Research Article

Whole-Genome Duplication and Purifying Selection Contributes to the Functional Redundancy of Auxin Response Factor (*ARF*) Genes in Foxtail Millet (*Setaria italica* L.)

You Chen [iD],[1] **Bin Liu** [iD],[2] **Yujun Zhao** [iD],[1] **Wenzhe Yu** [iD],[1] and **Weina Si** [iD][1]

[1]*National Engineering Laboratory of Crop Stress Resistance Breeding, School of Life Sciences, Anhui Agricultural University, Hefei 230036, China*
[2]*Precision Medicine Research Laboratory of Anhui Province, Hefei, Anhui 230088, China*

Correspondence should be addressed to Weina Si; weinasi@ahau.edu.cn

Auxin response factors (ARFs) play crucial roles in auxin-mediated response, whereas molecular genetics of *ARF* genes was seldom investigated in *Setaria italica*, an important crop and C$_4$ model plant. In the present study, genome-wide evolutionary analysis of *ARFs* was performed in *S. italica*. Twenty-four *SiARF* genes were identified and unevenly distributed on eight of the nine chromosomes in *S. italica*. Duplication mode exploration implied that 13 SiARF proteins were originated from whole-genome duplication and suffered purifying selection. Phylogeny reconstruction of SiARFs by maximum likelihood and neighbor-joining trees revealed SiARFs could be divided into four clades. SiARFs clustered within the same clade shared similar gene structure and protein domain composition, implying functional redundancy. Moreover, amino acid composition of the middle regions was conserved in SiARFs belonged to the same clade. SiARFs were categorized into either activators or repressors according to the enrichment of specific amino acids. Intrinsic disorder was featured in the middle regions of ARF activators. Finally, expression profiles of *SiARFs* under hormone and abiotic stress treatment not only revealed their potential function in stress response but also indicate their functional redundancy. Overall, our results provide insights into evolutionary aspects of *SiARFs* and benefit for further functional characterization.

1. Introduction

The phytohormone auxin plays remarkable roles in regulating diverse aspects of plant growth and development, through auxin-responsive signaling cascades and gene expressions [1]. Moreover, auxin/indole acetic acid (Aux/IAA) and auxin response factor (ARF) proteins are key factors to regulate the expression of auxin-response genes [2]. In the absence of auxin, Aux/IAA interacts with ARFs, inhibiting the transcriptional regulation of auxin-responsive genes. Accumulation of auxin could release and activate ARFs. Activity-elevated ARFs would further transcriptionally regulate the expression of auxin-related response genes, such as *small auxin up RNA* (*SAUR*) and *Gretchen Hagen 3* (*GH3*), by combining the TGTCTC auxin response elements (AuxREs) in promoters [3].

Typical ARF proteins are distinguished by an N-terminal DNA-binding domain (DBD), C-terminal Aux/IAA domains that are committed in homo- and heterointeraction, and a variable middle region (MR) [4]. The DBD generally consists of plant-specific B-type domain, which could bind specifically to AuxREs and the auxin_resp domain with unknown function [5]. The MR region determinates whether ARFs are transcriptional activators or repressors [6]. Considering its important roles in auxin regulation system in plant, *ARFs* have drawn dramatically attentions. Researchers have found that *osarf12* mutant plants exhibited premature senescence of leaves and resection of floral organs [7]. Similarly, rice plants with *OsARF12* silenced showed leaf curl, short stature, and reduced viability compared with wild type [8].

Increasing researches have proposed that *ARFs* play roles not only in auxin perception and regulation, but also in the crosstalk of auxin with other phytohormones, such as abscisic acid (ABA), irreplaceable in plant responses to environmental stress [9]. Consistently, many reports suggested that *ARFs* were also in responsive to various abiotic stress, such as drought, salt, and cold [10–13]. *AtARF2* in *Arabidopsis thaliana* was found to be responsive to low potassium stress by phosphorylation modification [14]. *TasiRNA-ARF* was reported to be involved in maintaining the normal morphogenesis of flowers under stress conditions by fine-tuning changes in the expression of floral development and auxin response-related genes in *A. thaliana* [15]. These results elucidate that *ARF* genes played important roles not only in regulating plant growth and development, but also in response to abiotic stress.

Fast development of sequencing technology has provided unprecedented chances and data basis for evolutionary and functional investigation. *ARFs* have been identified and reported in several plants, like *A. thaliana*, *Oryza sativa*, and *Zea mays* [16–19]. However, relevant reports about *ARFs* were seldom found in *Setaria italica*, which is an ideal C_4 model for genetics and molecular biology research [20]. In the present study, the family members of *ARFs* were explored in *S. italica* genome widely, followed by chromosomal location analysis. Furthermore, duplication modes involved in *SiARF* members and selection pressure underlying the origination of duplicated *SiARF* gene pairs were also investigated. In addition, phylogeny reconstruction of SiARFs was performed in two phylogenetic trees. Domain composition, amino acid composition of MR regions, gene structure, and tissue-specific expression patterns of *SiARFs* were carefully compared among members in different clades. Finally, hormone treatment, as well as abiotic stress including salinity and PEG, was applied to conduct expression profile analysis of *SiARFs*. Above all, our study serves as the first genome-wide evolutionary analysis of *SiARF* genes and underlies the basis of further analysis.

2. Materials and Methods

2.1. Identification of ARF Genes in S. italica Genome. By using the ARF HMM profile (ARF_resp, PF06507) [21], the *ARF* sequences of *O. sativa* and *A. thaliana* were used as a query to identify the homologous protein sequences in the foxtail millet protein database (http://www.phytozome.net/BlastP) [22]. The Pfam database (http://Pfam.sanger.ac.uk/) and SMART [23] were used to further confirm each predicted *ARF* gene.

2.2. Identifying Protein Characteristics and Mapping SiARF Genes on Chromosome. The amino acid (AA) composition, molecular weight (kDa), and theoretical pI (PI) were analyzed by ExPASy (http://web.expasy.org/protparam/). 24 SiARF proteins were accurately distinguished between 8 chromosomes by the MapChart software [24].

2.3. Gene Duplication Modes and Collinearity Estimation. MCScanX package was employed to characterize collinearity within *S. italica* genome and gene duplication modes *SiARFs* involved in [25]. MCScanX could implement whole-genome

BlastP to identify collinear blocks, referred to collinearity relationship within or between species (*E* value was set up to 1*e*-10). Additionally, MCScanX was applied to trace the evolutionary history of *ARF* gene family expansion and classify tandem and segmental/WGD duplicated *ARF* genes according to their copy number and chromosomal distribution.

2.4. Estimation of Synonymous and Nonsynonymous Substitution Rates. SiARF sequences were searched by BlastP against protein sequence of *A. thaliana*, *Z. mays*, and *O. sativa* (http://gramene.org/www.phytozome.net) to account the orthologous homology among *S. italica* and other gramineous plants. The multiple alignment of paralogous and orthologous gene pairs was carried out by Clustalx and submitted to PAL2NAL to calculate Ks and/or Ka value by the codeml program in PAML [26].

2.5. Sequence Alignment and Phylogenetic Analysis of SiARFs. Multiple sequence alignment was performed using the full-length amino acids of ARFs from *S. italica*, *Z. mays*, *O. sativa*, and *A. thaliana* by MAFFT [27]. TrimAl v1.2 was further employed to remove poorly aligned regions with the parameter of –automated1 [28]. Then, the trimmed alignments were submitted to PhyML to identify the best-fit amino acid substitution model. The best-fit amino acid substitution model was JTT+G (–lnL = 21729.13) [29]. Subsequently, maximum likelihood (ML) phylogenetic tree was constructed in PhyML according to the estimated best-fit mode, and fast approximate likelihood-based measures of branch supports (Shimodaira-Hasegawa approximate likelihood ratio test, SH-aLRT) were applied for branch. To build the neighbor-joining (NJ) phylogenetic tree, the full-length amino acids of SiARFs were aligned by MAFFT and submitted to MEGA 7.0 [30]. The significance of each node bootstrap analysis was carried out using 1000 replicates [31].

2.6. Expression Profiling of SiARF Genes Using Transcriptome Data. The European Nucleotide Archive provided all Illumina *S. italica* RNA-HiSeq reads for four tissues, namely, spica, stem, leaf, and root [SRX128226 (spica), SRX128225 (stem), SRX128224 (leaf), and SRX128223 (root)] [32]. The RNA-seq data removed low quality reads that were screened by NGS toolkit as standard (http://59.163.192.90:8080/ngsqctoolkit/) [33]. The RPKM (reads per kilobase per million) method was used for normalizing the number of reads mapped. According to RPKM values, the heat map demonstrated the specific gene expression in different tissues.

2.7. Plant Materials, Growth Conditions, and Stress Treatments. Seeds of *S. italica* were grown in an artificial climate chamber at $28 \pm 1°C$ day/$23 \pm 1°C$ night temperature with $70 \pm 5\%$ relative humidity and natural sunlight during 21 days (March to April, 2017). For drought stress treatment, seedlings were exposed to 20% polyethylene glycol (PEG 6000) or 100 mM ABA that was initiated on the 21th day of normal growth condition. The different samples were collected after 0, 1, 3, 6, 12, and 24 h treatments, respectively. The 20% PEG (dehydration), salinity, ABA, and indole-3-acetic acid (IAA) treatment were performed that counterfeit in previous studies [34–38]. The samples were frozen in liquid nitrogen and stored at

-80°C until RNA isolation. The sample experiments were repeated three times to ensure stability and precision.

