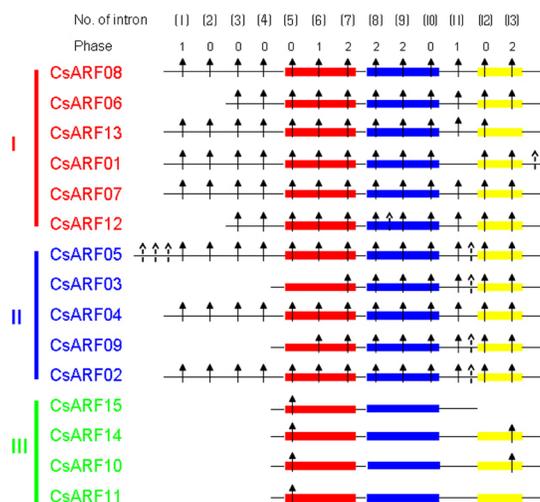


Figure 3. Scheme of the intron distribution patterns of 15 cucumber auxin response factor (*ARF*) genes. The red rectangle corresponds to the DNA-binding domain, the blue rectangle to the AUX_RESP domain, and the yellow rectangle to the carboxy-terminal domain. The upward arrows indicate the position of the intron corresponding to the example. The numbers above the arrows indicate the splicing phases, 0 refers to phase 0, 1 to phase 1, and 2 to phase 2. The numbers 1 to 13 beside the arrows mean different positions of the introns. Here the position of introns in the variable region has been adjusted manually to make them more contracted.

To gain insight into the evolution of 15 cucumber *ARF* genes, we analyzed their genomic distribution and found that they were unevenly distributed on chromosomes 1-7, except for 2 members lying within the unassembled Scaffold sequence (Figure 4; Table 1). There were 3 *CsARF*s present on chromosomes 5 and 6, 2 on chromosomes 1 and 2, and 1 on chromosomes 3, 4 and 7 (Figure 4; Table 1). With respect to the 3 classes, the class I genes were confined to chromosomes 1, 4, 5, and 7, the class II genes to chromosomes 1, 2, 3, 5, and 6, and the class III genes to chromosome 6. Interestingly, *CsARF03* and *CsARF04* as sister pairs were located on the same chromosome 5, as well as *CsARF10* and *CsARF11* on the same chromosome 6. However, three other sister pairs, *CsARF06/CsARF08*, *CsARF01/CsARF07*, *CsARF02/CsARF09*, were present on different chromosomes, suggesting that they may be derived from recent gene duplication events.

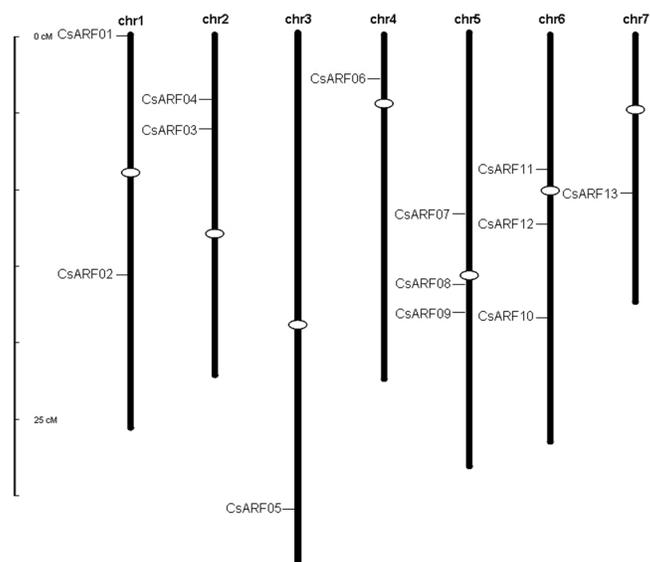


Figure 4. Chromosomal localization of 15 cucumber auxin response factor (*ARF*) genes. The scale is in megabases (Mb). The white ovals in the middle of the 7 chromosomes show the centromeric positions according to the sequencing result of cucumber genome (Huang et al., 2009). Two genes, *CsARF14* that lie within unassembled Scaffold000166 and *CsARF15* that lies within unassembled Scaffold000319, cannot be anchored on specific chromosome.

