

RESEARCH

Open Access



Genome-wide identification and expression analysis of the Auxin-Response factor (ARF) gene family in *Medicago sativa* under abiotic stress

Fenqi Chen¹, Jinqing Zhang¹, Xue Ha¹ and Hailing Ma^{1*}

Abstract

Background Alfalfa (*Medicago sativa*) is the most widely planted legume forage and one of the most economically valuable crops in the world. The periodic changes in its growth and development and abiotic stress determine its yield and economic benefits. Auxin controls many aspects of alfalfa growth by regulating gene expression, including organ differentiation and stress response. Auxin response factors (ARF) are transcription factors that play an essential role in auxin signal transduction and regulate the expression of auxin-responsive genes. However, the function of ARF transcription factors is unclear in autotetraploid-cultivated alfalfa.

Result A total of 81 *ARF* were identified in the alfalfa genome in this study. Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were analyzed, identifying that *ARF* genes are mainly involved in transcriptional regulation and plant hormone signal transduction pathways. Phylogenetic analysis showed that *MsARF* was divided into four clades: I, II, III, and IV, each containing 52, 13, 7, and 9 genes, respectively. The promoter region of the *MsARF* gene contained stress-related elements, such as ABRE, TC-rich repeats, MBS, LTR. Proteins encoded by 50 *ARF* genes were localized in the nucleus without guide peptides, signal peptides, or transmembrane structures, indicating that most *MsARF* genes are not secreted or transported but only function in the nucleus. Protein structure analysis revealed that the secondary and tertiary structures of the 81 *MsARF* genes varied. Chromosomal localization analysis showed 81 *MsARF* genes were unevenly distributed on 25 chromosomes, with the highest distribution on chromosome 5. Furthermore, 14 segmental duplications and two sets of tandem repeats were identified. Expression analysis indicated that the *MsARF* was differentially expressed in different tissues and under various abiotic stressors. The quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis showed that the expression profiles of 23 *MsARF* genes were specific to abiotic stresses such as drought, salt, high temperature, and low temperature, as well as tissue-specific and closely related to the duration of stress.

Conclusion This study identified *MsARF* in the cultivated alfalfa genome based on the autotetraploid level, which GO, KEGG analysis, phylogenetic analysis, sequence characteristics, and expression pattern analysis further confirmed. Together, these findings provide clues for further investigation of *MsARF* functional verification and molecular breeding of alfalfa. This study provides a novel approach to systematically identify and characterize ARF transcription factors in autotetraploid cultivated alfalfa, revealing 23 *MsARF* genes significantly involved in response to various stresses.

*Correspondence:

Hailing Ma

mahl@gsau.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords ARF genes, Cultivated alfalfa, Auxin, Abiotic stresses, Expression patterns

Introduction

Auxin is an important plant endogenous hormone that regulates various plant life processes and plays a key role, including tissue differentiation, organogenesis, apical dominance, root formation, tropism, and response to biotic and abiotic stresses [1, 2]. Numerous plants, including *Arabidopsis thaliana*, have a large number of functional genes that are auxin-regulated and play a significant role in growth and development [3–5]. Among these genes, auxin response factor (ARF) family members play a vital role in auxin signaling and regulating the expression of auxin-responsive genes [6, 7]. Most ARFs contain three conserved functional domains: the DNA binding domain (DBD) at the N terminus, the variable intermediate region (MR) acting as the activating domain (AD) or inhibitory domain (RD), and the C-terminal dimerization domain (CTD) [8]. DBD belongs to the b3-like family, which enables ARF to specifically bind to TGTCTC auxin response element (AuxRE) in various auxin response gene promoters, thus transcriptional regulation the expression of these genes widely involved in plant growth and development [9]. Depending on its amino acid composition, MR has fewer sequences and relies on ARF as a transcription activator or repressor. The AD is rich in glutamine, serine, and leucine residues, while the RD is rich in serine, proline, leucine, and glycine residues [8–10]. The CTD is responsible for protein-protein interactions, such as ARF homodimerization or ARF heterodimerization [5]. ARF transcription factors and auxin/indole-3-acetic acid (Aux/IAA) protein dimers can inhibit ARF-regulated transcriptional activity in the presence of low auxin concentrations, but the degree of expression inhibition is only moderate [11]. When the auxin concentrations are high, the 26 S proteasome releases the interacting ARF protein from the inhibitory dimer structure [12, 13].

Furthermore, recent studies have reported relevant information involving interacting proteins and target genes of ARF, elucidating the specific mechanisms by which ARF regulates plant developmental processes. *Arabidopsis* has 22 *AtARF* functional genes and one pseudogene [8, 14]. The expression of ARF is precisely and dynamically regulated in different developmental stages and tissues of *Arabidopsis*. For example, *AtARF2-4*, *AtARF3*, *AtARF5*, and *AtARF8* are crucial for plant floral organ differentiation and development [15–17]. *AtARF7* and *AtARF19* are essential for auxin-mediated plant development by regulating unique and partially overlapping target gene sets [18]. In addition

to regulating plant growth and development, the ARF gene family is also involved in plant responses to numerous abiotic stresses.

For example, overexpression of *AtARF3* stimulates the expression of drought-stress-responsive genes in *Arabidopsis*. However, under drought stress, these ARF-activated genes maintain very low expression levels in the *arf3* mutant [17]. Similarly, the expression levels of most *SlARF* were downregulated under salt stress, while some *SlARF* genes, such as *SlARF1*, *SlARF4*, and *SlARF19*, were significantly upregulated when undergoing salt stress [6]. In lettuce, the expression levels of most *LsARFs* are closely correlated with temperature changes, and *LsARF8a* is involved in regulating the timing of bolting [19]. In earlier studies, authors also found that the accumulation of *LsARF2* and *LsARF5* in flower stems continuously increased with the progress of lettuce bolting and discovered that the tryptophan metabolism pathway was involved in the early bolting of lettuce under high temperatures [20, 21]. In addition, Hu et al. [22] exhibited that *ARF2/26/27/43* actively participates in banana resistance (*Musa acuminata* L.) against cold stress damage. These previous studies have determined that the ARF gene family is essential for regulating plant growth, hormone responses, and responses to abiotic stresses, including drought, salt, and temperature changes, and is a crucial gene family for understanding plant biology.

Alfalfa (*Medicago sativa* L.), known as the “king of forage,” has a high biomass and crude protein content and is rich in digestible nutrients and mineral elements, which dramatically reduces the cost of feed supplements for livestock production [23, 24]. Alfalfa is widely cultivated in North America, Asia, and other continents and is one of the most economically valuable crops in the world [24, 25]. Alfalfa is the fourth most cultivated crop in the United States, after wheat, corn, and soybeans [26]. In China, alfalfa cultivation is distributed through 14 provinces in the northern region of the country, with a locally cultivated alfalfa variety (“Xinjiang Daye”) that has large leaves and an autotetraploid genome ($2n=4x=32$), which is widely cultivated due to its strong resistance [23].

The complete genome data of the autotetraploid variety Xinjiang Daye was released, resulting in a chromosomal-level genome assembly containing 32 high-quality chromosomes instead of the previously assembled 8 chromosomes [26, 27]. The updated assembly of this genome provides an important

foundation to facilitate scientific research on this economically important forage crop and improve the stress resistance of alfalfa through genetic engineering. To date, the function of the ARF gene family involved in plant development and abiotic stress has been widely reported in many plants, such as papaya (*Carica papaya*) [9], longan (*Dimocarpus longan* L.) [28], and apple (*Malus domestica*) [29]. Nonetheless, before the report of chen et al. [26], the genome that identified *MsARF* of the cultivated alfalfa is not available. Genomic data and many transcriptome analyses of cultivated alfalfa have been published, providing reliable experimental resources for systematic research on the *MsARF* [30–32]. Based on these data, our study used bioinformatics methods to identify *ARF* in autotetraploid Xinjiang Daye alfalfa at the genome-wide level. We analyzed the phylogeny, gene structure, motif composition, *cis*-regulatory elements, chromosome maps, tissue-specific expression patterns, and differential expression under various abiotic stresses. The results of this study lay the foundation for further investigation of the abiotic stress response and the creation of new germplasm of alfalfa with strong resistance through genetic engineering technology.

Results

Identification and functional analysis of *MsARF* family genes

After removing the redundant sequences, a total of 81 *MsARF* protein sequences were identified in the alfalfa genome and renamed from *MsARF001* to *MsARF081* according to the order of their occurrence in the genome (Table S1). The physicochemical properties of the *MsARF* proteins, including the gene name, gene ID, amino acid number, molecular weight, isoelectric point, and protein hydrophobicity coefficient, were investigated using ProtParam on the ExPasy website (Table S2). Significant differences exist in the number of amino acids, molecular weight, and isoelectric points of the 81 *MsARF* genes. *MsARF025* contains 1127 amino acids, making it the gene with the highest number of coding amino acids in the family, with a molecular weight of 126144.77 daltons (Da). *MsARF020*, with the family's lowest number of amino acids, encodes 116 amino acids with a molecular weight of 13091.05 Da. The theoretical isoelectric points of these *MsARF* proteins ranged from 4.32 (*MsARF054*) to 9.58 (*MsARF068*). The instability coefficient of the *MsARF* family ranged from 22.84 (*MsARF032*) to 98.98 (*MsARF039*), but the instability index of most genes (66 out of 81) was higher than 40, indicating that they were unstable proteins. The aliphatic amino acid index of 81 *MsARF* proteins ranged from 48.43 (*MsARF067*) to 93.76 (*MsARF001*), all less than 100, indicating that they were

hydrophilic proteins. Further subdivided with the help of protein hydrophobicity coefficients, 22 *MsARF* proteins with hydrophobicity coefficients less than -0.5 were categorized as hydrophilic, and 59 *MsARF* proteins were between -0.5 and 0.5 , belonging to the amphotropic protein group.

The GO and KEGG annotation analysis was performed using the Alfalfa Database to determine the functional classification of all *MsARF*. A total of 59 annotated *MsARF* were classified as biological processes, molecular functions, or cellular components (Table S3). Terms like “regulation of nuclear acid template translation,” “regulation of RNA biological process,” and “regulation of translation, DNA template” served as the main annotations for biological processes. In the cellular components, the *MsARF* genes were mainly enriched in terms such as “nucleus,” “internal membrane-bound organelle,” and “membrane-bound organelle”. KEGG enrichment analysis found that 20 *MsARF* genes were enriched in the “plant hormone signal transduction pathway,” and only *MsARF039* was enriched in “RNA polymerase,” indicating that *ARF* genes mainly function through the plant hormone signaling pathway (Table S4).

Phylogenetic analysis of the *MsARF* gene family

To study the phylogenetic relationship between *MsARF* proteins, 25 *ARF* proteins from rice, 23 from *Arabidopsis*, and 81 from alfalfa identified in this study were selected to further study the phylogenetic relationships of the *MsARF* gene family in alfalfa. The *ARF* gene family was divided into four major clades, of which clade I contained three *Arabidopsis* genes, six rice genes, and 52 alfalfa genes. Clade II contained 5 *Arabidopsis* genes, 9 rice genes, and 13 alfalfa genes. Clade III contains 2 *Arabidopsis* genes, 4 rice genes, and 7 alfalfa genes. Clade IV contains 13 *Arabidopsis* genes, 6 rice genes, and 9 alfalfa genes (Fig. 1).