2.8. RNA Extraction and Expression Profiling Analysis. Total RNA was extracted by using the RNAPlus reagent (Takara, Japan) method, followed by standard manufacturer's instructions [39]. The first cDNA strand was generated using the M-MuLV reverse transcriptase experimental kit (Takara Bio Inc., USA). All gene-specific primers were designed using the Primer Express 3.0 software (Applied Biosystems, USA) with default parameters. The product size ranges from 150 to 250 bp for each *SiARF* gene (Table S2). The PCR reaction condition was performed at 95°C for 10 min, followed by 40 cycles at 95°C for 15 s and 60°C for 1 min. The analysis of melting curve (60 to 95°C after 40 cycles) was done to reduce the experimental error in accordance with the principle of biological repetition [40]. *Actin 2* (PF00022) was used as an internal control. The three replicates were carried out. The statistics was analyzed using the DPS software [41].

2.9. Intrinsic Disorder Prediction in the SiARF Proteins. Full-length SiARF amino acid sequences were submitted in DisProt by the PONDR-FIT algorithm [42]. Disordered values were presented as a heat map in R using the gplot package with heat map.2 function.

3. Results

3.1. Identification, Chromosomal Distribution, and Homology Model Analysis of ARF Family in S. italica. In the present study, the ARF HMM profile was used to screen the protein database of *S. italica* by the BlastP program. Obtained proteins were further submitted to Pfam website and confirm the existence of ARF and B3 domain. As a result, a total of 24 ARF proteins were found (Table 1). These genes were found to be unevenly distributed on 8 chromosomes, except for chromosome 2 (Figure 1). Chromosome 3 contains seven *SiARFs* (29.1%), and chromosomes 1, 4, and 5 have four *SiARFs*, respectively. Two *SiARFs* were presented on chromosome 7, but only one *SiARF* was found on chromosomes 6, 8, and 9, respectively. Lastly, these *ARF* genes were named according to their chromosomal location from *SiARF01* to *SiARF24* (Table 1).

3.2. Duplication and Divergence Rate of the SiARF Genes. Gene duplication, generally related to whole genome/segmental duplication (WGD/SD) and tandem duplication (TD), played important roles in gene family expansion and evolution. Employing blast and MCScanX programs, WGD and TD gene pairs were successfully characterized. Intriguingly, no TD events were found in *SiARFs*, and 13 *SiARFs* were involved in WGD events, including ten WGD gene pairs (Figure 2). To estimate whether selective pressure exits during the evolution and expansion of *SiARFs*, the ratio of nonsynonymous (Ka)/synonymous (Ks) substitution of these WGD pairs was calculated. Generally speaking, the value of Ka/Ks ratio implies selection pressure: Ka/Ks < 1 indicates purifying selection, Ka/Ks = 1 stands for neutral selection, while Ka/Ks > 1 represents positive selection [43]. The maximum value of Ka/Ks was 0.4095, and the minimum value of Ka/Ks was 0.0092 with an average of 0.2129 in *SiARFs* (Table S1). In addition, to roughly deduce the origin of these WGD pairs, the distribution of Ks ratio of paralog in *S. italica* and the ortholog of *S. italica* with *Z. mays* and *O. sativa* was also compared. According our results, the frequency of Ks of paralogs in *S. italica* peaked between 0.8 and 0.9, whereas Ks frequency of ortholog genes between *S. italica* and *Z. mays* or *O. sativa* peaked between 0.6 and 0.8, 0.3, and 0.4, respectively. These results suggest that these WGD paralogs originated before the split of maize and *S. italica* or *S. italica* and *O. sativa* (Figure 3), and their origination time was closer to the split time of *S. italica* and *O. sativa*.

3.3. Phylogenetic and Domain Analysis of SiARFs. To further investigate the evolutionary fate of *SiARF* genes, we constructed an unrooted phylogenetic tree using protein sequences of ARFs from *S. italica*, and those from *O. sativa*, *Z. mays*, and *A. thaliana* were used as outgroups (Figure 4). Totally, 114 ARF members were applied, and a ML phylogenetic tree was successfully constructed. According to the topological structure and bootstrap values of the nodes, the phylogenetic tree could be divided into four clades, namely, clade I (37 members), clade II (33 members), clade III (15 members), and clade IV (29 members). Clade I harbored the most ARF members, while clade III had the least ARF members. Additionally, ARF proteins in *S. italica* were clustered evenly within the three clades, including 9 members in clade I (SiARF1, SiARF2, SiARF6, SiARF7, SiARF11, SiARF12, SiARF13, and SiARF15), 6 in clade II (SiARF3, SiARF10, SiARF16, SiARF19, SiARF21, and SiARF23), 4 SiARF proteins in clade III (SiARF8, SiARF9, SiARF17, and SiARF18), and 5 in clade IV (SiARF4, SiARF5, SiARF14, SiARF22, and SiARF24), respectively (Figure 4). Furthermore, no recent duplication events were detected within SiARF members, while several were organized in maize ARF proteins (Figure 4).

We further examined the protein domain composition of ARFs in these four clades (Figure 4). ARF members, which had closer evolutionary relationship and clustered within the same clade, exhibit conserved protein domain composition. Auxin_resp (PF06507), B3 (PF02362), and Aux/IAA domains are the most conserved domains in surveyed ARF proteins. Almost all the surveyed ARF proteins harbored the conserved auxin_resp (PF06507) and B3 (PF02362) domains. Two ARFs, AtARF23 and ZmARF21, only had the conserved domain auxin_resp, whereas distribution of Aux/IAA domain differed among genes in different clades. ARFs in clades I and II harbored one or two Aux/IAA domains, whereas almost all ARFs in clades III and IV harbored no Aux/IAA domain.

To better confirm the accuracy of the phylogeny reconstruction, an NJ tree was built by only SiARF proteins (Figure 5(a)). ARFs in the NJ tree showed the same topological relationships with the MJ tree. Our further analysis mainly focuses on SiARFs employing the NJ tree.

3.4. Amino Acid Composition of MR Regions and Gene Structure of SiARFs. The protein sequences of MRs between auxin_resp and Aux/IAA are reported to be high variable and associated with the transcriptional activity of *ARF* genes.

Table 1: The characteristics of identified *auxin response factor* gene family in *Setaria italica*.

Gene name[a]	Locus	Chr location	AA[b]	PI[c]	MW (kDa)[d]
SiARF1	Seita.1G077200	1 : 6988237…6994473	907	5.45	99.6988
SiARF2	Seita.1G090900	1 : 8014277…8022232	1133	5.9	125.4214
SiARF3	Seita.1G195500	1 : 27687425…27693427	673	5.76	74.9183
SiARF4	Seita.1G241500	1 : 31927733…31930605	696	6.86	74.5453
SiARF5	Seita.3G003300	3 : 165345…168218	502	5.75	54.4367
SiARF6	Seita.3G020000	3 : 1162915…1171149	835	6.3	92.7573
SiARF7	Seita.3G028100	3 : 1711500…1718231	937	5.79	103.1312
SiARF8	Seita.3G147000	3 : 10559752…10564250	577	9.14	64.2633
SiARF9	Seita.3G179300	3 : 13481507…13487361	579	8.33	64.254
SiARF10	Seita.3G321500	3 : 40156717…40163068	841	6.25	92.7379
SiARF11	Seita.3G394000	3 : 49527955…49534308	897	5.66	98.8577
SiARF12	Seita.4G040100	4 : 2680871…2687694	1054	6.09	116.3922
SiARF13	Seita.4G240600	4 : 36454278…36462936	1084	6.12	120.7337
SiARF14	Seita.4G257800	4 : 37696236…37700051	686	7.05	74.697
SiARF15	Seita.4G262300	4 : 38057919…38064263	931	5.95	102.7815
SiARF16	Seita.5G004300	5 : 322788…327036	687	5.59	76.8788
SiARF17	Seita.5G265200	5 : 32737893…32742338	667	6.42	73.7489
SiARF18	Seita.5G321300	5 : 37047017…37052916	685	6.84	74.7693
SiARF19	Seita.5G438100	5 : 45465923…45471577	809	6.01	90.7756
SiARF20	Seita.6G212000	6 : 32960180…32967624	1099	6.13	121.6518
SiARF21	Seita.7G108100	7 : 20883066…20888476	663	5.59	73.4238
SiARF22	Seita.7G169600	7 : 25114768…25117902	677	8.05	72.674
SiARF23	Seita.8G135700	8 : 26264640…26270855	811	6.6	90.0333
SiARF24	Seita.9G219800	9 : 16301628…16306053	684	6.71	74.705