Sequence analysis of CsARF proteins and conserved motif prediction

The open reading frame length of 15 *CsARF* genes varied from 1650 (*CsARF15*) to 3309 bp (*CsARF04*), encoding polypeptides of 549-1102 amino acids, with a predicted molecular weight range of 60.15-121.35 kDa. Multiple alignment demonstrated that all *CsARF* proteins contained a highly conserved region of about 320 amino acid residues in their N-terminal portion corresponding to DBD of the *Arabidopsis* ARF family (Table S2). Except for *CsARF15*, all *CsARF* proteins contained a CTD related to AUX/IAA protein domains III and IV (Table S2). It was believed that the middle region domain located between DBD and CTD was variable, may function as either AD or BD based on the amino acid composition. The previous study showed that ADs are rich in Q residues, while RDs are rich in PST residues. A detailed sequence analy-

sis revealed that PST-rich middle regions were found in CsARF07 and CsARF13, indicating that they were more likely to act as repressors. While Q-rich regions were found in CsARF02, CsARF03 and CsARF04, implying that they were probably transcriptional activators.

For further investigation of the characterized regions of cucumber ARF proteins, the online MEME utility was employed to analyze motif distribution in 15 cucumber ARFs. A total of 20 motifs named motifs 1-20 were identified (Figure 5; Table S3). Among them, the conserved motifs 1 and 15, specifying DBD, were observed in all cucumber ARF members, except for the lack of motif 1 in CsARF03 and CsARF09. Motifs 4, 5, 6, 9, and 11, specifying the AUX_RESP domain, were found in almost all members. Motifs 2 and 7, corresponding to the AUX_IAA domain, were observed in all cucumber ARF members except CsARF15. At the same time, it is worth noting that motif 20 was restricted to class I, motif 15 to class II, and motifs 16, 17 and 18 to class III. In addition, motifs 13 and 19 were confined to class I and II. Therefore, to a certain extent, the scheme of conserved motifs of each CsARF family protein demonstrated structural similarities among the proteins within one group, which supported the above subfamily designation of cucumber ARF family genes.

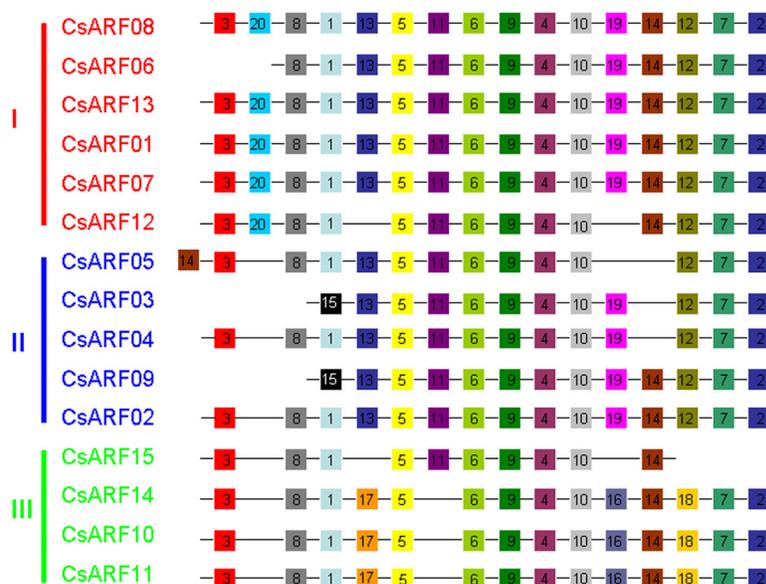


Figure 5. Distribution of conserved motifs in cucumber auxin response factor (ARF) proteins identified using the MEME search tool. Each motif is represented by a colored box and a number. Order of the motifs corresponds to position of motifs in individual protein sequence. For details of motifs, see Table S3.

Promoter regions and expression analysis of 15 *CsARF* genes

Using the online PLACEcare web tool (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>), we analyzed the putative cis-elements in 1.5 kb genomic DNA sequences upstream of the start codon (ATG). Surprisingly, among the 15 *CsARF* genes, auxin response elements (AuxRR-core) were only detected in promoter regions of *CsARF03*, *CsARF08* and *CsARF13* (Table S4), implying that the relationship between related response elements