Subcellular localization and structural analysis of *MsARF* protein

Subcellular localization prediction of the *MsARF* was performed, with 22 *MsARF* in the chloroplast, 9 *MsARF* in the cytoplasm, and the remaining 50 *MsARF* in the nucleus (Table S5). The results of the peptide, signal peptide, and transmembrane structure analysis showed that all *MsARF* had no peptides, signal peptides, or transmembrane structures, indicating that the 50 *MsARF* genes located in the nucleus acted as transcription factors, which are not secreted or transported but only function in the nucleus. Analysis of the secondary structures in the *MsARF* protein found that the largest proportion of the *MsARF* protein was a random coil, with the composition percentage of random coil > α -helix > extended

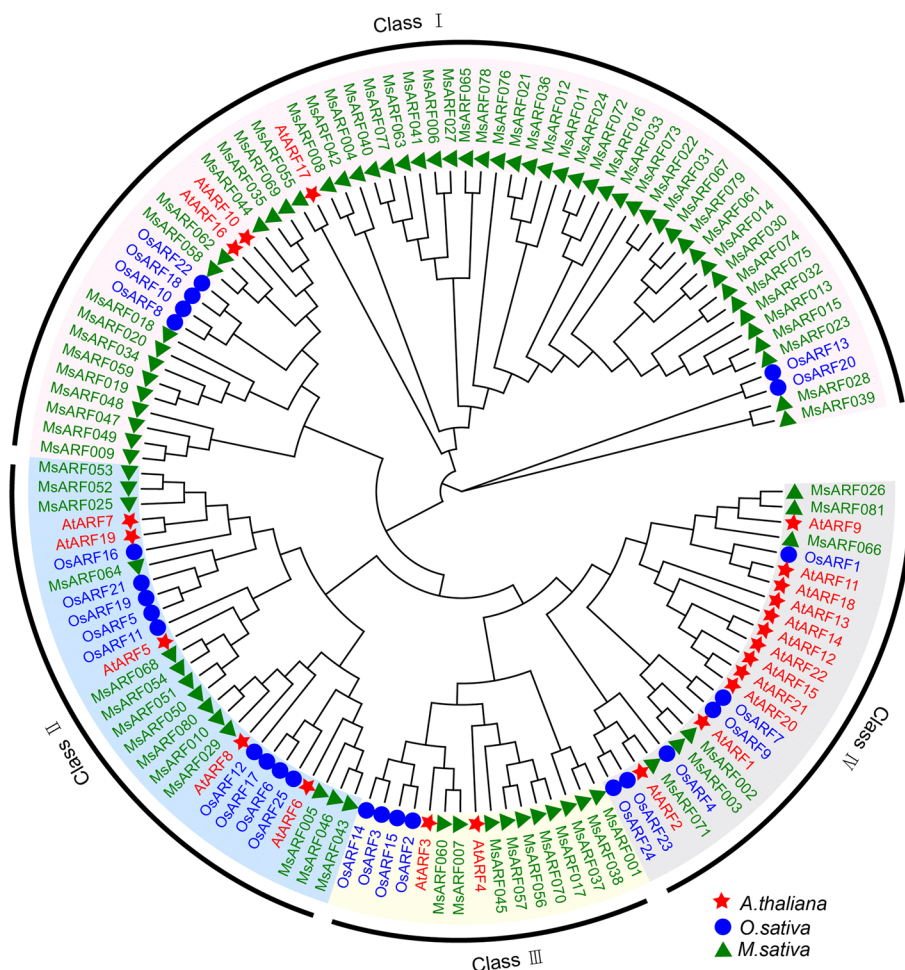


Fig. 1 Phylogenetic analysis of ARF protein in *A. thaliana*, *O. sativa*, and *M. sativa*

chain > β -turn angle (Table S5). Simultaneously, the tertiary structure of the MsARF protein was constructed (Fig. 2). Due to the different percentages of secondary structure composition, the tertiary structure also showed significant differences. The specific types of spatial structures in MsARF proteins were relatively few, and most types (46) were obtained through horizontal flipping and rotation, as shown in Fig. 2A C, as well as 2B and 2D, 2E, 2G, and 2I. Another 24 MsARF proteins shared a common spatial framework, as shown in Fig. 2F. In addition, the remaining 11 MsARF proteins all had their own unique spatial framework.

Conservative motif analysis of the MsARF gene

The MsARF domain is the core of ARF transcription factors, which can activate downstream genes by interacting with their promoters. MEME tools were used to explore the distribution and structural diversity of conserved motifs in the MsARF protein and to identify conserved motifs, most of which play important roles in

protein-protein interactions and transcriptional activity. A total of 10 conserved motifs were identified and renamed as motifs 1 to 10, with some of these motifs showing similar composition and position in the same subfamily of the MsARF (Fig. 3A, Fig. S1). Motif 8 was highly conserved in the MsARF gene family and was identified in all 80 MsARF except MsARF039, indicating that motif 8 is one of the most important motifs in the MsARF gene family (Fig. 3). Furthermore, motif 7 was highly conserved in the MsARF subfamilies II, III, and IV.

Analysis of the conserved domains of the MsARF protein sequences revealed that all MsARF sequences contained an auxin response domain (Auxin_resp). Among the 81 MsARF sequences, except for 10 MsARFs (MsARF013/018/028/032/039/049/051/054/072/079), all other sequences contain a typical DBD domain (B3-like). However, phylogenetic analysis revealed that the Aux/IAA binding domain is only distributed in the clade II, III, and IV subfamilies ($n=19$) (Figs. 1 and 3B). Furthermore, MsARF079 also contains a cyanobacterial

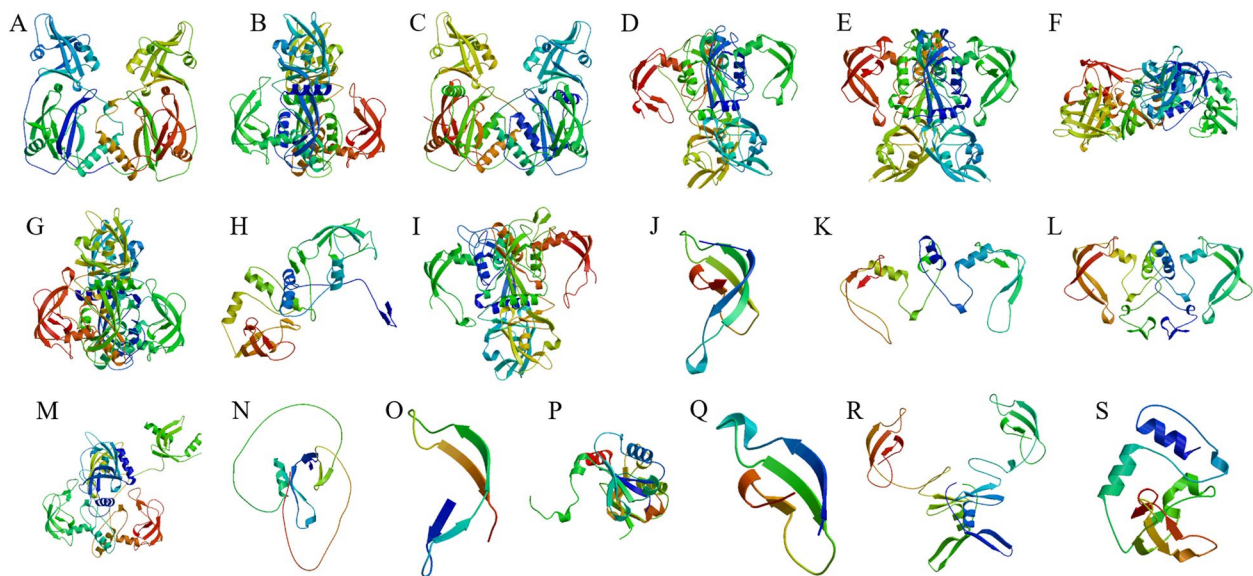


Fig. 2 Analysis of MsARF protein tertiary structures. **A** MsARF001; **B** MsARF002, 026, 034, 035, 045, 081; **C** MsARF003, 038, 066, 071; **D** MsARF004, 008, 009, 012, 019, 021, 025, 031, 043, 047, 059, 065, 078; **E** MsARF005, 010, 029, 046, 050, 052, 053, 064, 080; **F** MsARFS006, 011, 014, 015, 022, 023, 027, 032, 033, 036, 037, 040, 056, 057, 060, 061, 063, 067, 070, 074, 075, 076, 077, 079; **G** MsARF007, 017, 055; **H** MsARF013; **I** MsARF016, 024, 041, 042, 044, 048, 058, 062, 069, 073; **J** MsARF018; **K** MsARF020; **L** MsARF028; **M** MsARF030; **N** MsARF039; **O** MsARF049; **P** MsARF051; **Q** MsARF054; **R** MsARF068; **S** MsARF072

aminoacyl-tRNA synthetases appended domain (CAAD). To continue to determine the structural differences among *MsARFs* and reveal gene function, regulation, and evolution, a gene structure analysis of the *MsARF* family was performed. Gene structure analysis showed that the number of CDS in the *MsARF* gene ranged from 1 to 15 (Fig. 3C), and no UTR regions were identified. Among the 81 *MsARF* genes, *MsARF046* has the most CDS ($n=15$), followed by *MsARF005*, *MsARF026*, *MsARF029*, *MsARF043*, *MsARF050*, *MsARF064*, and *MsARF066* that have 14 CDS, while *MsARF020* and *MsARF030* have only one CDS. Phylogenetic analysis revealed that the *MsARF* genes differ in the number of CDS in different subfamilies.

Chromosomal localization and homology analysis of *MsARF* gene in alfalfa

This study identified and mapped the chromosomal positions of all *MsARF* genes to investigate whether there was a collateral homologous gene pair relationship between the *MsARFs* in alfalfa (Fig. 4). The 81 *MsARF* genes were unevenly distributed on 25 chromosomes (Chr) and seven other chromosomes (Chr 3.4, Chr 6.1, Chr 6.2, Chr 6.3, Chr 6.4, Chr 7.4, and Chr 8.3) (Fig. 4). The *MsARF* was distributed the most on Chromosome 5, with Chr 5.1 and Chr 5.2, each containing nine members, and Chr 5.3 and Chr 5.4, each containing seven members. Furthermore, five *MsARFs* were found on Chr 1.2, Chr 2.1,

and Chr 2.4, each with four members on Chr 2.3; three *MsARFs* each on Chr 1.4, Chr 1.1, and Chr 8.2; and two *MsARFs* on Chr 1.1, Chr 2.2, Chr 4.1, and Chr 8.1 each. There was only one *MsARF* on the remaining eight chromosomes.

We used TBtools software to conduct a collinearity analysis to detect gene duplication events in the *ARF* in alfalfa [33]. 14 pairs of segment repeats and two sets of tandem repeats in the *MsARF* were identified (*MsARF056/MsMASS057* and *MsARF074/MsMASS075*) (Fig. 5). The nonsynonymous substitution rate/synonymous substitution rate (Ka/Ks) ratio is usually an important indicator of selection pressure in evolution [34]. The analysis of homologous and homologous *MsARF* gene pairs showed that among the 14 pairs of segment repeats and two sets of tandem repeats of *MsARF* genes, the Ka/Ks ratio of two pairs of segment repeats (*MsARF008/MsMASS042* and *MsARF008/MsMASS042*) was greater than one, indicating a positive selection effect. The Ka/Ks ratios of the remaining 14 pairs of *MsARF* genes were all less than one, indicating they underwent strong purification selection after replication (Table S6).

Analysis of *Cis*-regulatory elements in the promoter region of the *MsARF* gene family

Cis-regulatory elements are specific deoxyribonucleic acid (DNA) sequences located upstream of gene coding sequences and regulate the expression of stress response



Fig. 3 Analysis of conserved motifs, domains, and gene structure of *MsARF* proteins. **A** Gene conservative motif analysis; **B** Protein conservative domain analysis; **C** Gene structure analysis

genes by binding to transcription factors. Therefore, we explored the distribution of seven *cis*-regulatory elements in the promoter regions of these altered *MsARF* genes during abiotic stress. We explored the distribution of *cis*-regulatory elements related to hormones and abiotic stress in the *MsARF* gene promoter region (Fig. 6). *cis*-regulatory elements related to hormones include abscisic acid response: ABRE ($n=118$), Methyl Jasmonate (MeJA) response: CGTCA-motif ($n=91$), auxin response: TGA-element ($n=38$) and AuxRR-core ($n=8$). *cis*-regulatory elements related to abiotic stress include defense and stress response: TC-rich repeats ($n=49$), drought induction: MBS ($n=66$), and low-temperature response: LTR ($n=27$). In addition, we found that there was only a *cis*-regulatory element in *MsARF012*, *MsARF021*, *MsARF022*, *MsARF025*, *MsARF031*, *MsARF036*, *MsARF042*, *MsARF045* and *MsARF051*.