[a]The name of a gene named according to its order on the chromosome. [b]Protein sequence length. [c]Isoelectric point. [d]Molecular weight.
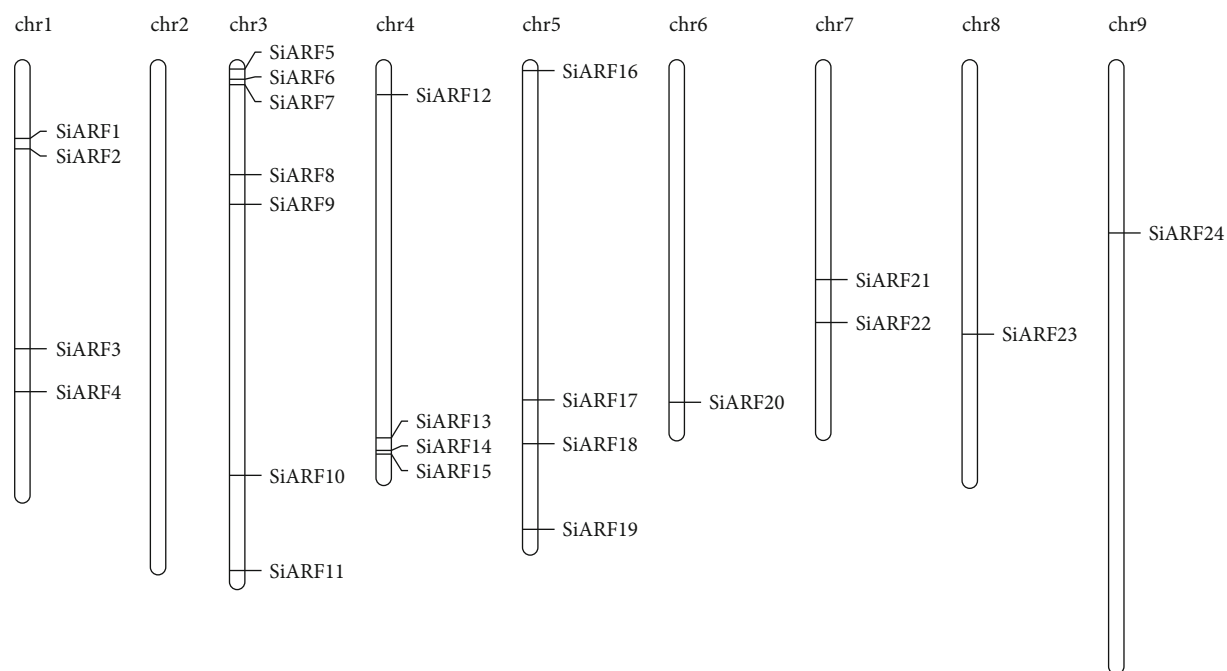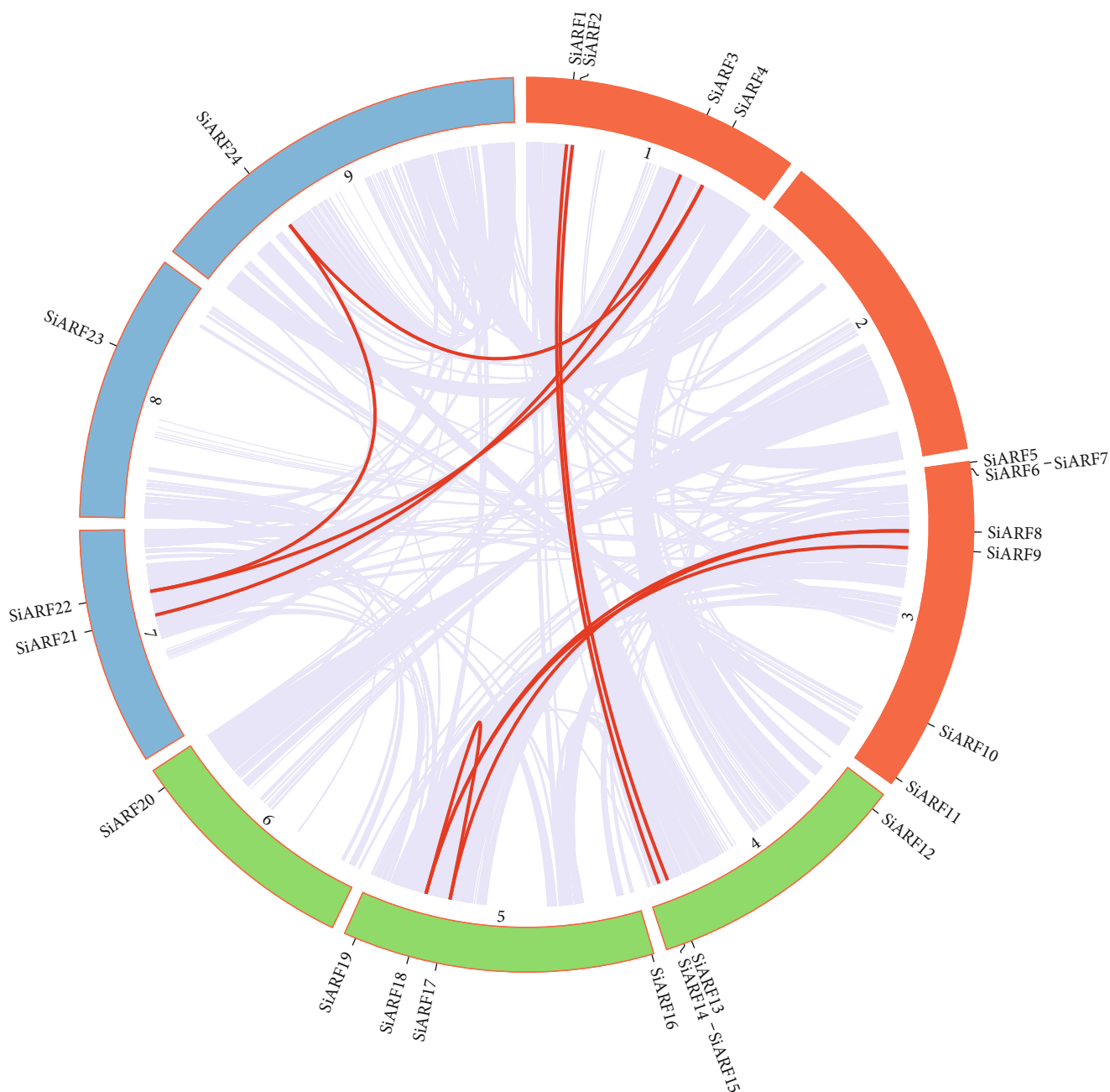

Figure 1: Chromosomal distribution of *SiARFs*.

FIGURE 2: Collinear analysis of *SiARF* genes. Genome-wide collinear genes in *Setaria italica* were marked with grey lines. *SiARF* whole-genome duplication gene pairs were emphasized by red lines.

We subsequently investigated the amino acid composition of MR regions in SiARFs (Figure 5(b)). As expected, SiARFs clustered within the same clades exhibited similar amino acid composition in MR regions. SiARFs in clade I are abundant in S, Q, and L, while SiARFs in clade II are abundant in S and P. In addition, S and G are enriched in the MR regions of SiARFs from clades III and IV.

We further analyzed the gene structure of *SiARF* genes (Figure 5(c)). As expected, the exon-intron structure varies among genes in different clades, while it is relative conserved among genes within in the same clades. Genes in clades I and II had more than 10 exons. Almost all genes in clade I har-

bored 14 exons, except for *SiARF20* and *SiARF7*, both of which had 13 exons. Five genes, including *SiARF21*, *SiARF3*, *SiARF19*, and *SiARF10* had 14 exons, while the other two genes, *SiARF16* and *SiARF23* have 12 exons. Finally, the five genes in clade IV had only three or two exons.

3.5. Tissue-Specific Expression Analysis of SiARFs. Tissue-specific expression profiles of *SiARFs* were estimated by RPKM values, generated from European Nucleotide Archive. The expression pattern of all *SiARF* genes was identified to simulate precise properties through the four tissues: root, stem, leaf, and spica. Results demonstrated the multifarious and
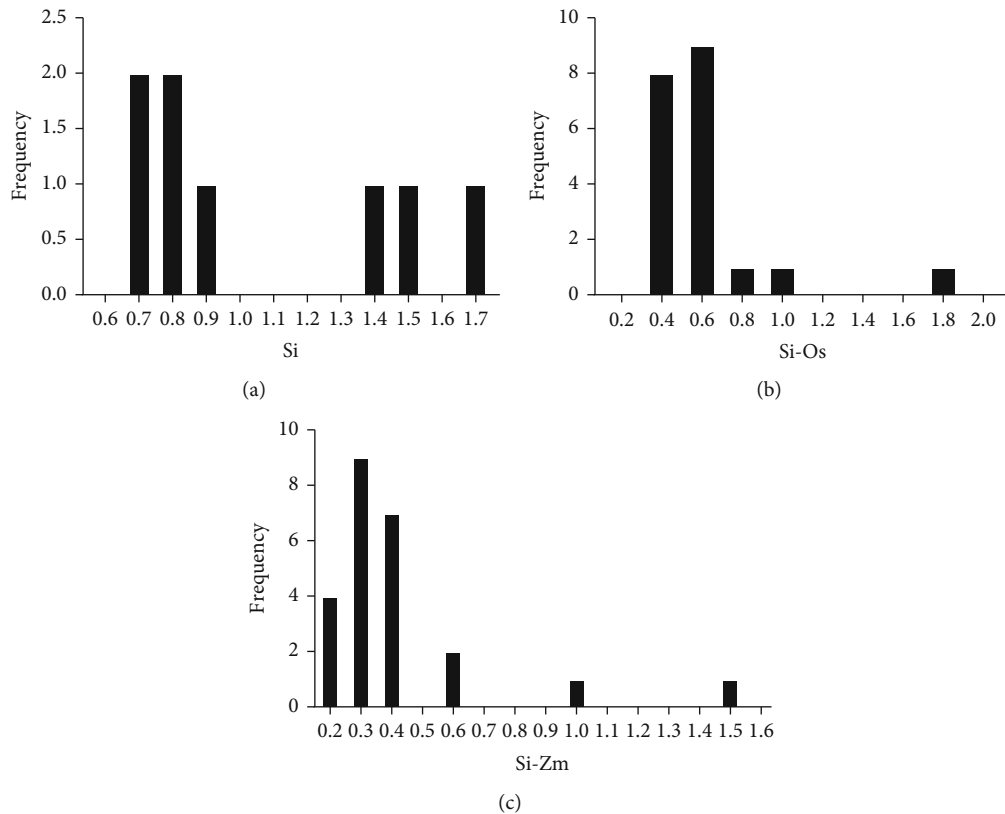
(a)

(b)

(c)

FIGURE 3: Ks distribution of paralog pairs of *SiARFs*. (a) Orthologous pairs between *S. italica* and *Oryza sativa* and between *S. italica* and *Zea mays*, respectively. The *x*-axis represents the Ks values of paralogous or orthologous gene pairs and was divided into different parts in unit of (a) 0.1 or (c) 0.2. (b) The *y*-axis means the frequency of Ks value relative to each unit. Si represents ARF paralogous in *S. italica*; Si-Os represents ARF orthologous between *S. italica* and *O. sativa*; Si represents ARF orthologous between *S. italica* and *Z. mays*.

dynamics expression profiles of all *SiARFs* in the four surveyed tissues (Figure 6). These 24 *SiARF* genes expressed in at least one tissue (PRKM > 1). *SiARF03*, *SiARF04*, *SiARF19*, *SiARF21*, and *SiARF23* expressed highly in all the four tissues (PRKM > 5). Furthermore, genes in the same clades showed dynamics expression patterns. Genes in clade I had distinct expression levels, while almost all genes in clade II are universally and highly expressed in four surveyed tissues. Remarkably, *SiARF23* exhibited high expression than other *SiARFs* in leaf. In addition, we also identified the expression between gene duplicated pairs. Six of the ten duplicate gene pairs (*SiARF1/15*, *SiARF2/13*, *SiARF22/24*, *SiARF8/17*, *SiARF9/18*, and *SiARF17/18*) resulted from WGD exhibited significant distinct expression profiles in the four surveyed tissues (two-way ANOVA test, $p < 0.05$), while four duplicate gene pairs (*SiARF3/21*, *SiARF4/22*, *SiARF4/24*, and *SiARF8/18*) shared similar expression profiles (two-way ANOVA test, $p > 0.05$), indicating functional redundancy.

