Genome wide identification of ARF transcription factor gene family and their expression analysis in sweet potato

By

Isaac Pratt

July 2021

Director of Thesis: Baohong Zhang

Department of Biology

ABSTRACT

Auxin response factors (ARFs) are a family of transcription factors that play an important role of auxin regulation through their binding with auxin response elements. ARF genes are represented by a large multigene family in plants; however, to our knowledge, the ARF gene family has not been well studied and characterized in sweet potatoes. In this study, a total of 26 possible ItfARF genes were identified in *Ipomea trifida*; however, ItfARF1a and ItfARF2a were found to be a single gene, resulting in a total of 25 ItfARF genes. Firstly, a comprehensive bioinformatics analysis of the ItfARF gene family was given, including conserved motifs, chromosomal locations, phylogenetic relationships, and a protein data breakdown. Furthermore, the expression patterns of ItfARF genes were analyzed within the storage roots and normal roots, at an early stage of development. ItfARF16b & ItfARF16c were both highly expressed in the storage root, with minimal to no expression in the normal root. ItfARF6a and ItfARF10a exhibited higher expression in the normal root, but not in the storage root. Subsequently, ItfARF1a, ItfARF2b, ItfARF3a, ItfARF6b, ItfARF8a, ItfARF8b and ItfARF10b were expressed in both root types with moderate to high expression for each. All ten of these ARF genes and their prominent expression signify their importance within the development of each respective root type. This study provides

comprehensive information regarding the ARF family in sweet potatoes, which will be useful for future research to discover further functional verification of these ItfARF genes.

Genome wide identification of ARF transcription factor gene family and their expression analysis in sweet potato

A Thesis

Presented to the Faculty of the Department of Biology

East Carolina University

In Partial Fulfillment of the Requirements for the Degree

Master of Science in Biology

By

Isaac Pratt

July 2021

Genome wide identification of ARF transcription factor gene family and their expression analysis in sweet potato

By

Isaac Pratt

APPROVED BY:

Director of Thesis _____

Baohong Zhang, PhD

Committee Member _____

Xiaoping Pan, PhD

Committee Member _____

John Stiller, PhD

Chair of the Department of Biology _____

David Chalcraft, PhD

Dean of the Graduate School _____

Paul J. Gemperline, PhD

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Indole-3-acetic acid (IAA), one type of auxin, is a plant growth hormone that elicits developmental growth throughout the plant's life cycle, including seed germination, vascular tissue formation, and reproductive and vegetative growth [1, 5]. These Auxin plant hormones work in conjunction with Auxin response factors (ARF), a set of transcription factors. When ARFs are bound to auxin response elements (AuxRE, TGTCTC), they work to activate or repress auxin gene expression and regulation [6]. ARF proteins can be broken down into three domain subunits: an amino-terminal DNA-binding domain (DBD), a conserved carboxy-terminal dimerization domain (CTD), and a non-conserved middle domain (MD) [2]. The DBD for ARF's are found in the N-terminal region and bind specifically to the AuxRE TGTCTC in promoters for the regulation of auxin gene expression [8]. Carboxy-terminal dimerization domains facilitate both homo- and hetero-dimerization through protein interactions, relating themselves to the domain III and IV in Aux/IAA proteins [8]. The MD has a double function, acting as either a repression domain (RD) or an activation domain (AD) [7,9]. ARF proteins that have a MD rich in glutamine (Q) are constituted as an activator, this can be seen in AtARF5,6,7,8, and 19 [10, 11], while ARF proteins that have a MD rich in Proline (P), Serine (S), and Threonine (T) are considered repressors, these repressors were found in AtARF1, 2, 3, 4 and 9 [9].ARF proteins have been identified in a variety of plants, 22 ARF genes and one pseudogene were found in *Arabidopsis* and 25 ARF genes were found in rice [7, 12]. Subsequently, there have been more ARFs identified in a multitude of other plant species, such as tomato (*Solanum lycopersicum*) [3], sorghum (*Sorghum bicolor*) [16], soybean (*Glycine max*) [17,18], maize (*Zea maize)* [13, 19, 20], Chinese cabbage (*Brassica rapa*) [21], cotton (*Gossypium raimondii*) [22], sweet orange (*Citrus sinensis*) [14], and alfalfa (*Medicago truncatula)* [15]. However, with all these

identifications of auxin response factors, there has not been the identification of auxin response factors in sweet potato.