Expression pattern analysis of the *MsARF* gene family

Tissue-specific expression is related to the specific function of *ARF* genes in specific tissues. Therefore, 81 *MsARF* genes were compared and identified using

'blastn' on the Alfalfa Database, and expression patterns of 45 *MsARFs* were obtained using gene expression profile data (Table S7). TBtools software was used to generate *MsARF* genes (gene expression data was 0 when mapping). The *MsARF* was expressed in eight specific tissues, including flowers, leaves, roots, post-elongating stems, nodules, elongating stems, young leaves, mature leaves, and senescent leaves of alfalfa. Group a genes were highly expressed in flowers, leaves, roots, pre-elongated stems, nodules, and elongated stems. Group b genes were highly expressed in flowers and leaves. Group c genes were highly expressed in flowers, and group d genes were highly expressed in young, mature, and senescent leaves, which may play an important role in leaf development. The two genes in group e, *MsARF025* and *MsARF049*, were highly expressed in senescent leaves and considered putative key regulatory genes for this specific tissue type (Fig. 7A). To identify the expression patterns of *MsARF* family genes in response to abiotic stress, we analyzed the transcriptional expression profiles of *MsARF* family genes under salt stress, mannitol

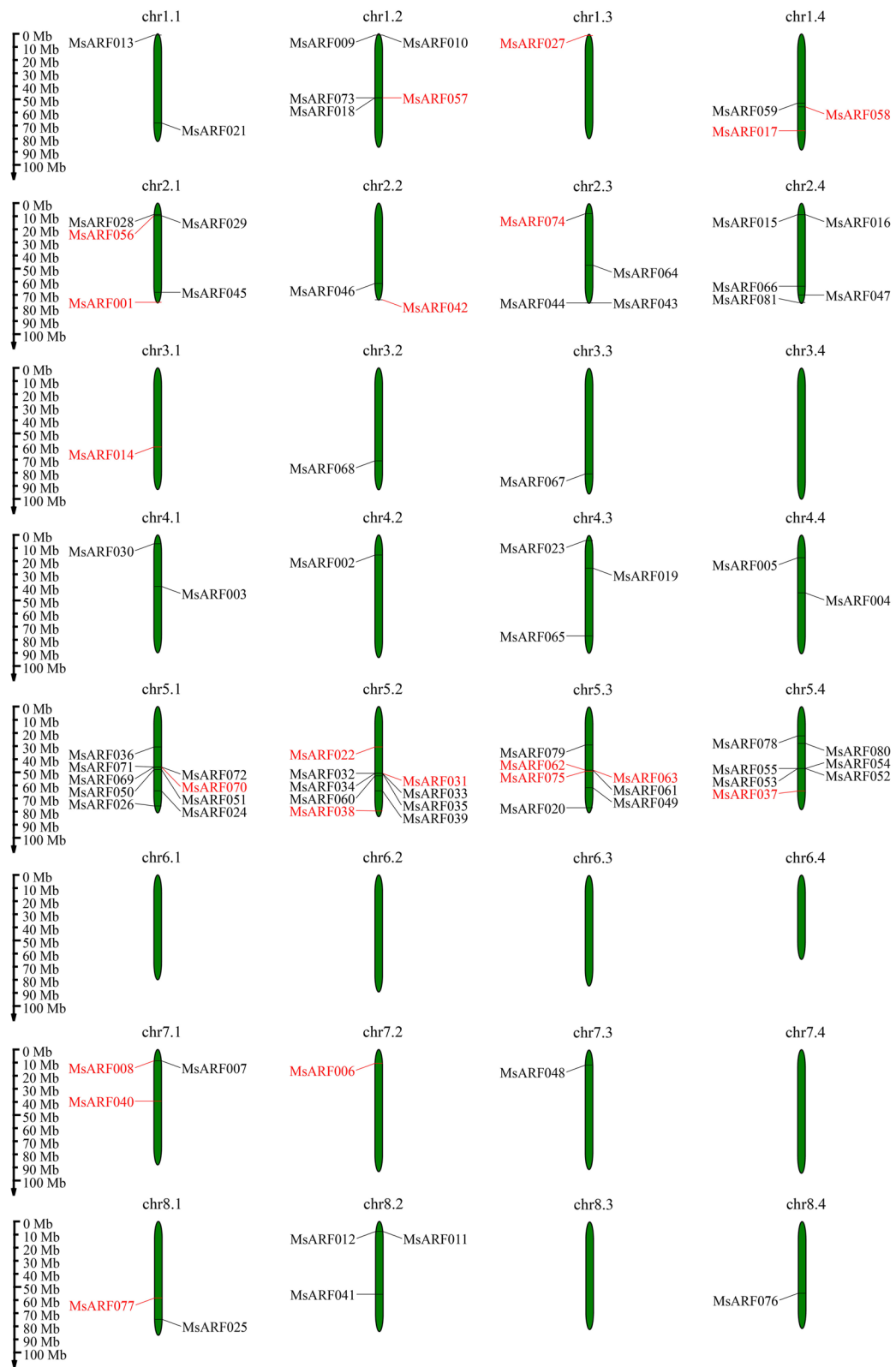


Fig. 4 Distribution and location of the *MsARF* gene on alfalfa chromosomes. Chr 1.1 to Chr 8.4 represent the linkage group of “Xinjiang Daye” alfalfa. Each black line indicates the location of the *ARF* gene. The red line represents tandem duplicates of the *ARF* gene

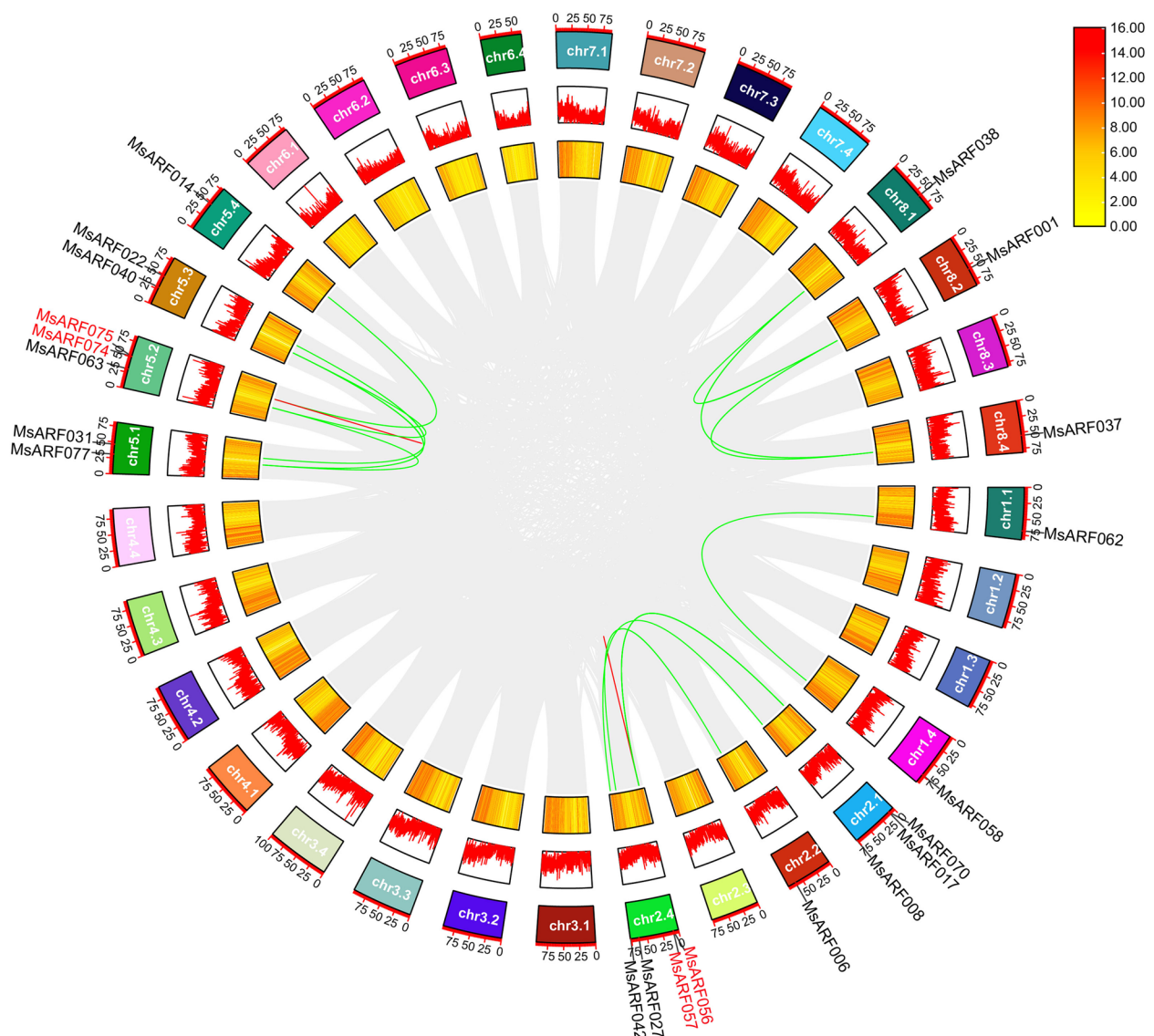


Fig. 5 Synteny analysis of the *MsARF* gene. The gray line represents all the synteny blocks in the alfalfa genome. The green and red lines represent gene pairs, and tandem repeat genes in the *ARF* gene, respectively. The yellow heat map and the broken red line map represent the gene density and expression level of the *MsARF* gene, respectively

treatment, cold stress, and metal ion stress (aluminum and lead) (Fig. 7B, C). In Fig. 7B, group a genes were mainly upregulated after 48 h of cold stress, participating in alfalfa's response to cold stress. *MsARF012* and *MsARF015* were mainly overexpressed after 3 h of salt stress. Five *MsARF* genes (*MsARF016*, *MsARF014*, *MsARF075*, *MsARF049*, and *MsARF013*) were overexpressed after 6 h of cold stress, and *MsARF011* was overexpressed after 6 h of mannitol treatment, suggesting that these genes may play a vital role in alfalfa's response to abiotic stress.

Real-time quantitative PCR analysis of the *MsARF* family genes

To verify the function of *MsARF*, a total of 25 *MsARF* genes (*MsARF005*, *MsARF007*, *MsARF010*, *MsARF017*, *MsARF018*, *MsARF025*, *MsARF026*, *MsARF029*, *MsARF035*, *MsARF039*, *MsARF041*, *MsARF043*, *MsARF044*, *MsARF045*, *MsARF047*, *MsARF050*, *MsARF051*, *MsARF053*, *MsARF055*, *MsARF060*, *MsARF064*, *MsARF066*, *MsARF071*, *MsARF080*, *MsARF081*) were selected to have their expression patterns analyzed during salt, drought, high-temperature,

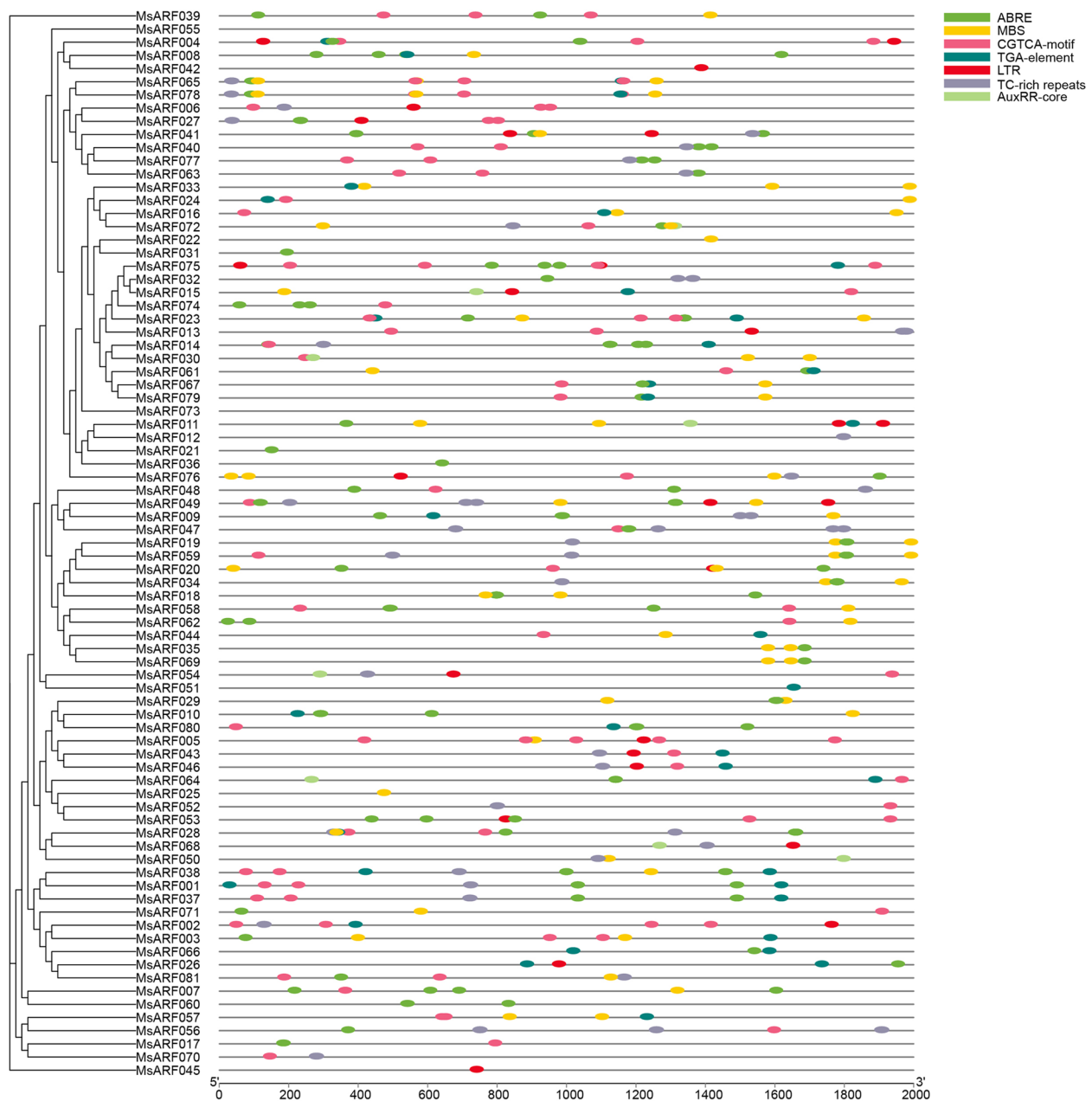


Fig. 6 Analysis of *Cis*-regulatory elements in the upstream promoter of the *MsARF* gene

and low-temperature stresses at different time points (Table S8), based on RNA-seq data (Fig. 7B). *MsARF041* and *MsARF047* showed no or very low expression and were not further analyzed. The expression pattern of *MsARFs* under salt stress is depicted in Fig. 8. Except for *MsARF051*, which is upregulated under salt stress, other genes showed a downregulated trend compared to the control (CK). In leaf tissue, *MsARF051* at 24 h, *MsARF081* and *MsARF029* at 12 h, *ARF080* at 12 and 24 h, *ARF010* at 6 h, *MsARF018* at 6 and 24 h, and