3.6. Expression Patterns of SiARFs under Hormone and Abiotic Stress Treatments. It is generally known that the regulation pathway of auxin and the expression level of *ARF* genes have an important correlation. The induction treatment of exogenous auxin was beneficial to the expression of *ARF* family gene in plant. Previous experimental results reveal that *SiARF* members have complex biological effects

on the secretion of auxin. In the present study, under the treatment of 1 μm IAA, the real-time quantitative PCR analysis showed that huge variations exist in all *SiARF* genes (Figure 7(a)). For example, in IAA treatment condition, the expression of almost half of all *SiARF* genes (*SiARF03*, *SiARF05*, *SiARF07*, *SiARF08*, *SiARF09*, *SiARF11*, *SiARF14*, *SiARF17*, *SiARF18*, *SiARF20*, *SiARF20*, and *SiARF24*) were subdued continually. Nevertheless, 8 *ARF* genes (*SiARF04*, *SiARF06*, *SiARF10*, *SiARF12*, *SiARF19*, *SiARF21*, *SiARF22*, and *SiARF23*) were upregulated by exogenous IAA treatment. The remaining *ARF* gene expression levels show irregular expression profiles.

To further investigate whether the *SiARFs* were involved in ABA signaling pathway, seedlings were used to be disposed ABA. The results showed that almost all *SiARFs* were constantly upregulated by ABA stimuli until at 12 h (Figure 7(b)), besides *SiARF04* expressed highest at 1 h. In addition to this, the expression level of *SiARF22* showed a downregulation tendency. Comparing the relative expression of the *SiARF* family in ABA stimuli, It is not hard to find that *SiARF11* was evidenced to superlatively expressed (>70-fold) than others. In summary, *SiARF* genes had largely response effect on plant hormones in theory.

To investigate the response mechanism of *SiARF* transcription family members, the real-time quantitative PCR was used to analyze the potential functions of *SiARFs* during
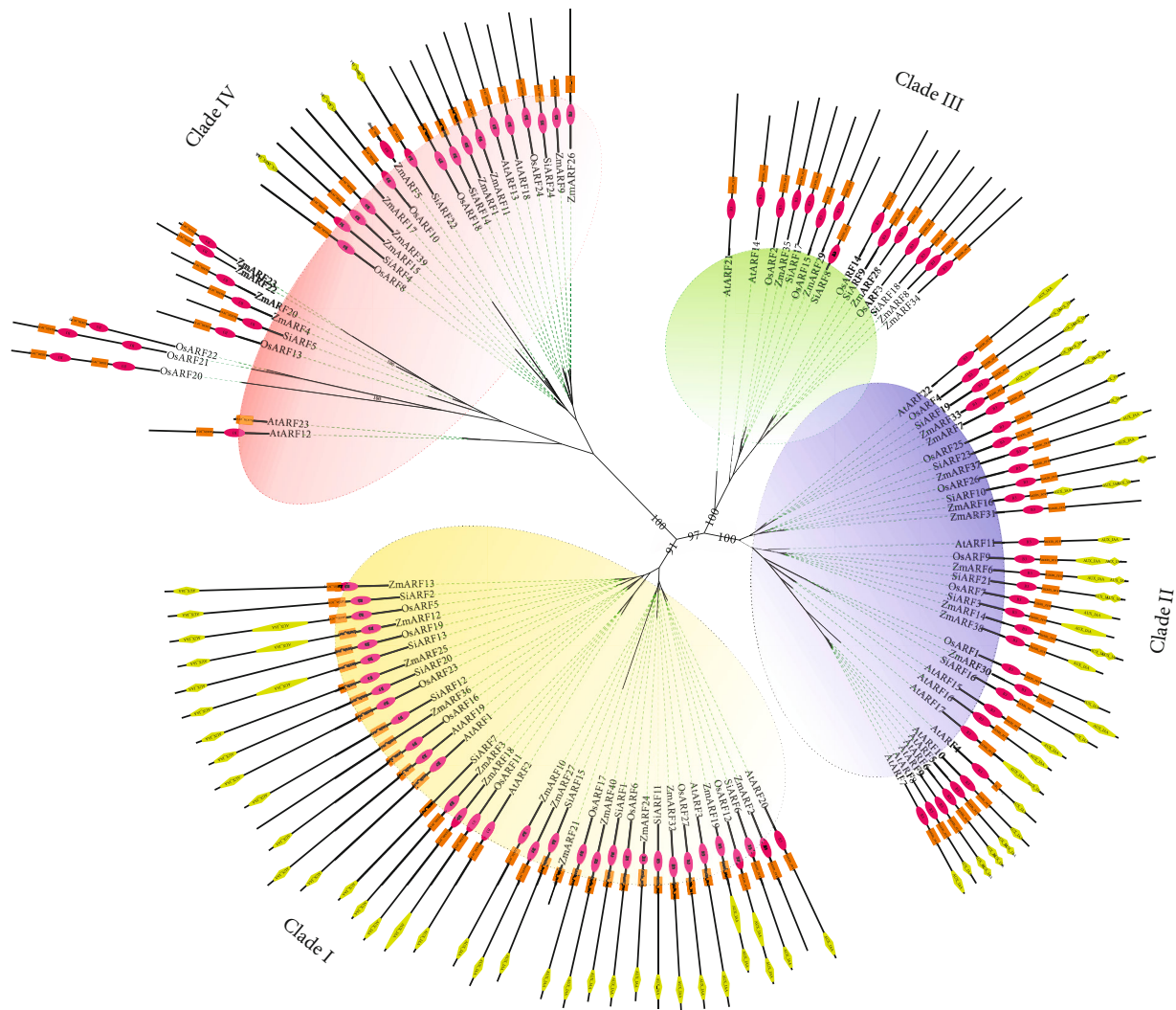
Figure 4: The ML phylogenetic tree constructed by ARF proteins from four plant species. Prefixes in the protein names, including Si, Os, Zm, and At, were used as to denote the species they belong to and represent *S. italica*, *O. sativa*, *Z. mays*, and *Arabidopsis thaliana*, respectively. The phylogenetic tree could be classified into four clades, and each clade was marked with different colors. Yellow, blue, green, and red represented clades I to IV, respectively. Protein domain composition of each ARF members was also exhibited. Red diamonds represented B3 domain. Orange squares represented auxin_resp domain, and yellow diamonds indicated Aux/IAA domain.

abiotic stress (salinity and dehydration). Firstly, during the different period (1 h, 3 h, 6 h, 12 h, and 24 h) duration of PEG treatment, the results were visualized by histogram. Interestingly, we found that all *SiARF* genes were upregulated at the bulk of time points (Figure 8(a)). The quantitative expression of all *SiARFs* found that *SiARF01*, *SiARF02*, and *SiARF14* peaked at 1 h, and *SiARF03*, *SiARF04*, and *SiARF15* were highest expressed after 6 h. The expression of *SiARF22* and *SiARF23* was increased until at 3 h, and *SiARF07* and *SiARF18* peaked at 24 h. In addition, the transcription level of *SiARF05*, *SiARF06*, *SiARF08*, *SiARF09*, *SiARF10*, *SiARF11*, *SiARF12*, *SiARF13*, *SiARF16*, *SiARF17*, *SiARF19*, *SiARF20*, *SiARF21*, and *SiARF24* reached the expression peak at 12 h; subsequently, the expression levels were reduced. Contrarily, the relative expression of *SiARF11* steadily increased more than 50-fold at 12 h.