and auxin inducibility was unclear and needed to be further investigated. On the other hand, putative cis-elements involved in light signaling perception were observed in all ARF members (Table S5), where the possible reason was likely the strong crosstalk between auxin and light signals (Colon-Carmona et al., 2000; Liscum and Reed, 2002). In addition, putative cis-elements involved in hormones such as abscisic acid, gibberellin acid, ethylene and methyl jasmonic acid were found in a series of CsARF promoter regions (Table 6). Also, putative cis-elements associated with biotic and/or abiotic stress adaptive elements, cold and dehydration-responsive element, low temperature-responsive element, and heat stress-responsive element were identified in most ARF members (Table S7). The evidence of diverse cis-elements in promoter regions of CsARFs supported the fact that ARF proteins can be linkers between auxin response and other important metabolic pathways, and thus, the possible regulatory network deserves further research (Li et al., 2006; Vert et al., 2008). To get insight into the functions of the 15 *CsARF* genes during different cucumber development processes, we performed RT-PCR analysis of RNA from root, stem, leaf, flower, and fruit. Similar to a previous study in *Arabidopsis* and rice (Wang et al., 2007), we found that except for *CsARF03* (expressed in stem, leaf, flower and fruit) and *CsARF04* (expressed in stem and fruit), 13 other *CsARF* genes could be detected in all 5 tissues investigated, suggesting the widespread function of the *CsARF* gene family within plant tissues (Figure 6). Among the 13 *CsARF* genes, 2 genes (*CsARF05* and *CsARF06*) showed a similar expression pattern with a high transcript signal in root, flower and fruit, 4 genes (*CsARF01*, *CsARF02*, *CsARF09*, and *CsARF13*) displayed similar patterns with a high transcript signal in shoot and flower or fruit, and the remaining 7 genes (*CsARF07-08*, *CsARF10-12*, and *CsARF14-15*) exhibited similar transcript levels in all 5 tissues investigated.

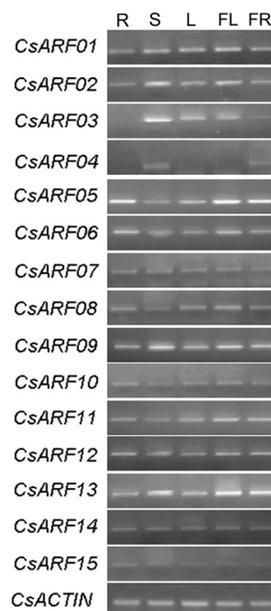


Figure 6. Expression analysis of cucumber *ARF* genes in different tissues using RT-PCR. RT-PCR was performed using primers specific for the *CsARF* genes. PCR products were run on 1.5% agarose gels. *CsACTIN* primers that generated a 161-bp product were used as the inner standard for each gene. Sources of the samples are as follows: root (R), stem (S), leaf (L), flower (FL), and fruit (FR).

DISCUSSION

Characterization of an expanded ARF gene family in cucumber

A previous study indicated that 7 duplicated AtARFs (12, 13, 14, 15, 20, 21, 22, and 23) in a single cluster (corresponding to group Ib in our study) are physically located near each other in a region proximal to the centromere of chromosome 1. They are probably originally derived from one *AtARF* gene. This evident tandem duplication was also another force contributing to increase the number of AtARFs except for the whole genome and segmental duplication events (Theologis et al., 2000; Hagen and Guilfoyle, 2002; Remington et al., 2004; Okushima et al., 2005). Huang et al. (2009) believed that in the cucumber genome, the event of recent whole-genome duplication is absent, but a number of tandem duplications have occurred. However, whether or not a segment duplicated event existed in the cucumber ARF family was unknown, and thus, we performed segment duplication analysis on 100-kb DNA segments flanking each *CsARF* gene and found that no gene was attributed to segmental duplication. Taken together, tandem duplication most likely played a crucial role in the observed gene expansion of the cucumber *ARF* genes.