MsARF050 at 6 and 12 h were all upregulated, other genes at other times were all downregulated. In root tissue, only *MsARF044* and *MsARF029* at 12 h, and *MsARF025* at 24 h were upregulated. In stem tissue, the expressions of *MsARF* genes were diverse. Figure 9 shows the expression patterns of 23 *MsARF* genes under drought stress. *MsARF051* was upregulated under drought stress in all three tissues compared to CK. The changes in other genes were relatively diverse and tissue-specific, as well as related to the duration of treatment. Addition to, most

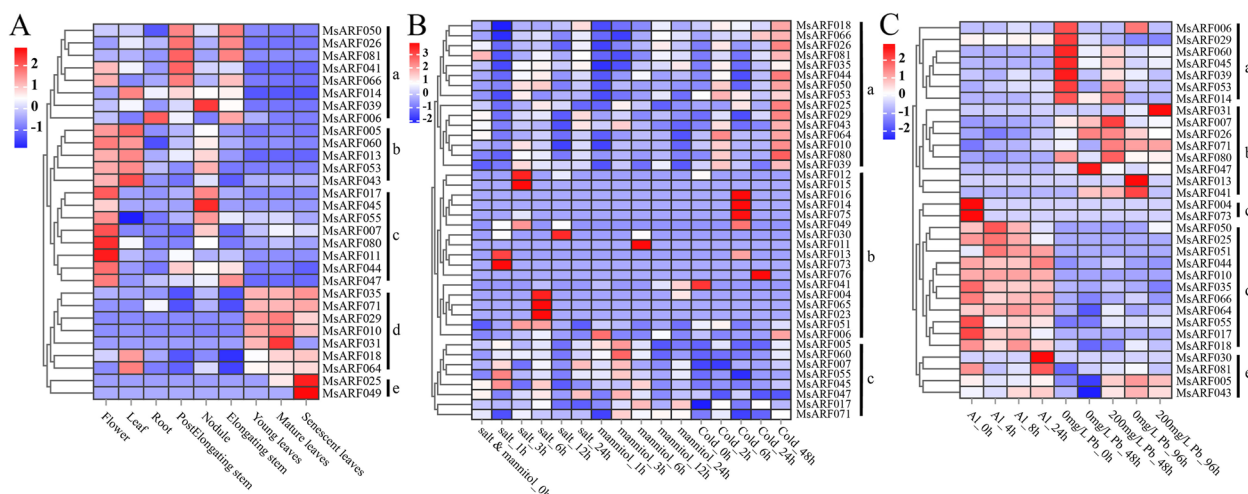


Fig. 7 Analysis of the expression pattern of the *MsARF* gene family. **A** Tissue-specific expression pattern analysis; **B** Analysis of expression patterns under salt, drought, and cold stress; **C** Analysis of expression patterns under aluminum (Al) and lead (Pb) stress

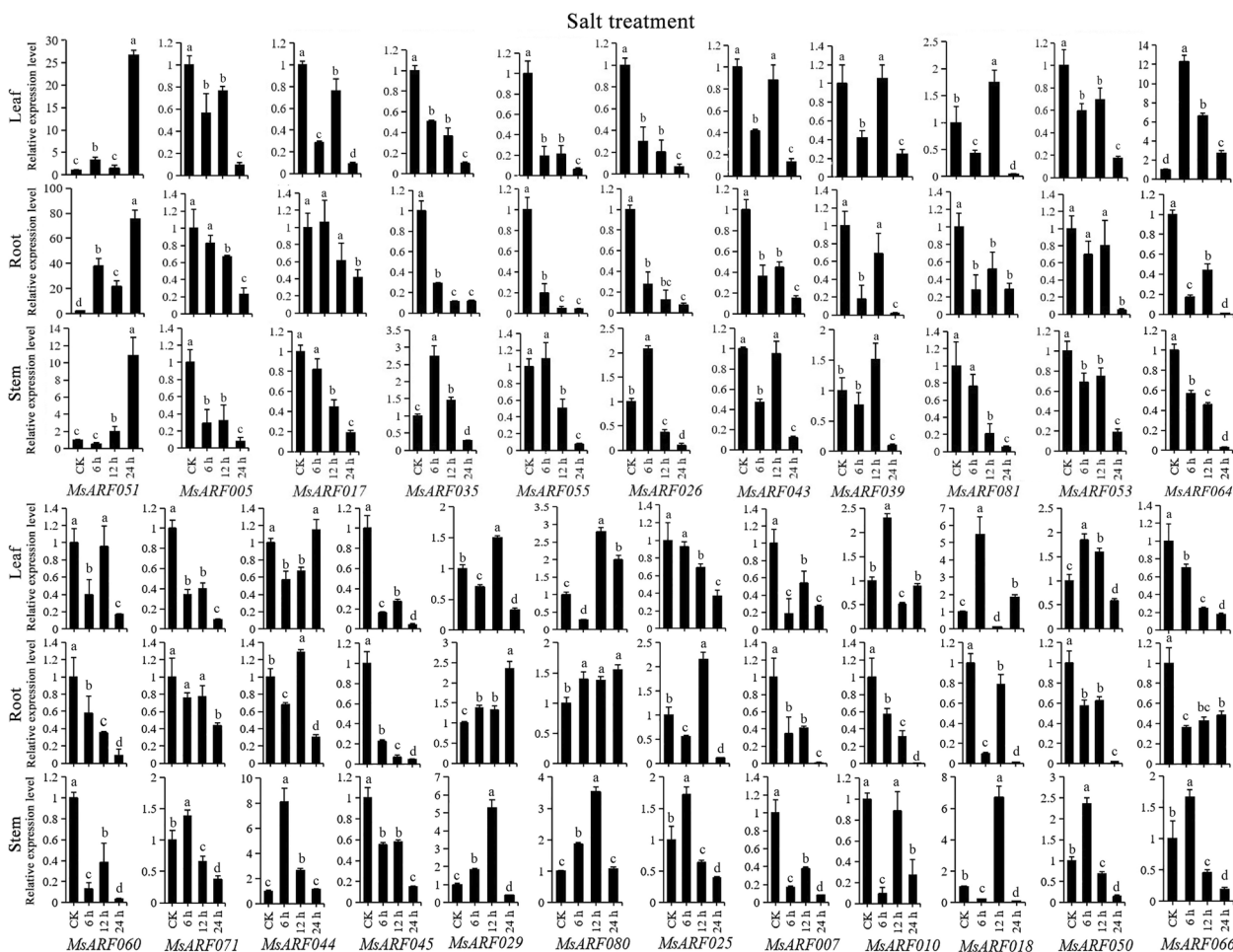


Fig. 8 The expression pattern analysis of *ARF* genes in diverse organs undergoing salt stress treatment for 0, 6, 12, and 24 h using qRT-PCR. Data are presented as the mean \pm standard deviation. Lowercase letters indicate a significant difference at the 5% level. The same below

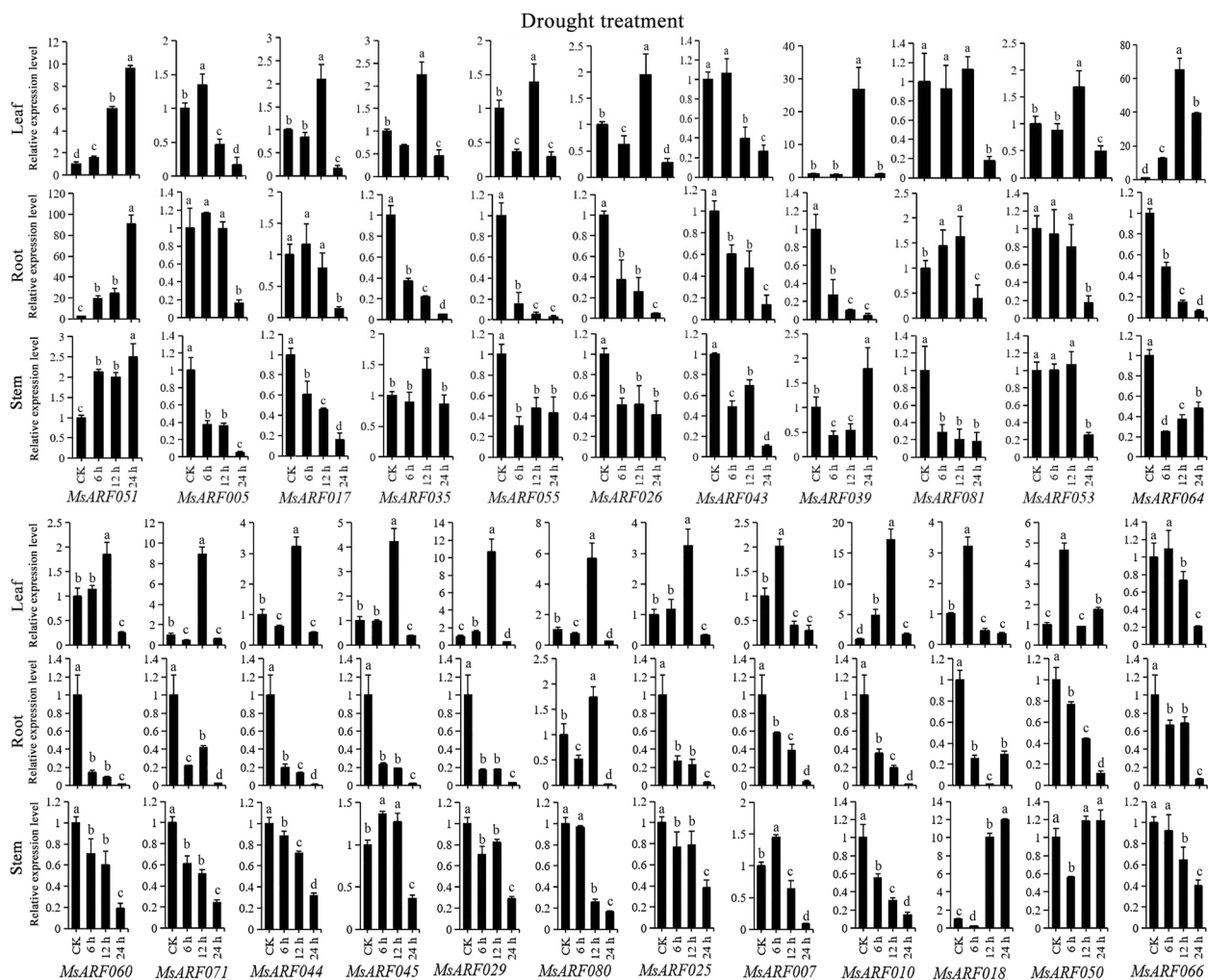


Fig. 9 The expression pattern analysis of *ARF* genes in diverse organs undergoing drought stress treatment for 0, 6, 12, and 24 h using qRT-PCR

genes (*MsARF007*, *MsARF010*, *MsARF017*, *MsARF018*, *MsARF025*, *MsARF029*, *MsARF035*, *MsARF039*, *MsARF044*, *MsARF045*, *MsARF050*, *MsARF051*, *MsARF053*, *MsARF055*, *MsARF060*, *MsARF064*, *MsARF071* and *MsARF080*) were mainly upregulated in the leaves at 12 h.

Under high-temperature treatment, *MsARF051* was upregulated in all three tissues, and treatment time lengths compared with CK (Fig. 10). *MsARF017*, *MsARF039* and *MsARF071* were upregulated at 12 h, indicating that their heat stress regulation was tissue-independent and mainly acted at 12 h. Meanwhile, we also found that most *MsARF* genes were upregulated at 12 h in leaves and stems undergoing heat stress, while the expression pattern was quite varied in roots. Under low-temperature treatment, *MsARF051* was upregulated in all three tissues and treatments compared

with CK (Fig. 11). *MsARF026*, *MsARF039*, *MsARF071* and *MsARF081* were upregulated at 12 h in leaf and stem tissues, with significantly distinct expression patterns in the roots. The changes in the remaining genes were tissue-specific and related to the treatment time length, and most genes were upregulated in specific tissues only at 12 h, suggesting that the regulation of cold stress by these genes was significantly correlated with the stress length. Compared with CK, *MsARF017*, *MsARF035*, *MsARF026*, and *MsARF043* were found in the leaves, and *MsARF017* was upregulated in the stem at 12 h and was downregulated at 6 and 24 h (Fig. S2). And compared to CK, *MsARF060* in roots and stems, while *MsARF044* in all three tissues were downregulated. The dynamic expression patterns of the remaining genes showed diversity.