Salinity stress changes the permeability of plant cells to affect the growth and development of plants by osmotic stress

and the ion toxicity. Hence, we also designed salinity treatment expression profiles to get insight into potential tolerance effects of foxtail millet *ARF* genes. The results found that all *SiARF* genes have incurred variations in different treatment phases (Figure 8(b)). It was observed that all *SiARF* genes appear transcriptional level increased in different time points. Strikingly, only *SiARF04* was accompanied by constant expression until 6 h reaches maximum value, and three genes (*SiARF21*, *SiARF22*, and *SiARF24*) showed highest expression at 3 h. However, all of other genes (*SiARF01*, *SiARF02*, *SiARF03*, *SiARF04*, *SiARF05*, *SiARF06*, *SiARF07*, *SiARF08*, *SiARF09*, *SiARF10*, *SiARF11*, *SiARF12*, *SiARF13*, *SiARF14*, *SiARF15*, *SiARF16*, *SiARF17*, *SiARF18*, *SiARF19*, *SiARF20*, and *SiARF23*) were measured that all were accidentally upregulated in treatment time and peaked at 12 h. Noteworthy, five genes (*SiARF05*, *SiARF11*, *SiARF17*, *SiARF22*, and *SiARF24*) appeared the idiocratic higher expression at the time point than other treatment duration. The *SiARF22*

(a)

Ala (A)
Cys (C)
Asp (D)
Glu (E)
Phe (F)
Gly (G)
His (H)
Ile (I)
Lys (K)
Leu (L)

Met (M)
Asn (N)
Pro (P)
Gln (Q)
Arg (R)
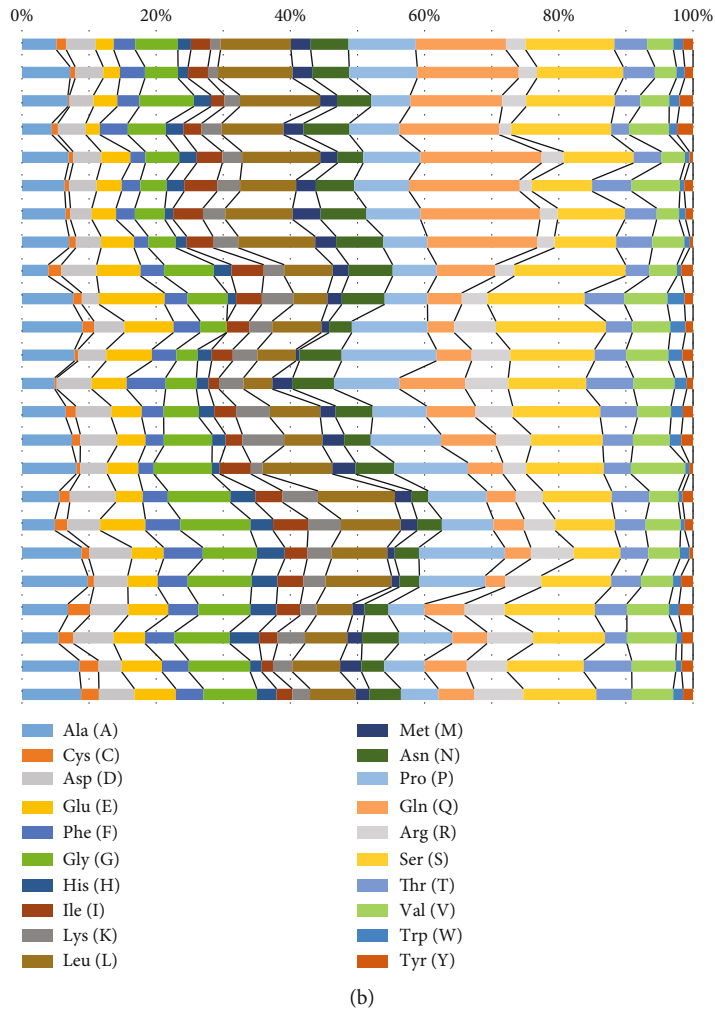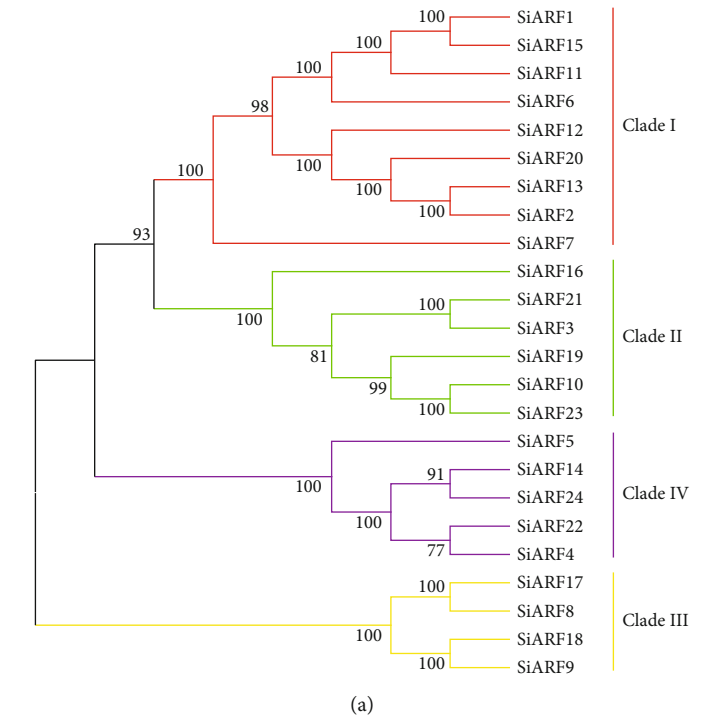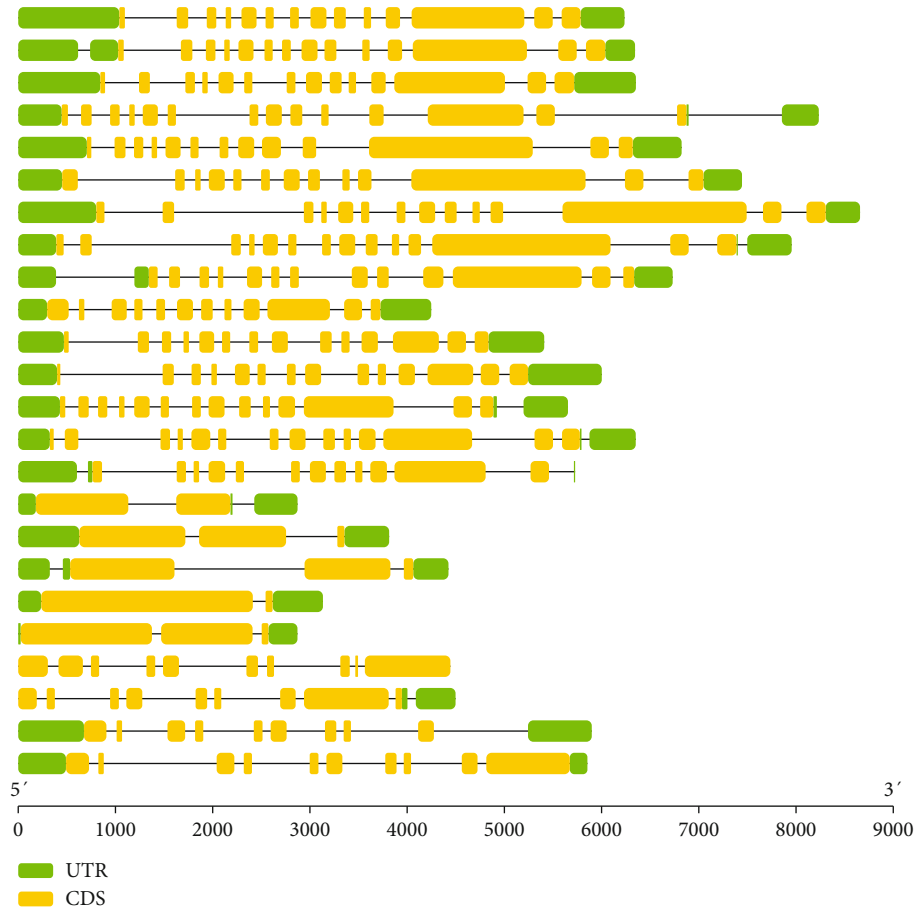Ser (S)
Thr (T)
Val (V)
Trp (W)
Tyr (Y)

(b)

FIGURE 5: Continued.

(c)