Phylogenetic relationship between CsARF, AtARF, OsARF, and ZmARFs

On the basis of phylogenetic analysis, the organization of cucumber ARF proteins was very similar to that in other species (Wang et al., 2007; Xing et al., 2011), implying that most of the ARFs derived from a common ancestor. Class Ib was a special group that only contained AtARFs, where no CsARF, OsARF and ZmARF members were included, and the segregation to a separate group suggested that the function of these proteins may be species-specific. Their role in *Arabidopsis* deserves further research, and at the same time, it would be interesting to characterize the possible mechanism underlying their absence in the cucumber, rice and maize genomes. The ARFs in class IV were all from the monocots maize and rice, indicating that this class may be lost in dicots after divergence of monocots and dicots or evolved solely in monocots after the divergence. In the combined phylogenetic tree, there were 3 kinds of orthologous relationships. The first 1:1 orthologous relationship was observed in 33 sister pairs including 21 interspecies sister pairs, indicating that these genes were descended from a common ancestor and corresponded to well-conserved functions (Wang et al., 2007; Wu et al., 2011). Interestingly, the 21 interspecies sister pairs were only observed between the monocots maize and rice and dicots cucumber and *Arabidopsis*, while no sister pair was found between monocots and dicots, suggesting that the common ancestor of these sister pairs more likely appeared after the divergence of monocots and dicots (Wang et al., 2007; Wu et al., 2011). It was worth mentioning that 8 cucumber ARF members (*CsARF01-03*, *CsARF05*, *CsARF07*, *CsARF09*, *CsARF12*, and *CsARF15*) showed a sister pair relationship with *Arabidopsis*-related ARF members, i.e., *CsARF01/AtARF02*, *CsARF02/AtARF06*, *CsARF03/AtARF07*, *CsARF05/AtARF05*, *CsARF07/AtARF01*, *CsARF09/AtARF08*, *CsARF12/AtARF04*, and *CsARF15/AtARF17*. In view that *Arabidopsis AtARF1* and *AtARF2* participate in the regulation of seed size and weight (Ellis et al., 2005; Schruoff et al., 2006), combined with the occurrence of *CsARF01* and *CsARF07* mRNA transcripts in fruit, we propose that *CsARF01* and *CsARF07* play similar roles in cucumber seed development. Similarly, the possible functions of *CsARF02*

and *CsARF09* in flower maturation like its sister pair members *AtARF06* and *AtARF08* (Goetz et al., 2006), *CsARF05* in vascular development like *AtARF05* (Hardtke and Berleth, 1998; Vidaurre et al., 2007), *CsARF12* in leaf identity and morphogenesis like *AtARF04* (Hunter et al., 2006), and *CsARF15* in root development like *AtARF17* were expected (Rhoades et al., 2002). The 1:n (*AtARF09/CsARF06/CsARF08*) and n:1 (*AtARF19/CsARF03/AtARF07*) orthologous relationships in *Arabidopsis* and cucumber ARF members suggested that function might have begun to diversify in one species due to gene duplication (Wang et al., 2007; Wu et al., 2011). It was reported that *AtARF9* negatively regulates glucosinolate accumulation and positively regulates camalexin accumulation in pathogen interactions (Ulmasov et al., 1999; Hagen and Guilfoyle, 2002). *AtARF19* acts redundantly with *atNPH4/ARF7* in controlling leaf expansion and lateral root growth (Li et al., 2006). The roles of their orthologous members in cucumber need to be further investigated. The third n:n orthologous relationships (*ZmARF8/ZmARF15/OsARF08/OsARF10* and *ZmARF05/ZmARF31/OsARF13/OsARF20*) suggested the function had started diversifying in both species (Wang et al., 2007; Wu et al., 2011).

Conserved domains and motif analysis

In general, the features and number of domains present in protein sequences is always helpful for predicting the function of unknown genes, a careful analysis of the identified protein sequences is the first attempt to find useful clues about the roles of the corresponding *CsARF* genes. As we know, the DBD of a typical ARF protein plays a critical role in regulating the expression of auxin response genes. Here, the DBD was observed in all 15 *CsARF* proteins, although *CsARF03*, *CsARF06* and *CsARF09* contained a truncated DBD. Contrarily, the 4 genes *CsARF10*, *CsARF11*, *CsARF14*, and *CsARF15*, like *Arabidopsis AtARF10*, *AtARF16* and *AtARF17* and maize *ZmARF1*, were found to have additional sequences in the DBD (Table S2, indicated in red text). We believe that the less conserved features of the DNA-binding domain resulted from a rapid divergence during evolution, which was likely the reason for DBD divergence in cucumber ARF members.

Previous studies considered that the middle regions of ARFs function as ADs or RDs according to the amino acid composition. Our analysis revealed that the 2 members *CsARF07* and *CsARF13* are more likely to act as repressors, while 3 other members, *CsARF02*, *CsARF03* and *CsARF04*, as possible activators. Moreover, the phylogenetic analysis of *CsARF* proteins demonstrated that the activators and repressors were appropriately placed in separate groups, for example, the activator ARFs were present in class II, and the repressors in class I. Undoubtedly, the suppressive or stimulatory role of cucumber ARF proteins should be further confirmed by relevant experiments.