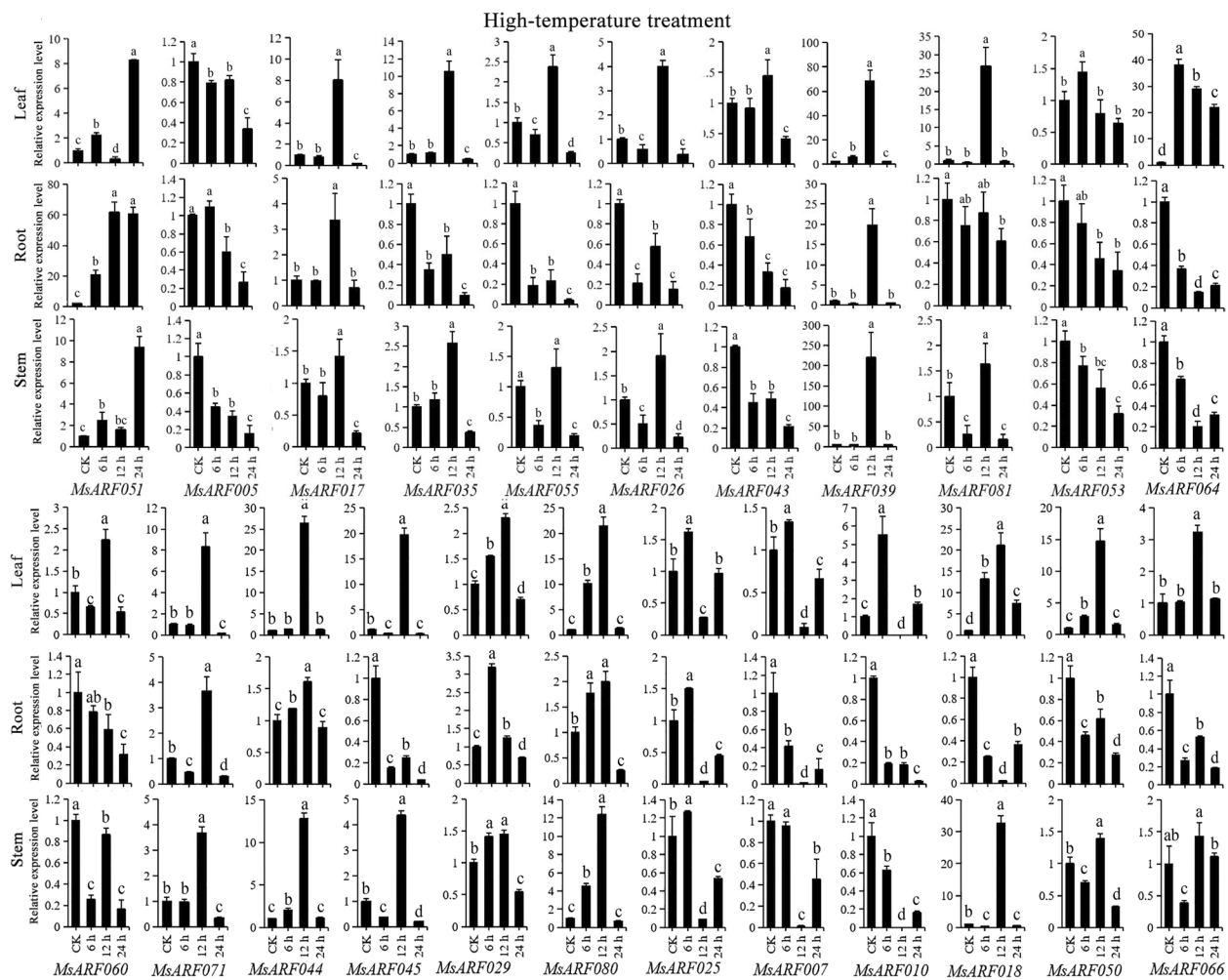


Fig. 10 The expression pattern analysis of *ARF* genes in diverse organs undergoing high-temperature stress treatment for 0, 6, 12, and 24 h using qRT-PCR

Discussion

Extreme weather conditions like drought, salt stress, and temperature have had a significant negative impact on alfalfa, one of the most economically important leguminous forages in the world. As a result, yield and geographic distribution have become major bottlenecks. As an important component of the auxin signaling pathway, *ARF* directly binds to and regulates the specific expression of downstream target genes during the auxin response [35]. Previous studies have shown that the *ARF* family plays a crucial role in plant growth, development, hormone response, and stress response [36, 37], making it a key gene family for understanding plant biology. Understanding the dynamics of this gene family provides an important resource for improving yield through variety improvement. However, there are no studies on the molecular function of alfalfa *ARF* transcription factors in response to abiotic stress. Therefore, studying the main

structure and expression characteristics of *ARF* genes in alfalfa helps us gain insight into the regulatory role of *ARF* genes in alfalfa resistance to abiotic stress, growth, and development. This study identified 81 *ARF* family members in the alfalfa genome and analyzed the possible role of *MsARF* in different tissues and abiotic stress responses in alfalfa for further studies on the function of *ARF* genes.

Members of the *ARF* family have revealed the evolution of tetraploid cultivated alfalfa

As a prominent effector of many aspects of the auxin response in plants, *ARF* transcription factors convert chemical signals into the transcriptional regulation of a group of specific genes [19]. Due to its functional importance in the auxin cascade, the *ARF* gene family has been deeply studied in model plants and major crops. The genome size of lettuce is 2.5 gigabases (Gb), which

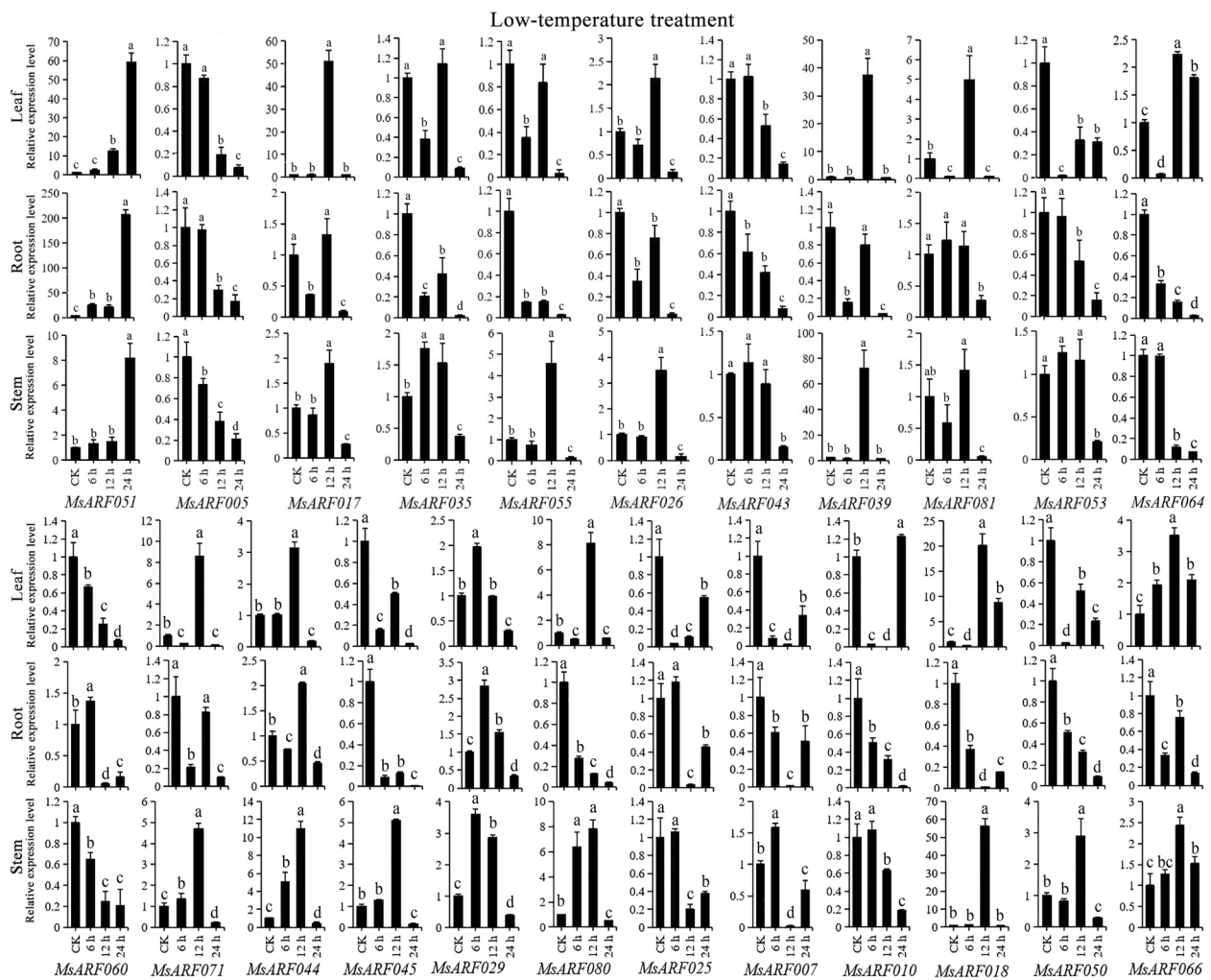


Fig. 11 The expression pattern analysis of *ARF* genes in diverse organs undergoing low-temperature stress treatment for 0, 6, 12, and 24 h using qRT-PCR

is considered one of the largest assembled plant genomes to date [38]. Compared to lettuce, the reference genome of autotetraploid alfalfa is larger, reaching up to 2.738 Gb [26]. Autotetraploid alfalfa can encode a larger *ARF* gene family based on such a large genome. This study identified 81 members of the *MsARF* family, which was greater than several reported crops, such as soybean (51), rice (25), and maize (36) [39–41]. Previous studies have shown that the genetic relationship between alfalfa and *Medicago truncatula* is closest [26]. A total of 40 *MtARF* genes were previously identified from the genome of diploid alfalfa [42], which is precisely half of the results of this study. These results may be related to the polyploidy of alfalfa (diploid genome), similar to the MADS-box genome-wide analysis of wheat [43]. Genome doubling (plant polyploidization) provides primitive genetic material for biological evolution, improves species diversity and environmental adaptability [44], and plays a vital role

in the adaptive evolution of angiosperms [45]. Therefore, identifying the tetraploid cultivated alfalfa *ARF* family from the genome level provides essential theoretical guidance and practical application value for research that examines the biological evolution, species protection, and genetic breeding of subsequent alfalfa.

Analysis of physicochemical properties of *MsARF* gene encoding proteins

This study's identification and analysis of the *MsARF* gene family was completed using the whole genome data of the "Xinjiang Daye" alfalfa cultivar. The hydrophilicity analysis of 81 *MsARF* proteins showed that the instability coefficient of the *MsARF* protein family ranged from 22.84 (*MsARF032*) to 98.98 (*MsARF039*), but the instability index of most genes (66) was higher than 40, indicating that they were not stable proteins. The aliphatic amino acid index of the 81 *MsARF* proteins ranged from

48.43 (MsARF067) to 93.76 (MsARF001), all of which were less than 100, indicating that they were hydrophilic proteins. The further subdivision of the protein family was carried out using the protein hydrophobicity coefficient. Among them, 22 MsARF proteins had a hydrophobicity coefficient less than -0.5 , indicating that they were hydrophilic proteins, while the other 59 MsARF proteins were between -0.5 and 0.5 , indicating that they were amphiphilic proteins. This result indicates that the MsARF family proteins have better hydrophilicity. As transcription factors, ARF proteins usually function in the nucleus [7]. Consistent with the subcellular localization results reported in other species, such as eggplant [7], tomato [46], and maize [41], the subcellular predictions indicate that most of the *MsARF* genes were localized to the nucleus. In addition, no *MsARF* genes were found to have any conducting peptides, signal peptides, or transmembrane structures, indicating that the 50 *MsARF* genes located in the nucleus did not secrete or transport but only regulated gene expression in the nucleus.