FIGURE 5: (a) Phylogeny reconstruction and (b) amino acid composition of (c) middle regions and gene structure of *SiARFs*.
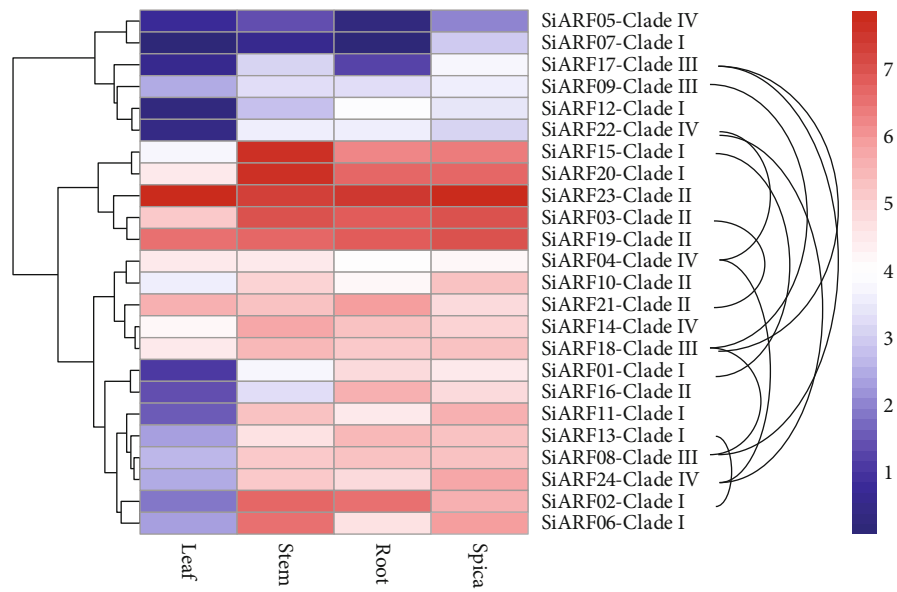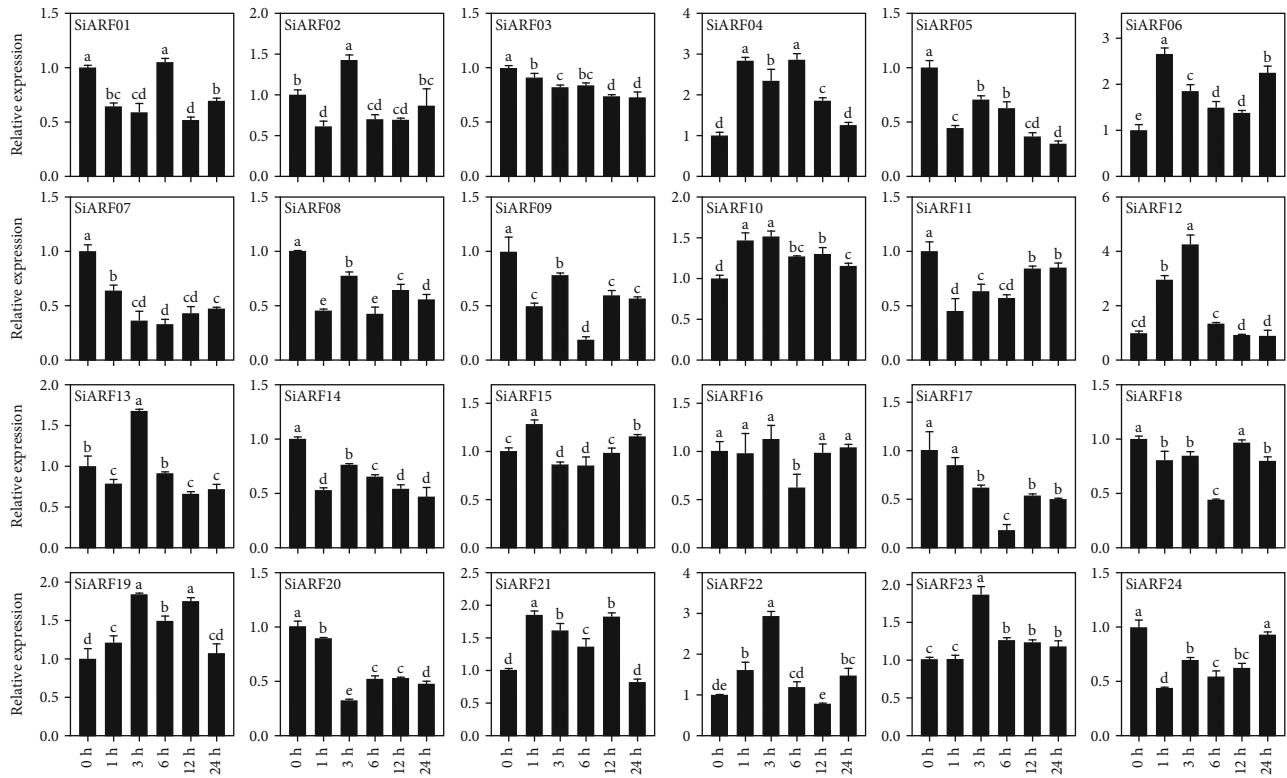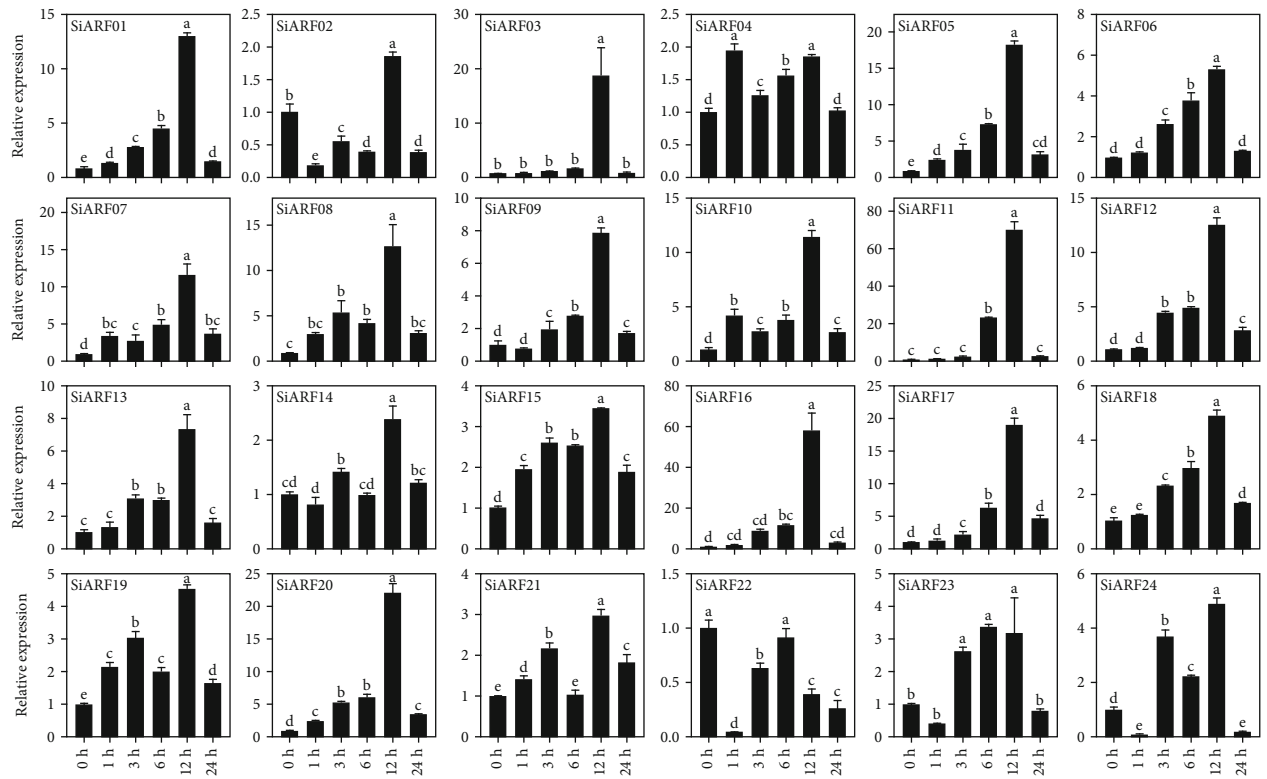

FIGURE 6: The heat map of the expression level of *SiARFs* in leaf, root, stem, and spica. The color from blue (1) to red (7) represents an increase in the level of gene expression.

FIGURE 7: The relative expression of 24 *SiARFs* analyzed using real-time quantitative PCR under (a) indole-3′-acetic acid (IAA) treatment and (b) abscisic acid (ABA) treatment after 0 h, 1 h, 3 h, 6 h, 12 h, and 24 h. All the experimental results were based on three biologic replicates. Error bars imply ±SD for three biological replicates. Different letters indicate significant differences ($p < 0.05$).
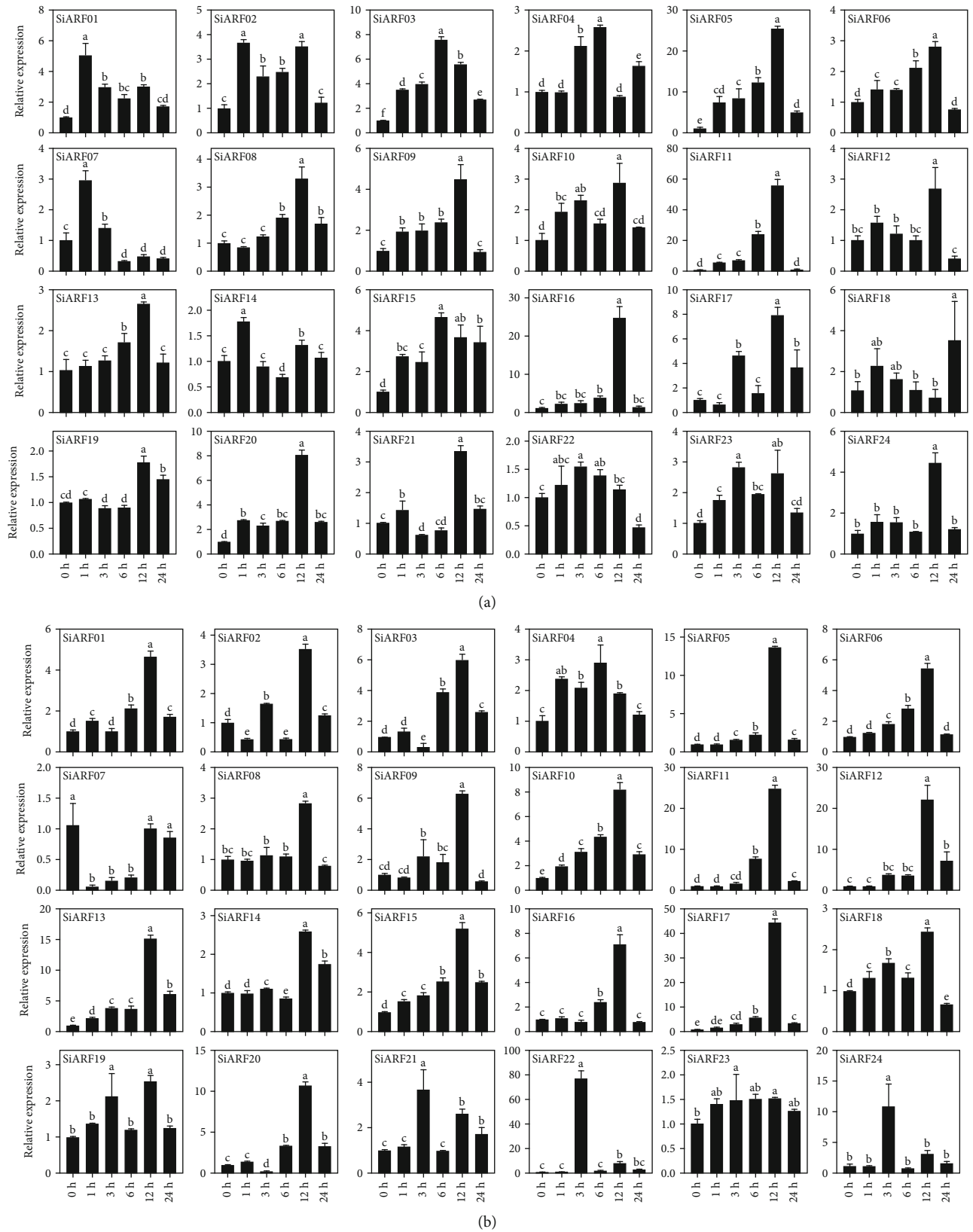
FIGURE 8: The relative expression of 24 *SiARFs* analyzed using real-time quantitative PCR under (a) PEG stress treatment and (b) salinity stress treatment after 0 h, 1 h, 3 h, 6 h, 12 h, and 24 h. All the experimental results were based on three biologic replicates. Error bars imply ±SD for three biological replicates. Different letters indicate significant differences (*p* < 0.05).

was even upregulated more than 70-fold at 3 h. Data results indicate that *SiARFs* play an important role in regulating the dehydration stress and salinity stress.