CTD is a protein-protein interaction domain showing high similarity to Aux/IAA protein domains III and IV, which are required for ARF-ARF and ARF-IAA dimerization. Interestingly, previous studies demonstrated a group of CTD-truncated ARFs in *Arabidopsis*, rice, maize and tomato, where these CTD-truncated ARFs may regulate gene expression in an auxin-independent manner (Wang et al., 2007; Kumar et al., 2011; Xing et al., 2011). Here, of all the 15 *CsARFs* examined, only *CsARF15* lacked the C-terminal domains, which was dramatically less than those in *Arabidopsis* (4 out of 23 ARFs), tomato (6 out of 21 ARFs), and maize (9 out of 31 ARFs), suggesting that the CTD is very important in cucumber ARF members. As expected, motif analysis revealed that most of the motifs were selectively dis-

tributed among the specific groups in the phylogenetic tree, demonstrating common structural features within the same group. In view that commonly shared motifs tend to be of functional and structural importance, it will be worthwhile to characterize the functions of the specific motifs, e.g., motif 20 in class I, motif 15 in class II and motifs 16-18 in class III.

In summary, we performed a comprehensive genome-wide analysis on 15 cucumber *ARF* genes, with regard to phylogeny, genome distribution, gene structures, conserved motifs, and expression analysis. The 15 *CsARF* genes could be divided into classes I to III, which was supported to a certain extent by the presence and position of introns and the distribution of conserved amino acid sequence motifs. Genome distribution of *CsARF* genes revealed that tandem duplication may contribute to the expansion of *CsARF* genes. The comparative and phylogenetic analyses of the *CsARF* gene family in cucumber, rice, *Arabidopsis*, and maize and the data of *CsARF* gene expression will facilitate a comprehensive functional characterization of the *ARF* gene family. The results presented here will provide useful clues for elucidating the detailed roles of CsARFs in plant growth and development in future studies.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#31060262), the Natural Science Foundation of Jiangxi, China (#2009GQN0034) and the Foundation of Jiangxi Agricultural University, Jiangxi, China (2009).

[Supplementary material](#)

REFERENCES

- Bailey TL and Elkan C (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2: 28-36.
- Colon-Carmona A, Chen DL, Yeh KC and Abel S (2000). Aux/IAA proteins are phosphorylated by phytochrome *in vitro*. *Plant Physiol.* 124: 1728-1738.
- Durbin R, Eddy S, Krogh A and Mitchison G (1998). Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. A Tutorial Introduction to Hidden Markov Models and Other Probabilistic Modelling Approaches in Computational Sequence Analysis. Cambridge University Press, Cambridge.
- Ellis CM, Nagpal P, Young JC, Hagen G, et al. (2005). AUXIN RESPONSE FACTOR1 and AUXIN RESPONSE FACTOR2 regulate senescence and floral organ abscission in *Arabidopsis thaliana*. *Development* 132: 4563-4574.
- Goetz M, Vivian-Smith A, Johnson SD and Koltunow AM (2006). AUXIN RESPONSE FACTOR8 is a negative regulator of fruit initiation in *Arabidopsis*. *Plant Cell* 18: 1873-1886.
- Guilfoyle TJ and Hagen G (2007). Auxin response factors. *Curr. Opin. Plant Biol.* 10: 453-460.
- Hagen G and Guilfoyle T (2002). Auxin-responsive gene expression: genes, promoters and regulatory factors. *Plant Mol. Biol.* 49: 373-385.
- Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41: 95-98.
- Hardtke CS and Berleth T (1998). The *Arabidopsis* gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO J.* 17: 1405-1411.
- Harper RM, Stowe-Evans EL, Luesse DR, Muto H, et al. (2000). The NPH4 locus encodes the auxin response factor ARF7, a conditional regulator of differential growth in aerial *Arabidopsis* tissue. *Plant Cell* 12: 757-770.
- Huang S, Li R, Zhang Z, Li L, et al. (2009). The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* 41: 1275-1281.
- Hunter C, Willmann MR, Wu G, Yoshikawa M, et al. (2006). Trans-acting siRNA-mediated repression of ETTIN and ARF4 regulates heteroblasty in *Arabidopsis*. *Development* 133: 2973-2981.
- Kalluri UC, Difazio SP, Brunner AM and Tuskan GA (2007). Genome-wide analysis of Aux/IAA and ARF gene families in *Populus trichocarpa*. *BMC Plant Biol.* 7: 59.