MsARF gene homology analysis

A phylogenetic tree was constructed to analyze the relationship between the *ARF* gene family in rice, *Arabidopsis*, and cultivated alfalfa (Fig. 2). The *ARF* gene family was divided into four large clades, where clade I contained the most *MsARF* (52) genes and was closely related to *AtARF10/16/17*. *AtARF10/16* controls root crown cell formation, limiting the stem cell niche and promoting columella cell differentiation [47]. *AtARF10* plays an important role in regulating the formation of the primary outer wall and pollen development in *Arabidopsis* [48]. Therefore, we hypothesized that these 52 *MsARF* genes have similar functions to *AtARF10/16/17*. Clade II contained 13 alfalfa genes that were closely related to *AtARF5/6/7/8/19*, and it has been speculated that these 13 *MsARF* genes may be auxin-promoting factors in carrot protoplasts [49]. Clade III contained seven alfalfa genes, which were closely related to *AtARF3* and *AtARF4*, and it has been speculated that these seven *MsARF* may be auxin suppressors [49]. Clade IV contained nine alfalfa genes, which were closely related to 13 *Arabidopsis* genes such as *AtARF1*, *AtARF2*, and *AtARF11*, suggesting that these nine *MsARF* may act as auxin repressors involved in the Abscisic Acid (ABA) pathway regulating seed germination and primary root growth [50].

The analysis of the MsARF gene structure

Based on the phylogenetic tree, analysis of ten conserved motifs of *MsARF* genes revealed different clades of *MsARF* genes containing common or specific motifs (Fig. 3). Although the number of motif members varied

among clades, the group motif patterns were strongly conserved. The comparison of novel functional domains or motif sequences across multiple homologous proteins has been a widely used method to predict protein function based on evolutionary conservation [7]. We found that motif 8 was highly conserved in the *MsARF* gene family. In addition, we also found that motif 7 was highly conserved in the *MsARF* subfamily in Clades II, III, and IV. Although most of the motifs in *MsARF* were conserved, other motifs may have been related to novel plant functions and should be studied further. By analyzing the conserved domains of *MsARF* protein sequences, we found that all *MsARFs* contain an auxin resp domain, and 71 *MsARFs* contain a typical DBD domain (B3-like), while only 19 (23.46%) *MsARFs* contain an auxin/IAA binding domain and are mainly distributed in clades II, III, and IV. This result was similar to the conserved domain of the ARF protein sequence in *Medicago truncatula*, indicating that auxin can also regulate *MsARF* genes independently [51].

To further determine the structural differences between *MsARF* genes, gene structure analysis was conducted on the *MsARF* family. The number of CDS in the *MsARF* gene ranged from 1 to 15, and no UTR region was identified. This finding was consistent with the previous reports in papaya [9], maize [41], and barley [52], indicating that the gene structure of the *ARF* family was relatively simple. Gene structure analysis determined the structural differences among the *MsARF* gene family. Analysis of the distribution of 7 *cis*-regulatory elements in the 2000 bp (bp) upstream promoter region of the *MsARF* revealed that ABRE had the most *cis*-regulatory elements, followed by CGTCA motifs, MBS, TC-rich repeats, TGA elements, LTR, and AuxRR cores (Fig. 5), indicating that *MsARF* played a positive role in hormone responses to plant stressors.

Chromosomal localization and collinearity analysis of the MsARF

We found that 81 *MsARF* genes were unevenly distributed on 25 chromosomes (Chr) of alfalfa and were the most widely distributed on chromosome 5, indicating that *MsARF* genes mainly perform biological functions on chromosome 5. Gene replication is traditionally considered a way to expand and obtain functional diversity during evolution [53]. Gene duplication analysis in the *MsARF* gene family showed that fragment duplication had a greater impact on *ARF* genes, which was consistent with the results observed in rice and *Arabidopsis* in previous studies [54], suggesting that replication events in the *ARF* family within a plant genome may be a general evolutionary mechanism. Ka/Ks ratios are usually an important indicator

of selection pressure in evolution [34]. The analysis of homologous *MsARF* gene pairs showed that only 2 pairs of fragment replicates (*MsARF008/MsMASS042* and *MsARF008/MsMASS042*) had a Ka/Ks ratio greater than 1, while the remaining 14 pairs of *MsARF* had a Ka/Ks ratio less than 1, indicating that they experienced strong purification selection after duplication.

The *MsARF* genes function in response to abiotic stress

Since *ARF* genes play a vital role in controlling phytohormone signaling, studying each family member's function is necessary. ARF transcription factors are crucial in plant growth and development and actively participate in plant resistance to stress [6, 17, 19, 20]. This study combined transcriptome data with a novel bioinformatic approach to search for *MsARF* genes that have functions related to the response and regulation of drought stress in alfalfa. We compared the *MsARF* genes with known transcriptome data and initially selected genes that were significantly differentially expressed under drought, salt, and cold stress as *ARF* candidate genes responding to abiotic stress. Subsequently, alfalfa seedlings were treated with salt, drought, cold, and heat stress, and qRT-PCR experimentally verified their expression patterns. As reported by Zhou et al. [55], *ARF12* might play central roles in the regulation of NaCl-responsive genes. In addition, *AcARF5* can positively regulate salt stress and drought stress in kiwifruit (*Actinidia chinensis*) [56]. When compared to CK, *MsARF051* was upregulated in all treatment groups, which helped us elucidate that *MsARF051* had great potential for transformation in the cultivation of reverse-resistant alfalfa varieties. In addition, *MsARF017* and *MsARF039*, *MsARF026*, *MsARF071* and *MsARF081* were upregulated during temperature stress, and *MsARF039* could regulate heat stress and cold stress in alfalfa. Furthermore, we found that most of the *ARF* genes were upregulated at 12 h under all abiotic stresses involved in this study, and we speculated that these *ARF* genes were mainly active at 12 h and belong to short-acting genes, especially in the leaves and stems. Future studies focusing on alfalfa should examine these genes and conduct further research through genetic engineering and editing techniques in our research. The results of this study on the alfalfa *ARF* gene family at the whole genome level create a foundation for mining stress resistance genes for alfalfa and provide an in-depth analysis of abiotic stress theory and guidance for the creation of new germplasm using genetic engineering technology for an alfalfa variety that has increased stress resistance.

Conclusion

Based on the whole genome of cultivated alfalfa, this study identified the *ARF* gene family, predicted the basic physical and chemical properties of the genes, and analyzed the phylogenetic relationship, gene structures, chromosomal localization, paralogous and homologous genes, and expression patterns of *ARF* genes in response to abiotic stress and different tissues. A total of 81 *MsARF* genes were identified and phylogenetically divided into four branches, analyzing 14 discovered segmental repeat pairs and two sets of tandem repeats. Compared with the transcriptomic data of alfalfa, the expression of *MsARF051* under all stress treatments can be used as the candidate gene of alfalfa. *MsARF039* is mainly involved in the response of alfalfa to temperature stress and can be used to improve new alfalfa varieties cultivated in areas with extreme temperatures. Future research will examine the functional verification of these genes and analyze the genetic transformation of these two genes to lay a foundation for further investigation of the abiotic stress theory and the creation of new germplasm using genetic engineering technology for an alfalfa variety that has increased stress resistance.

Materials and methods

Materials

The material used in this experiment was cultivated "Xinjiang Daye" alfalfa, and its seeds were provided by the National Livestock Husbandry Station, Ministry of Agriculture and Rural Affairs of The People's Republic of China.

Identification of the *MsARF* gene family and analysis of its physical and chemical properties

The Plant Transcription Factor Database (<http://planttfdb.gao-lab.org/>) was used to download the *ARF* protein sequences from Arabidopsis and rice model plants. The genome data and protein-nucleic acid sequence of "Xinjiang DaYe" alfalfa used in this experiment were downloaded from the alfalfa genome website (https://figshare.com/projects/whole_genome_sequencing_and_assembly_of_Medicago_sativa/66380) [26]. In addition, the hidden Markov model (HMM) profile of ARF (PF06507) was downloaded from the Pfam database (<http://pfam.xfam.org/>), and *MsARF* family proteins from alfalfa genome were identified using HMMER 3.0 with an E-value set to 1.0 as the threshold and the remaining parameters set to default. The redundant protein was removed using the Expasy [57] online database (https://web.expasy.org/decrease_redundancy), with all parameters set to default values. Further characterization of the ARF domain in the *MsARF* protein sequences was performed using the Pfam

website and NCBI-CD Search (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.Cgi>) to verify the accuracy of the MsARF protein sequences after the removal of the redundancy. Sequences without the ARF domain were removed [58]. A total of 81 MsARF protein sequences were subsequently analyzed for physicochemical properties using the online network ProtParam (<https://web.expasy.org/protparam/>) tools, including molecular weight (MW), theoretical isoelectric point (pI), instability index, and the grand average of hydropathicity (GRAVY) index. Subcellular localization (https://www.genscript.com/wolf_psort.html) was predicted for all 81 *MsARF* genes by the online tool WoLF PSORT. The predicted genes were named *MsARF001* through *MsARF081* according to their order of occurrence in the genome. In addition, GO and KEGG annotations were performed using the Alfalfa Database (<http://47.92.172.28:12088/>) to investigate the functional role of the *MsARF* genes [59, 60].

Phylogenetic analysis, gene structure and motif composition of the *MsARFs*

To analyze the evolutionary relationship between MsARF proteins, multiple sequence alignments of MsARF, AtARF, and OsARF protein sequences were performed using MEGA 7.0 software [61] and automatically trimmed with TrimAL [62]. The maximum likelihood phylogenetic tree was constructed from IQ-TREE version 1.6.12, and the JTT + R6 model was determined to be the best sub-phylogenetic model using ModelFinder, with the Ultrafast bootstrap set to 1000 [63, 64]. Based on the evolutionary relationship between *MsARF* genes and *AtARF*, *MsARF* genes were classified into a clade I-IV type [65]. The MEME 4.12.0 online tool (<http://meme-suite.org/>) was used to identify the conserved motifs of the 81 alfalfa MsARF proteins. The parameters were set as follows: The maximum number of motifs was 10; each sequence selected a zero or one occurrence as the site distribution; the minimum and maximum motif widths were set to 6 and 200, respectively; other parameters were set to default values; and the motif features were visualized with TBtools [33].

Gene duplication and gene structure analysis

CDS sequences and gene sequences corresponding to all *MsARF* genes were obtained from the alfalfa genome file for prediction analysis of intron and exon structures of *MsARF* genes using GSDS2.0 (gene structure display server) (<http://gsds.gao-lab.org/>). MapGene2Chrome V2.0 (http://mg2c.iask.in/mg2c_v2.0/) was used to map chromosome localization based on the alfalfa genome annotation file (GFF3) data to explore the distribution characteristics of *ARF* genes on chromosomes. Based on

the alfalfa genome sequence file in fasta format and the corresponding gene structure annotation file in GFF3/GTF format, referring to the method of Chen et al. [33], TBtools software was used to complete the chromosome location of genes and the collinearity analysis of gene duplications. The specific steps: first, the “Fasta Stats” function was opened, the alfalfa genome file was input, and the chromosome length information file was obtained. Next, the “Gene Density Profile” function was used, and gene structure annotation files were input to obtain gene density information. Then, by using the “one step MCScanX-super fast” function, the alfalfa genome file and corresponding gene structure annotation file were input to obtain the collinearity file of genes within the species. Subsequently, the association file between genes was obtained through the “File Merge for MCScanX” and “File Transform for MicroSynteny Viewer” functions, and the green RGB value was added to the *MsARF* gene pairs, as well as the red RGB value was added to the tandem repeat sequence genes. Finally, the “Advanced Circos” function was used to input the file obtained in the previous steps to draw the circos diagram.

Protein signal peptide, guide peptide, transmembrane structure and protein secondary and tertiary structure analysis

The signal peptide and transmembrane structures of the *MsARF* were analyzed using SignalP 5.0 (<https://services.healthtech.dtu.dk/services/SignalP-5.0/>) and TMHMM 2.0 (<https://services.healthtech.dtu.dk/services/TMHMM-2.0/>), respectively. The conducting, or guide, peptide analysis of the *MsARF* was also performed using the Target P-2.0 Server (<https://services.healthtech.dtu.dk/services/TargetP-2.0/>). The MsARF protein secondary structure was analyzed using SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.plpage=npsa_sopma.html) to obtain the percentage of different conformations, including α -helix, extended chain, β -turn angle, and random coil. Using SWISS-MODEL (<http://swissmodel.expasy.org/interactive>), this structural analysis was subsequently used to analyze the tertiary structure of MsARF proteins.

Cis-regulatory element analysis

The PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) analyzed *cis*-regulatory elements in the 2000-bp sequence upstream of the upstream transcription start site of the 81 *MsARF* genes. In the promoter region of alfalfa *ARF* genes, seven *cis*-regulatory elements known to associate with stress and hormones were identified, including ABRE, TC-rich repeats, MBS, LTR, CGTCA-motif, AuxRR-core, and TGA-element, which are involved in the abscisic acid

response, defense and stress response, drought induction, low-temperature response, Methyl Jasmonate (MeJA) response, auxin response, and auxin response *cis*-elements.