4. Discussion

ARF genes are key factors in the auxin-induced signaling cascades and gene expression, regulating all aspects of plant metabolism progress and participating in plant stress response [14]. *ARF* families in plant species have multiple gene members, which, to a certain extent, could explain why simple auxin-to-ARF pathway could manipulate such many aspects of physiological metabolism [2]. A broad exploration to the evolutionary patterns and phylogenic relationship between *ARFs* could benefit for deciphering the functional specialization in *ARFs*. Hence, in this study, a genome-wide survey of *ARF* genes in *S. italica* was presented.

In *S. italica*, twenty-four *SiARF* genes were identified, number of which was similar to that in *A. thaliana* (23) and *O. sativa* (25), but less than that in *Z. mays* (36) [16, 17, 19]. Gene duplication, mainly including WGD and TD, was generally believed to involve in gene expansion of plant multiple gene families. Moreover, genes generated by different duplication modes had different functional fates. WGD duplicates were supposed to contribute to functional redundancy, while TD duplicates could cope with rapidly changing environments and gain functional novelty [44]. In the present study, 13 out of 24 *SiARFs* were deduced to experience WGD events and 10 WGD gene pairs were identified (Figure 2). We also found evolutionary force underlying WGD gene pairs was purifying selection. Ka distribution implied the ancient origin of *SiARF* pairs, which were originated before the split of *S. italica* with *O. sativa* or with *Z. mays*. Considering broad impact of auxin-induced signaling cascades and the upstream roles of *ARFs*, these WGD pairs may contribute to the function redundancy or specification.

Phylogenetic analysis of *SiARFs* was also performed. Firstly, an ML tree was constructed, applying ARFs from other three genes as outgroups (Figure 4). In the ML tree, ARFs could apparently be classified into four clades. Moreover, genes within the same clades shared similar protein domain composition. We also constructed an NJ tree with only SiARFs (Figure 5). As expected, SiARFs in the NJ tree exhibited almost the same phylogenic relationship with those in the ML tree. The consistence between the NJ and ML trees confirmed the accuracy and repeatability of the phylogeny reconstruction. With restriction to SiARFs, proteins in clade II and clade I harbored integrated functional domain, including B3, Aux_resp, and Aux/IAA domains, while majority of the proteins in clade III and clade IV lacked Aux/IAA domain. Considering the homo- or heterointeraction of Aux/IAA domain, proteins in clades III and IV may function as monomers. In addition, the amino acid composition of MR regions was reported to determine whether an ARF protein is a transcriptional activator or repressor [3]. In previous advances, Q, S, and L are enriched in the MR regions of ARF activators. Other ARFs without Q-rich MR regions are almost probably transcriptional repressors. In our results, MR regions of SiARFs in clade I are enriched in Q, S, and L

(Figure 5(b)). Furthermore, those OsARFs in clade I had been experimentally confirmed to be transcriptional activators and interact with OsAux/IAA proteins [6]. These outcomes strongly imply that SiARFs in clade I are transcriptional activators and could interact with SiAux/IAA. SiARFs in clades II to IV are most likely to be transcriptional repressors (Figure S2). Moreover, previous advances predicted the presence of intrinsic disorder (ID) in the MR of ARF activators [2]. ID was characterized by lack of 3D structure and enriched with charged and polar amino acid residues. Our results proved that intrinsic disorder was conserved character in the clade I SiARFs (Figure S2), consistent with those in previous research [2]. ID contributes to protein-protein interaction, conditional DNA binding, and posttranslational modifications, through specific and rapid conformational changes [45]. As majority research were concerned on the conserved B3, Aux_resp, and Aux/IAA domains, the discovery of ID could provide new aspect of auxin output control through *ARF*-related pathway.

It is well known that *ARFs* play central in auxin-related pathway. Concentration of auxin could regulate the activity of *ARFs* [3]. In our results, *SiARFs* are responsive to exogenous IAA treatment, which may be owing to that exogenous IAA could influence the homeostasis of auxin (Figure 7(a)). Additionally, alteration of the concentration of auxin could also induce the ABA-related pathway [46]. Moreover, ABA is another important hormone and involves in response to diverse stress [47]. In previous researches, *ARFs* were believed to be promising candidates involving in environmental stress and hormone signaling. *IbARF5* from sweet potato could significantly improve the tolerance to salt and drought in transgenic *A. thaliana*, associated with ABA biosynthesis [48]. *ARF4* in poplar inhibited auxin signaling in lateral root (LR) formation under salt stress [49]. In the present study, we found that almost all *SiARFs* were upregulated in leaves by exogenous ABA. Subsequently with abiotic stress treatment (NaCl and PEG), majority of *SiARFs* were also found responsive, exhibiting either upregulated or downregulated expression tendency (Figures 7(b) and 8). These findings strongly imply that *SiARFs* are associated with abiotic stress, in an ABA-dependent pathway, showing the functional conversation or redundancy. Our results could provide some bases for further functional characterization, especially those related to stress and hormone signaling.

5. Conclusions

ARFs are key output control of auxin-related pathway. In the present study, evolutionary patterns of *ARFs* genes in *S. italica* were genome widely performed. 13 out of the 24 SiARF proteins were originated from whole-genome duplication, suffering purifying selection. Phylogeny reconstruction of *SiARFs* by ML and NJ trees revealed that *SiARFs* could be categorized into four clades. *SiARFs* clustered within the same clade shared similar gene structure and protein domain composition, implying functional redundancy. Moreover, amino acid composition of the middle regions (MR) was conserved in *SiARFs* belonged to the same clade. *SiARFs* in clade I are suggested to be transcriptional activators and may interact with SiIAA/Aux, while *SiARFs* in clades II to

IV are most likely to be transcriptional repressors. Intrinsic disorder was featured in the MR regions of *ARF* activators. In addition, various *SiARFs* are induced or repressed under hormone and abiotic stress treatments, which revealed their potential function in stress response and functional redundancy. Overall, the present study provides a "mugshot" of *SiARFs* and provides a theoretical basis for further functional and molecular mechanism analysis.

Data Availability

Our research data has been submitted as supplementary materials.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Authors' Contributions

You Chen and Bin Liu contributed equally to this work.

Acknowledgments

Supplementary Materials

See Figure S1-S2 and Table S1-S3 in the Supplementary Material 1. Figure S1: the ML phylogenetic tree constructed by ARF proteins from four plant species. Figure S2: intrinsic disorder and transcriptional activity prediction of SiARFs. Green indicated unstructured protein scores. Table S1: synonymous and nonsynonymous values of *SiARF* WGD duplicate pairs. Table S2: synonymous values of orthologous *ARF* gene pairs between *S. italica* and *O. sativa* and between *S. italica* and *Z. mays*. Table S3: primers used in the real-time quantitative PCR assay. See coding sequencing sequences of *ARFs* of four surveyed species in the Supplementary Material 2. See full-length amino acid sequences of ARFs of four surveyed species in the Supplementary Material 3. *(Supplementary Materials)*

References

[1] S.-B. Li, Z.-Z. Xie, C.-G. Hu, and J.-Z. Zhang, "A review of auxin response factors (ARFs) in plants," *Frontiers in Plant Science*, vol. 7, 2016.

[2] M. Roosjen, S. Paque, and D. Weijers, "Auxin response factors: output control in auxin biology," *Journal of Experimental Botany*, vol. 69, no. 2, pp. 179–188, 2018.

[3] S. B. Tiwari, G. Hagen, and T. Guilfoyle, "The roles of auxin response factor domains in auxin-responsive transcription," *Plant Cell*, vol. 15, no. 2, pp. 533–543, 2003.

[4] S. Song, L. Hao, P. Zhao et al., "Genome-wide Identification, Expression Profiling and Evolutionary Analysis of Auxin Response Factor Gene Family in Potato (*Solanum tuberosum* Group Phureja)," *Scientific Reports*, vol. 9, no. 1, p. 1755, 2019.

[5] D. R. Boer, A. Freire-Rios, W. A. M. van den Berg et al., "Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors," *Cell*, vol. 156, no. 3, pp. 577–589, 2014.

[6] C. Shen, S. Wang, Y. Bai et al., "Functional analysis of the structural domain of ARF proteins in rice (*Oryza sativa* L.)," *Journal of Experimental Botany*, vol. 61, no. 14, pp. 3971–3981, 2010.

[7] S. K. Wang, S. N. Zhang, C. D. Sun et al., "Auxin response factor (OsARF12), a novel regulator for phosphate homeostasis in rice (*Oryza sativa*)," *The New Phytologist*, vol. 201, no. 1, pp. 91–103, 2014.

[8] K. A. Attia, A. F. Abdelkhalik, M. H. Ammar et al., "Antisense phenotypes reveal a functional expression of OsARF1, an auxin response factor, in transgenic rice," *Current Issues in Molecular Biology*, vol. 11, Suppl 1, pp. i29–i34, 2009.