- Kumar R, Tyagi AK and Sharma AK (2011). Genome-wide analysis of auxin response factor (ARF) gene family from tomato and analysis of their role in flower and fruit development. *Mol. Genet. Genomics* 285: 245-260.
- Larkin MA, Blackshields G, Brown NP, Chenna R, et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948.
- Letunic I, Copley RR, Schmidt S, Ciccarelli FD, et al. (2004). SMART 4.0: towards genomic data integration. *Nucleic Acids Res.* 32: D142-D144.
- Li J, Dai X and Zhao Y (2006). A role for auxin response factor 19 in auxin and ethylene signaling in *Arabidopsis*. *Plant Physiol.* 140: 899-908.
- Liscum E and Reed JW (2002). Genetics of Aux/IAA and ARF action in plant growth and development. *Plant Mol. Biol.* 49: 387-400.
- Nagpal P, Ellis CM, Weber H, Ploense SE, et al. (2005). Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. *Development* 132: 4107-4118.
- Nemhauser JL, Feldman LJ and Zambryski PC (2000). Auxin and ETTIN in *Arabidopsis gynoecium* morphogenesis. *Development* 127: 3877-3888.
- Nishimura T, Wada T, Yamamoto KT and Okada K (2005). The *Arabidopsis* STV1 protein, responsible for translation reinitiation, is required for auxin-mediated gynoecium patterning. *Plant Cell* 17: 2940-2953.
- Okushima Y, Overvoorde PJ, Arima K, Alonso JM, et al. (2005). Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in *Arabidopsis thaliana*: unique and overlapping functions of ARF7 and ARF19. *Plant Cell* 17: 444-463.
- Page RD (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12: 357-358.
- Quint M and Gray WM (2006). Auxin signaling. *Curr. Opin. Plant Biol.* 9: 448-453.
- Remington DL, Vision TJ, Guilfoyle TJ and Reed JW (2004). Contrasting modes of diversification in the Aux/IAA and ARF gene families. *Plant Physiol.* 135: 1738-1752.
- Rhoades MW, Reinhart BJ, Lim LP, Burge CB, et al. (2002). Prediction of plant microRNA targets. *Cell* 110: 513-520.
- Schruff MC, Spielman M, Tiwari S, Adams S, et al. (2006). The AUXIN RESPONSE FACTOR 2 gene of *Arabidopsis* links auxin signalling, cell division, and the size of seeds and other organs. *Development* 133: 251-261.
- Sessions A, Nemhauser JL, McColl A, Roe JL, et al. (1997). ETTIN patterns the *Arabidopsis* floral meristem and reproductive organs. *Development* 124: 4481-4491.
- 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.
- Tatematsu K, Kumagai S, Muto H, Sato A, et al. (2004). *MASSUGU2* encodes *Aux/IAA19*, an auxin-regulated protein that functions together with the transcriptional activator NPH4/ARF7 to regulate differential growth responses of hypocotyl and formation of lateral roots in *Arabidopsis thaliana*. *Plant Cell* 16: 379-393.
- Theologis A, Ecker JR, Palm CJ, Federspiel NA, et al. (2000). Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. *Nature* 408: 816-820.
- Ulmasov T, Hagen G and Guilfoyle TJ (1999). Activation and repression of transcription by auxin-response factors. *Proc. Natl. Acad. Sci. U. S. A.* 96: 5844-5849.
- Vert G, Walcher CL, Chory J and Nemhauser JL (2008). Integration of auxin and brassinosteroid pathways by Auxin Response Factor 2. *Proc. Natl. Acad. Sci. U. S. A.* 105: 9829-9834.
- Vidaurre DP, Ploense S, Krogan NT and Berleth T (2007). AMP1 and MP antagonistically regulate embryo and meristem development in *Arabidopsis*. *Development* 134: 2561-2567.
- Wang D, Pei K, Fu Y, Sun Z, et al. (2007). Genome-wide analysis of the auxin response factors (ARF) gene family in rice (*Oryza sativa*). *Gene* 394: 13-24.
- Wang JW, Wang LJ, Mao YB, Cai WJ, et al. (2005). Control of root cap formation by MicroRNA-targeted auxin response factors in *Arabidopsis*. *Plant Cell* 17: 2204-2216.
- Wu J, Wang F, Cheng L, Kong F, et al. (2011). Identification, isolation and expression analysis of auxin response factor (ARF) genes in *Solanum lycopersicum*. *Plant Cell Rep.* 30: 2059-2073.
- Xing H, Pudake RN, Guo G, Xing G, et al. (2011). Genome-wide identification and expression profiling of auxin response factor (ARF) gene family in maize. *BMC Genomics* 12: 178.