***MsARF* gene in response to abiotic stress**

To investigate the expression level of *MsARF* under abiotic stress and various tissue sites, BLASTn alignment with the Alfalfa Database produced the expression profile data of *MsARF* genes under abiotic stress and various tissue sites. To deeply investigate the expression patterns of *MsARF* genes in different tissues, expression profile data from eight tissues were analyzed, including flowers, leaves, roots, post-elongating stems, nodules, elongating stems, young leaves, mature leaves, and senescent leaf tissues of alfalfa. The expression analysis of *MsARF* genes under abiotic stress included the transcriptional expression profiles of alfalfa under cold, drought, salt, aluminum, and lead stress. OmicShare Tools (<https://www.omicshare.com/tools/>), a freely available data analysis platform, was used for cluster analysis and expression quantity heat map mapping, and the expression level was expressed by Z-core normalized calculation results for the expression data.

Growth and stress treatment of plant materials

The plant material used in this experiment was “Xinjiang DaYe” alfalfa. First, we chose seeds with healthy and identical shapes. After surface sterilization of alfalfa seeds, they were planted in pots with vermiculite: nutrient soil = 1:1 [20 centimeters (cm) diameter, 20 cm high] with 5 biological replicates per treatment group. The seedlings were cultured in a 16-hour light/8-hour dark cycle with 60% relative humidity and a temperature of 22 °C. After 15 days, the alfalfa seedlings were inoculated with different stress treatments. A 15% polyethylene glycol-6000 solution was added to the seedling pots by root irrigation (until thoroughly poured) to simulate drought. To simulate salt stress, 220 micromolar (mM) sodium chloride solution was added to the seedling pots by root irrigation (until thoroughly poured). The whole alfalfa seedlings were harvested at different stress time points (6, 12, and 24 h) for the drought treatment group and the salt stress treatment group. To simulate cold stress, alfalfa seedlings to be treated were transferred to a 16-hour light/8-hour dark cycle in an artificial climate chamber with 60% relative humidity and temperatures at 5 °C. To simulate high temperature stress, alfalfa seedlings to be treated were transferred to a 16-hour light (32 °C)/8-hour dark (28 °C) cycle in an artificial climate chamber with 60% relative humidity. Whole alfalfa seedlings were collected after 6, 12, and 24 h of incubation. All samples were flash-frozen

in liquid nitrogen and stored at 80 °C for downstream analysis.

RNA extraction and qRT-PCR

To verify the response pattern of *MsARF* genes to abiotic stress, twelve genes were selected for qRT-PCR analysis based on the expression heatmap. According to the instructions of the manufacturer, total ribonucleic acid (RNA) from the root, stem, and leaf tissues of “Xinjiang Daye” alfalfa under different stress treatment times were extracted using an RNA Simple Total RNA kit (Tiangen, Shanghai, China) and were reverse transcribed into cDNA using a FastKing cDNA kit (Tiangen, Shanghai, China). Using the Primer-BLAST tool at NCBI (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>), twelve gene-specific PCR primers were designed (Table S9) for gene expression analysis. The QuantStudio 5 real-time PCR system (Thermo Scientific, Massachusetts, USA) was used to amplify samples and standards with Hyperreal Premix Plus (SYBR Green) (Tiangen, Shanghai, China). Three biological replicates were used. Chen et al. described the experimental process that we used as a model for our experimentation [61]. The relative transcription levels of the selected genes were calculated with the $2^{-\Delta\Delta CT}$ method [62] and normalized to the expression levels of the *Medicago truncatula* ACTIN gene (AES78237. 1). The qRT-PCR data were collated and plotted in Excel 2016, and the one-way ANOVA ($p < 0.05$) was analyzed by IBM SPSS 24.0 software.

Abbreviations

ARF	Auxin-Response Factor
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
DBD	DNA binding domain
AuxRE	Auxin response elements
MR	Variable intermediate region
AD	Activating domain
RD	Inhibitory domain
CTD	C-terminal dimerization domain
Chr	Chromosome
Ka/Ks	Nonsynonymous substitution rate/synonymous substitution rate
FPKM	Fragments Per Kilobase of transcript per Million mapped reads

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09610-z>.

Additional file 1: Table S1. List of the ARF sequences in Alfalfa. **Table S2.** Protein property of ARF proteins. **Table S3.** Gene ontology (GO) annotation results of ARF genes. **Table S4.** Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation results of ARF genes. **Table S5.** The subcellular localization and secondary structure analysis of ARF proteins. **Table S6.** Ka and Ks of *MsARF* gene pairs. **Table S7.** Expression (FPKM) of ARF genes in various alfalfa tissues in response to cold, drought, salt, aluminum, and lead treatment. **Table S8.** Relative expression data of 23

MsARF genes related to abiotic stress. **Table S9.** Sequences of ARF primers used in qRT-PCR.

Additional file 2: Fig. S1. Conserved Motif of ARF proteins. **Fig. S2.** Relative expression level analysis of ten ARF genes without stress treatment for 0, 6, 12, and 24 hours using qRT-PCR.

Acknowledgements

We would like to thank Dr. Joseph Elliot at the University of Kansas for his assistance with English language and grammatical editing of the manuscript.

Authors' contributions

HLM and FQC conceived and designed the experiment. FQC, JQZ and XH performed the experiments. FQC and JQZ analyzed all the data. FQC wrote the manuscript. HLM revised the manuscript. All of the authors read and approved the final manuscript.

Funding

This research was supported by the alfalfa breeding project (2022ZD0401102).

Availability of data and materials

All data generated or analyzed in this study are included in this published article and its supplementary material. The draft genome data of autotetraploid cultivated ('Xinjiang Daye') alfalfa was obtained from figshare (https://figshare.com/projects/whole_genome_sequencing_and_assembly_of_Medicago_sativa/66380). The *Arabidopsis* and rice ARF protein sequences were all available in the NCBI SRA repository (<http://plantfdb.gao-lab.org/>). Genome-wide transcriptome data of different alfalfa tissues were acquired from the *Medicago* Analysis Portal (<https://medicago.legumeinfo.org/>). All transcriptome sequencing data analysed during the current study are available in the NCBI SRA repository (<https://www.ncbi.nlm.nih.gov/sra/>): SRR7091780-SRR7091794 (cold treatment), SRR7160322-SRR7160357 (drought and salt treatments), SRR22519684-SRR22519695 (aluminum stress), and SRR5279707-SRR5279711 (lead stress).

Declarations

Ethics approval and consent to participate

Experimental research and studies on plants in this study, including the collection of plant material, are complied with the institutional, national, and international guidelines and legislation. We conducted the experimental research on cultivated alfalfa in accordance with the IUCN Policy Statement on Research Involving Species at Risk of Extinction and the Convention on the Trade in Endangered Species of Wild Fauna and Flora. Alfalfa seeds of "Xinjiang Daye" was gifted by the National Livestock Husbandry Station, Ministry of Agriculture and Rural Affairs of The People's Republic of China.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹College of Pratacultural Science, Gansu Agricultural University, Key Laboratory of Grassland Ecosystem, Ministry of Education, Pratacultural Engineering Laboratory of Gansu Province, Sino-US. Center for Grazingland Ecosystem Sustainability, Yingmencun, Anning District, Gansu province, Lanzhou, Gansu 730070, China.

Received: 19 May 2023 Accepted: 20 August 2023

Published online: 29 August 2023

References

- Su YH, Liu YB, Bai B, Zhang XS. Establishment of embryonic shoot-root axis is involved in auxin and cytokinin response during *Arabidopsis* somatic embryogenesis. *Front Plant Sci.* 2014;5:792.

- Goldental-Cohen S, Israeli A, Ori N, Yasuor H. Auxin response dynamics during wild-type and entire flower development in tomato. *Plant Cell Physiol.* 2017;58(10):1661–72. <https://doi.org/10.1093/pcp/pcx102>.
- Liu X, Dinh TT, Li D, Shi B, Li Y, Cao X, Guo L, Pan Y, Jiao Y, Chen X. AUXIN RESPONSE FACTOR 3 integrates the functions of AGAMOUS and APETALA2 in floral meristem determinacy. *Plant J.* 2014;80(4):629–41. <https://doi.org/10.1111/tpj.12658>.
- Di GQ. The biosynthesis of auxin: how many paths truly lead to IAA? *Plant Growth Regul.* 2016;78(3):275–85.
- Guilfoyle TJ. The PB1 domain in auxin response factor and Aux/IAA proteins: a versatile protein interaction module in the auxin response. *Plant Cell.* 2015;27(1):33–43. <https://doi.org/10.1105/tpc.114.132753>.
- Bouzroud S, Gouiaa S, Hu N, Bernadac A, Mila I, Bendaou N, Smouni A, Bouzayen M, Zouine M. Auxin Response Factors (ARFs) are potential mediators of auxin action in tomato response to biotic and abiotic stress (*Solanum lycopersicum*). *PLoS One.* 2018;13(2): e0193517. <https://doi.org/10.1371/journal.pone.0193517>.
- Chen J, Wang S, Wu F, Wei M, Li J, Yang F. Genome-wide identification and functional characterization of Auxin Response Factor (ARF) genes in eggplant. *Int J Mol Sci.* 2022;23(11): 6219. <https://doi.org/10.3390/ijms23116219>.
- Guilfoyle TJ, Hagen G. Auxin response factors. *Curr Opin Plant Biol.* 2007;10(5):453–60. <https://doi.org/10.1016/j.pbi.2007.08.014>.
- Liu K, Yuan C, Li H, Lin W, Yang Y, Shen C, Zheng X. Genome-wide identification and characterization of auxin response factor (ARF) family genes related to flower and fruit development in papaya (*Carica papaya* L.). *BMC Genomics.* 2015;16(1):901. <https://doi.org/10.1186/s12864-015-2182-0>.
- Shen C, Wang S, Bai Y, Wu Y, Zhang S, Chen M, Guilfoyle TJ, Wu P, Qi Y. Functional analysis of the structural domain of ARF proteins in rice (*Oryza sativa* L.). *J Exp Bot.* 2010;61(14):3971–81. <https://doi.org/10.1093/jxb/erq208>.
- Vernoux T, Brunoud G, Farcot E, Morin V, Van den Daele H, Legrand J, Oliva M, Das P, Larrieu A, Wells D, et al. The auxin signalling network translates dynamic input into robust patterning at the shoot apex. *Mol Syst Biol.* 2011;7:508.
- Finet C, Berne-Dedieu A, Scutt CP, Marlétaz F. Evolution of the ARF gene family in land plants: old domains, new tricks. *Mol Biol Evol.* 2013;30(1):45–56. <https://doi.org/10.1093/molbev/mss220>.
- De Smet I, Lau S, Voss U, Vanneste S, Benjamins R, Rademacher EH, Schlereth A, De Rybel B, Vassileva V, Grunewald W, et al. Bimodular auxin response controls organogenesis in *Arabidopsis*. *Proc Natl Acad Sci USA.* 2010;107(6):2705–10.
- Liscum E, Reed JW. Genetics of Aux/IAA and ARF action in plant growth and development. *Plant Mol Biol.* 2002;49(3–4):387–400.
- Liu Z, Miao L, Huo R, Song X, Johnson C, Kong L, Sundaresan V, Yu X. ARF2-ARF4 and ARF5 are essential for female and male gametophyte development in *Arabidopsis*. *Plant Cell Physiol.* 2018;59(1):179–89. <https://doi.org/10.1093/pcp/pcx174>.
- Ghelli R, Brunetti P, Napoli N, De Paolis A, Cecchetti V, Tsuge T, Serino G, Matsui M, Mele G, Rinaldi G, Ghelli R, Brunetti P, Napoli N, De Paolis A, Cecchetti V, Tsuge T, Serino G, Matsui M, Mele G, Rinaldi G, Palumbo GA, Barozzi F, Costantino P, Cardarelli M. A newly identified flower-specific splice variant of AUXIN RESPONSE FACTOR8 regulates stamen elongation and endothecium lignification in *Arabidopsis*. *Plant Cell.* 2018;30(3):620–37. <https://doi.org/10.1105/tpc.17.00840>.
- Zheng Y, Zhang K, Guo L, Liu X, Zhang Z. AUXIN RESPONSE FACTOR3 plays distinct role during early flower development. *Plant Signal Behav.* 2018;13(5): e1467690. <https://doi.org/10.1080/15592324.2018.1467690>.
- Okushima Y, Overvoorde PJ, Arima K, Alonso JM, Chan A, Chang C, Ecker JR, Hughes B, Lui A, Nguyen D, Okushima Y, Overvoorde PJ, Arima K, Alonso JM, Chan A, Chang C, Ecker JR, Hughes B, Lui A, Nguyen D, Onodera C, Quach H, Smith A, Yu G, Theologis A. Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in *Arabidopsis thaliana*: unique and overlapping functions of ARF7 and ARF19. *Plant Cell.* 2005;17(2):444–63. <https://doi.org/10.1105/tpc.104.028316>.
- Hu M, Qi Z, Ren Z, Tong J, Wang B, Wu Z, Hao J, Liu N. Genome-wide analysis of Auxin Response Factors in lettuce (*Lactuca sativa* L.) reveals the positive roles of LsARF8a in Thermally Induced Bolting. *Int J Mol Sci.* 2022;23(21): 13509. <https://doi.org/10.3390/ijms232113509>.