[9] R. R. Finkelstein, S. S. Gampala, and C. D. Rock, "Abscisic acid signaling in seeds and seedlings," *Plant Cell*, vol. 14, suppl 1, pp. S15–S45, 2002.

[10] M. Jain and J. P. Khurana, "Transcript profiling reveals diverse roles of auxin-responsive genes during reproductive development and abiotic stress in rice," *The FEBS Journal*, vol. 276, no. 11, pp. 3148–3162, 2009.

[11] S. Wang, Y. Bai, C. Shen et al., "Auxin-related gene families in abiotic stress response in *Sorghum bicolor*," *Functional & Integrative Genomics*, vol. 10, no. 4, pp. 533–546, 2010.

[12] C. V. Ha, D. T. le, R. Nishiyama et al., "The auxin response factor transcription factor family in soybean: genome-wide identification and expression analyses during development and water stress," *DNA Research*, vol. 20, no. 5, pp. 511–524, 2013.

[13] M. K. Sekhwal, A. K. Swami, V. Sharma, and R. Sarin, "Identification of drought-induced transcription factors in Sorghum bicolor using GO term semantic similarity," *Cellular & Molecular Biology Letters*, vol. 20, no. 1, pp. 1–23, 2015.

[14] S. Zhao, M. L. Zhang, T. L. Ma, and Y. Wang, "Phosphorylation of ARF2 relieves its repression of transcription of the K+ transporter gene HAK5 in response to low potassium stress," *Plant Cell*, vol. 28, no. 12, pp. 3005–3019, 2016.

[15] A. Matsui, K. Mizunashi, M. Tanaka et al., "tasiRNA-ARF pathway moderates floral architecture in Arabidopsis plants subjected to drought stress," *BioMed Research International*, vol. 2014, Article ID 303451, 10 pages, 2014.

[16] Y. Okushima, P. J. Overvoorde, K. Arima et al., "Functional genomic analysis of the *auxin response factor* gene family members in Arabidopsis thaliana: unique and overlapping functions of *ARF7* and *ARF19*," *Plant Cell*, vol. 17, no. 2, pp. 444–463, 2005.

[17] D. K. Wang, K. M. Pei, Y. P. Fu et al., "Genome-wide analysis of the *auxin response factors* (ARF) gene family in rice (*Oryza sativa*)," *Gene*, vol. 394, no. 1-2, pp. 13–24, 2007.

[18] H. Y. Jiang, "Genome-wide analysis and evolution of the auxin response factor (ARF) gene family in *Sorghum bicolor*," *Journal of Anhui Agricultural University*, vol. 37, no. 3, pp. 395–400, 2010.

[19] Y. Liu, H. Y. Jiang, W. J. Chen et al., "Genome-wide analysis of the auxin response factor (ARF) gene family in maize (*Zea mays*)," *Plant Growth Regulation*, vol. 63, no. 3, pp. 225–234, 2011.

[20] P. Li and T. P. Brutnell, "Setaria viridis and Setaria italica, model genetic systems for the Panicoid grasses," *Journal of Experimental Botany*, vol. 62, no. 9, pp. 3031–3037, 2011.

[21] A. Bateman, L. Coin, R. Durbin et al., "The Pfam protein families database," *Nucleic Acids Research*, vol. 32, no. 90001, pp. 138D–1141, 2004.

[22] D. M. Goodstein, S. Q. Shu, R. Howson et al., "Phytozome: a comparative platform for green plant genomics," *Nucleic Acids Research*, vol. 40, no. D1, pp. D1178–D1186, 2012.

[23] I. Letunic, S. Khedkar, and P. Bork, "SMART: recent updates, new developments and status in 2020," *Nucleic Acids Research*, vol. 49, no. D1, pp. D458–d460, 2021.

[24] Y. Cheng, X. Y. Li, H. Y. Jiang et al., "Systematic analysis and comparison of nucleotide-binding site disease resistance genes in maize," *The FEBS Journal*, vol. 279, no. 13, pp. 2431–2443, 2012.

[25] Y. P. Wang, H. B. Tang, J. D. DeBarry et al., "MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity," *Nucleic Acids Research*, vol. 40, no. 7, p. e49, 2012.

[26] M. Suyama, D. Torrents, and P. Bork, "PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments," *Nucleic Acids Research*, vol. 34, no. Web Server, pp. W609–W612, 2006.

[27] K. Katoh, J. Rozewicki, and K. D. Yamada, "MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization," *Briefings in Bioinformatics*, vol. 20, no. 4, pp. 1160–1166, 2019.

[28] S. Capella-Gutierrez, J. M. Silla-Martinez, and T. Gabaldon, "trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses," *Bioinformatics*, vol. 25, no. 15, pp. 1972-1973, 2009.

[29] S. Guindon, J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel, "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0," *Systematic Biology*, vol. 59, no. 3, pp. 307–321, 2010.

[30] S. Kumar, G. Stecher, and K. Tamura, "MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets," *Molecular Biology and Evolution*, vol. 33, no. 7, pp. 1870–1874, 2016.

[31] K. Tamura, J. Dudley, M. Nei, and S. Kumar, "MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0," *Molecular Biology and Evolution*, vol. 24, no. 8, pp. 1596–1599, 2007.

[32] G. Cochrane, B. Alako, C. Amid et al., "Facing growth in the European Nucleotide Archive," *Nucleic Acids Research*, vol. 41, no. D1, pp. D30–D35, 2013.

[33] R. K. Patel and M. Jain, "NGS QC toolkit: a toolkit for quality control of next generation sequencing data," *PLoS One*, vol. 7, no. 2, p. e30619, 2012.

[34] Y. Xiang, Y. M. Huang, and L. Z. Xiong, "Characterization of stress-responsive *CIPK* genes in rice for stress tolerance improvement," *Plant Physiology*, vol. 144, no. 3, pp. 1416–1428, 2007.

[35] G. Lu, C. X. Gao, X. N. Zheng, and B. Han, "Identification of *OsbZIP72* as a positive regulator of ABA response and drought tolerance in rice," *Planta*, vol. 229, no. 3, pp. 605–615, 2009.

[36] Y. Zhao, Y. Q. Zhou, H. Y. Jiang et al., "Systematic analysis of sequences and expression patterns of drought-responsive members of the *HD-Zip* gene family in maize," *PLoS One*, vol. 6, no. 12, article e28488, 2011.

[37] A. K. Mishra, M. Muthamilarasan, Y. Khan, S. K. Parida, and M. Prasad, "Genome-wide investigation and expression analyses of WD40 protein family in the model plant foxtail millet (*Setaria italica* L.)," *PLoS One*, vol. 9, no. 1, article e86852, 2014.

[38] S. B. Li, W. Z. OuYang, X. J. Hou, L. L. Xie, C. G. Hu, and J. Z. Zhang, "Genome-wide identification, isolation and expression analysis of *auxin response factor* (ARF) gene family in sweet orange (Citrus sinensis)," *Frontiers in Plant Science*, vol. 6, 2015.

[39] J. Logemann, J. Schell, and L. Willmitzer, "Improved method for the isolation of RNA from plant tissues," *Analytical Biochemistry*, vol. 163, no. 1, pp. 16–20, 1987.

[40] K. Kumar, M. Muthamilarasan, and M. Prasad, "Reference genes for quantitative real-time PCR analysis in the model plant foxtail millet (Setaria italica L.) subjected to abiotic stress conditions," *Plant Cell, Tissue and Organ Culture (PCTOC)*, vol. 115, no. 1, pp. 13–22, 2013.

[41] Q. Y. Tang and C. X. Zhang, "Data Processing System (DPS) software with experimental design, statistical analysis and data mining developed for use in entomological research," *Insect Science*, vol. 20, no. 2, pp. 254–260, 2013.

[42] B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, and V. N. Uversky, "PONDR-FIT: a meta-predictor of intrinsically disordered amino acids," *Biochimica et Biophysica Acta*, vol. 1804, no. 4, pp. 996–1010, 2010.

[43] Z. Yang, S. Gu, X. Wang, W. Li, Z. Tang, and C. Xu, "Molecular evolution of the *CPP-like* gene family in plants: insights from comparative genomics of Arabidopsis and rice," *Journal of Molecular Evolution*, vol. 67, no. 3, pp. 266–277, 2008.

[44] X. Qiao, Q. Li, H. Yin et al., "Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants," *Genome Biology*, vol. 20, no. 1, p. 38, 2019.

[45] J. G. Liu, N. B. Perumal, C. J. Oldfield, E. W. Su, V. N. Uversky, and A. K. Dunker, "Intrinsic disorder in transcription factors," *Biochemistry-Us*, vol. 45, no. 22, pp. 6873–6888, 2006.

[46] C. D. Rock and X. Sun, "Crosstalk between ABA and auxin signaling pathways in roots of *Arabidopsis thaliana* (L.) Heynh," *Planta*, vol. 222, no. 1, pp. 98–106, 2005.

[47] A. S. Raghavendra, V. K. Gonugunta, A. Christmann, and E. Grill, "ABA perception and signalling," *Trends in Plant Science*, vol. 15, no. 7, pp. 395–401, 2010.

[48] C. Kang, S. He, H. Zhai, R. Li, N. Zhao, and Q. Liu, "A sweet-potato *auxin response factor* gene (IbARF5) is involved in carotenoid biosynthesis and salt and drought tolerance in transgenic Arabidopsis," *Frontiers in Plant Science*, vol. 9, p. 1307, 2018.

[49] F. He, C. Z. Xu, X. K. Fu et al., "TheMicroRNA390/TRANS-ACTING short interfering RNA3Module mediates lateral root growth under salt stress via the auxin pathway," *Plant Physiology*, vol. 177, no. 2, pp. 775–791, 2018.