20. Wang Y, Li B, Li Y, Du W, Zhang Y, Han Y, Liu C, Fan S, Hao J. Application of exogenous auxin and gibberellin regulates the bolting of lettuce (*Lactuca sativa* L). *Open life Sciences*. 2022;17(1):438–46.
21. Hao JH, Zhang LL, Li PP, Sun YC, Li JK, Qin XX, Wang L, Qi ZY, Xiao S, Han YY, Hao J-H, Zhang L-L, Li P-P, Sun Y-C, Li J-K, Qin X-X, Wang Lu, Qi Z-Y, Xiao S, Han Y-Y, Liu C-J, Fan S-X. Quantitative proteomics analysis of lettuce (*Lactuca sativa* L.) reveals molecular basis-associated auxin and photosynthesis with bolting induced by high temperature. *Int J Mol Sci*. 2018;19(10): 2967. <https://doi.org/10.3390/ijms19102967>.
22. Hu W, Zuo J, Hou X, Yan Y, Wei Y, Liu J, Li M, Xu B, Jin Z. The auxin response factor gene family in banana: genome-wide identification and expression analyses during development, ripening, and abiotic stress. *Front Plant Sci*. 2015;6:742.
23. Dong X, Deng H, Ma W, Zhou Q, Liu Z. Genome-wide identification of the MADS-box transcription factor family in autotetraploid cultivated alfalfa (*Medicago sativa* L.) and expression analysis under abiotic stress. *BMC Genomics*. 2021;22(1):603. <https://doi.org/10.1186/s12864-021-07911-9>.
24. Liu W, Xiong C, Yan L, Zhang Z, Ma L, Wang Y, Liu Y, Liu Z. Transcriptome analyses reveal candidate genes potentially involved in AI stress response in alfalfa. *Front Plant Sci*. 2017;8:26.
25. Liu Z, Chen T, Ma L, Zhao Z, Zhao PX, Nan Z, Wang Y. Global transcriptome sequencing using the Illumina platform and the development of EST-SSR markers in autotetraploid alfalfa. *PLoS One*. 2013;8(12): e83549. <https://doi.org/10.1371/journal.pone.0083549>.
26. Chen H, Zeng Y, Yang Y, Huang L, Tang B, Zhang H, Hao F, Liu W, Li Y, Liu Y, Chen H, Zeng Y, Yang Y, Huang L, Tang B, Zhang He, Hao F, Liu W, Li Y, Liu Y, Zhang X, Zhang Ru, Zhang Y, Li Y, Wang K, He H, Wang Z, Fan G, Yang H, Bao A, Shang Z, Chen J, Wang W, Qiu Q. Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nat Commun*. 2020;11(1):2494. <https://doi.org/10.1038/s41467-020-16338-x>.
27. Shen C, Du H, Chen Z, Lu H, Zhu F, Chen H, Meng X, Liu Q, Liu P, Zheng L, Shen C, Du H, Chen Z, Lu H, Zhu F, Chen H, Meng X, Liu Q, Liu P, Zheng L, Li X, Dong J, Liang C, Wang T. The chromosome-level genome sequence of the autotetraploid alfalfa and resequencing of core germplasm provide genomic resources for alfalfa research. *Mol Plant*. 2020;13(9):1250–61. <https://doi.org/10.1016/j.molp.2020.07.003>.
28. Peng Y, Fang T, Zhang Y, Zhang M, Zeng L. Genome-wide identification and expression analysis of Auxin Response Factor (ARF) gene family in Longan (*Dimocarpus longan* L.). *Plants*. 2020;9(2): 221. <https://doi.org/10.3390/plants9020221>.
29. Luo XC, Sun MH, Xu RR, Shu HR, Wang JW, Zhang SZ. Genomewide identification and expression analysis of the ARF gene family in apple. *J Genet*. 2014;93(3):785–97. <https://doi.org/10.1007/s12041-014-0462-0>.
30. Zhou Q, Chen T, Wang Y, Liu Z. The development of 204 novel EST-SSRs and their use for genetic diversity analyses in cultivated alfalfa. *Biochem Syst Ecol*. 2014;57:227–30.
31. Luo D, Wu Y, Liu J, Zhou Q, Liu W, Wang Y, Yang Q, Wang Z, Liu Z. Comparative transcriptomic and physiological analyses of *Medicago sativa* L. indicates that multiple regulatory networks are activated during continuous ABA treatment. *Int J Mol Sci*. 2018;20(1): 47. <https://doi.org/10.3390/ijms20010047>.
32. Luo D, Zhou Q, Wu Y, Chai X, Liu W, Wang Y, Yang Q, Wang Z, Liu Z. Full-length transcript sequencing and comparative transcriptomic analysis to evaluate the contribution of osmotic and ionic stress components towards salinity tolerance in the roots of cultivated alfalfa (*Medicago sativa* L.). *BMC Plant Biol*. 2019;19(1):32. <https://doi.org/10.1186/s12870-019-1630-4>.
33. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant*. 2020;13(8):1194–202. <https://doi.org/10.1016/j.molp.2020.06.009>.
34. Roth C, Liberles DA. A systematic search for positive selection in higher plants (Embryophytes). *BMC Plant Biol*. 2006;6(1): 12. <https://doi.org/10.1186/1471-2229-6-12>.
35. Cho H, Ryu H, Rho S, Hill K, Smith S, Audenaert D, Park J, Han S, Beeckman T, Bennett MJ, Cho H, Ryu H, Rho S, Hill K, Smith S, Audenaert D, Park J, Han S, Beeckman T, Bennett MJ, Hwang D, De Smet I, Hwang I. A secreted peptide acts on BIN2-mediated phosphorylation of ARFs to potentiate auxin response during lateral root development. *Nat Cell Biol*. 2014;16(1):66–76. <https://doi.org/10.1038/ncb2893>.
36. Tang Y, Bao X, Liu K, Wang J, Zhang J, Feng Y, Wang Y, Lin L, Feng J, Li C. Genome-wide identification and expression profiling of the auxin response factor (ARF) gene family in physic nut. *PLoS ONE*. 2018;13(8):e0201024.
37. Jain M, Khurana JP. Transcript profiling reveals diverse roles of auxin-responsive genes during reproductive development and abiotic stress in rice. *FEBS J*. 2009;276(11):3148–62. <https://doi.org/10.1111/j.1742-4658.2009.07033.x>.
38. Reyes-Chin-Wo S, Wang Z, Yang X, Kozik A, Arikiti S, Song C, Xia L, Froenicke L, Lavelle DO, Truco MJ, et al. Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. *Nat Commun*. 2017;8:14953.
39. Wang D, Pei K, Fu Y, Sun Z, Li S, Liu H, Tang K, Han B, Tao Y. Genome-wide analysis of the auxin response factors (ARF) gene family in rice (*Oryza sativa*). *Gene*. 2007;394(1–2):13–24.
40. Ha CV, Le DT, Nishiyama R, Watanabe Y, Suliman S, Tran UT, Mochida K, Dong NV, Yamaguchi-Shinozaki K, Shinozaki K et al: The auxin response factor transcription factor family in soybean: genome-wide identification and expression analyses during development and water stress. *DNA research* 2013, 20(5):511–524.
41. Xing H, Pudake RN, Guo G, Xing G, Hu Z, Zhang Y, Sun Q, Ni Z. Genome-wide identification and expression profiling of auxin response factor (ARF) gene family in maize. *BMC Genomics*. 2011;12(1): 178. <https://doi.org/10.1186/1471-2164-12-178>.
42. Liu R, Guo Z, Lu S. Genome-wide identification and expression analysis of the Aux/IAA and auxin response factor gene family in *Medicago truncatula*. *Int J Mol Sci*. 2021;22(19): 10494. <https://doi.org/10.3390/ijms221910494>.
43. Schilling S, Kennedy A, Pan S, Jermin LS, Melzer R. Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. *New Phytol*. 2020;225(1):511–29.
44. Jiao Y. Double the genome, double the fun: genome duplications in angiosperms. *Mol Plant*. 2018;11(3):357–8. <https://doi.org/10.1016/j.molp.2018.02.009>.
45. Wu S, Han B, Jiao Y. Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol Plant*. 2020;13(1):59–71. <https://doi.org/10.1016/j.molp.2019.10.012>.
46. Zouine M, Fu Y, Chateigner-Boutin AL, Mila I, Frasse P, Wang H, Audran C, Roustan JP, Bouzayen M. Characterization of the tomato ARF gene family uncovers a multi-levels post-transcriptional regulation including alternative splicing. *PLoS One*. 2014;9(1): e84203. <https://doi.org/10.1371/journal.pone.0084203>.
47. Wang JW, Wang LJ, Mao YB, Cai WJ, Xue HW, Chen XY. Control of root cap formation by MicroRNA-targeted auxin response factors in *Arabidopsis*. *Plant Cell*. 2005;17(8):2204–16. <https://doi.org/10.1105/tpc.105.033076>.
48. Yang J, Tian L, Sun MX, Huang XY, Zhu J, Guan YF, Jia QS, Yang ZN. AUXIN RESPONSE FACTOR17 is essential for pollen wall pattern formation in *Arabidopsis*. *Plant Physiol*. 2013;162(2):720–31. <https://doi.org/10.1104/pp.113.214940>.
49. Tiwari SB, Hagen G, Guilfoyle T. The roles of auxin response factor domains in auxin-responsive transcription. *Plant Cell*. 2003;15(2):533–43. <https://doi.org/10.1105/tpc.008417>.
50. Wang L, Hua D, He J, Duan Y, Chen Z, Hong X, Gong Z. Auxin Response Factor2 (ARF2) and its regulated homeodomain gene HB33 mediate abscisic acid response in *Arabidopsis*. *PLoS Genet*. 2011;7(7): e1002172. <https://doi.org/10.1371/journal.pgen.1002172>.
51. Shen C, Yue R, Sun T, Zhang L, Xu L, Tie S, Wang H, Yang Y. Genome-wide identification and expression analysis of auxin response factor gene family in *Medicago truncatula*. *Front Plant Sci*. 2015;6:73.
52. Tombuloglu H. Genome-wide analysis of the auxin response factors (ARF) gene family in barley (*Hordeum vulgare* L.). *J Plant Biochem Biotechnol*. 2019;28(1):14–24. <https://doi.org/10.1007/s13562-018-0458-6>.
53. Liu Y, Guan X, Liu S, Yang M, Ren J, Guo M, Huang Z, Zhang Y. Genome-wide identification and analysis of TCP transcription factors involved in the formation of leafy head in Chinese cabbage. *Int J Mol Sci*. 2018;19(3): 847. <https://doi.org/10.3390/ijms19030847>.
54. Yao X, Ma H, Wang J, Zhang D. Genome-wide comparative analysis and expression pattern of TCP gene families in *Arabidopsis thaliana* and *Oryza sativa*. *J Integr Plant Biol*. 2007;49(006):885–97.

55. Zhou B, Gao X, Zhao F. Integration of mRNA and miRNA analysis reveals the post-transcriptional regulation of salt stress response in *Hemerocallis fulva*. *Int J Mol Sci*. 2023;24(8): 7290. <https://doi.org/10.3390/ijms24087290>.
56. Su L, Xu M, Zhang J, Wang Y, Lei Y, Li Q. Genome-wide identification of auxin response factor (ARF) family in kiwifruit (*Actinidia chinensis*) and analysis of their inducible involvements in abiotic stresses. *Physiol Mol Biol Plants*. 2021;27(6):1261–76.
57. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res*. 2012;40(Web Server issue):W597–603.
58. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer EL L, Hirsh L, Paladin L, Piovesan D, Tosatto SC E, Finn RD. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47(D1):D427–d432. <https://doi.org/10.1093/nar/gky995>.
59. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 2008;36(10):3420–35. <https://doi.org/10.1093/nar/gkn176>.
60. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27–30.
61. Newman L, Duffus A, Lee C. Using the free program MEGA to build phylogenetic trees from molecular data. *Am Biol Teacher*. 2016;78:608–12.
62. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinf (Oxford England)*. 2009;25(15):1972–3.
63. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32(1):268–74. <https://doi.org/10.1093/molbev/msu300>.
64. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018;35(2):518–22. <https://doi.org/10.1093/molbev/msx281>.
65. Mei M, Ai W, Liu L, Xu X, Lu X. Genome-wide identification of the auxin response factor (ARF) gene family in *Magnolia sieboldii* and functional analysis of MsARF5. *Front Plant Sci*. 2022;13: 958816. <https://doi.org/10.3389/fpls.2022.958816>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