The sweet potato (*Ipomoea batatas*) of the family Convolvulaceae is a highly desired food crop throughout the world and has been ranked the 7th largest crop. It serves as a staple diet in developing countries for its overall high nutritional and caloric values, ability to grow in most climates and conditions and having a high production yield. The sweet potato (*Ipomoea batatas*) is a hexaploid with 90 chromosomes, making genomic research on the crop highly complicated [26]. However, there is an alternative to performing genomic research on sweet potatoes, and that is to use a diploid relative. The diploid *Ipomoea trifida* is a relative to the hexaploid *I. batatas* and is a model species of genomic research due to its small genome size and number of chromosomes [27, 28, 29]. In this study, all the potential ARF transcription factors from the *I. trifida* genome were identified using various bioinformatics tools. ItfARF proteins, once identified, underwent a phylogenetic analysis where they were compared among themselves, and then amongst the model plant species: *Arabidopsis, O. sativa,* and *G. raimondii.* Finally, the expression profiles of each ItfARF gene was analyzed using tissue samples of both the normal and storage roots of sweet potato plant *I. trifida*.

Chapter 2: Materials and Methods

Identification of ARF genes in *I. trifida*

The genome sequence of *Ipomoea trifida* was downloaded from the Sweet potato genomic resource database (http://sweetpotato.plantbiology.msu.edu/) [39]. For the identification of the genomic members of the ARF gene family, both HMMER 3.3.1 (http://hmmer.org/download.html) and Pfam 33.1 (http://pfam.xfam.org/) were used to search whole genome sequences with Pfam's ARF (PF06507). After identification of the ItfARF gene family, information on the ItfARF's genome location, protein and CDS length, were obtained along with the gene, protein and CDS sequences from the Sweet potato database. In addition, Expasy (https://web.expasy.org/compute_pi/) was used to compute the isoelectric point (PI) and the molecular weight (Mw) of the ItfARF proteins. To analyze the identified ItfARF genes for conserved motifs, the protein sequences were examined using the software MEME (Multiple Expectation maximizations for Motif Elicitation) (https://meme-suite.org/meme/tools/meme). The sequence search options were as follows: motif distribution among sequences was zero to one occurrence, the motif width range was from 6-50 amino acids, and the maximum motifs per sequences was 20.

Phylogenetic analysis of ARF genes

To better understand the relationship and homology between the sequences, a phylogenetic analysis was performed using the protein sequence data [4]. For this analysis, the ItfARF protein sequences were analyzed with Molecular Evolutionary Genetics Analysis (MEGA-X 10.2) using the neighbor-joining method and its built-in sequence alignment tool, Clustal-W [25]. A second phylogenetic analysis was conducted to compare the evolutionary

relationship to different plant species. The ARF protein sequences from *Arabidopsis*, *O. sativa*, and *G. raimondii* were also used to construct the tree.

Sweet potato culture, tissue collection and RNA extraction

The sweet potato (*Ipomoea batatas*) were cultivated in the Greenhouse with normal agronomic practices. Both storage root and normal root samples were collected from 5-weeks grown sweet potato plants (Fig. 6). Root tissues were quickly sampled from the plants, immediately frozen in liquid nitrogen and stored at -80 ℃ until RNA extraction. At least six biological replicates were collected for each type of root. Total RNAs were extracted from each root sample by using mirVana^{TM} miRNA isolation Kit (Ambion, Austin TX) as performed in previous studies [23, 24]. Briefly, the tissues were ground into a fine powder in a mortar and pestle, and then transferred into a 2ml centrifuge tube with the Lysis/Binding buffer. For the normal roots, 400 μL of the Lysis/Binding buffer was used, and for the storage roots, 700 μL buffer, this is because the storage root contained higher levels of starch that absorbed large quantities of liquid. The samples were then sonicated for 15-20 seconds on ice. Post sonication, each sample was inverted for accurate mixture of the Lysis buffer, and the plant material. After a 10-minute ice bath, 400 μL of Acid-Phenol: Chloroform was added for the separation of RNA from its cellular components. Carefully following the manufactures protocols, consisting of aqueous phase extraction and several wash cycles, 95 ℃ nuclease-free water was added to the filter cartridge medium, for the conclusion of RNA isolation. To test the concentration and purity of the RNA isolated, Nanodrop ND-1000 was used. Samples NR1, NR2, NR3, SR1, SR2, and SR3 were all collected, each having a ng/μL concentration in the range of 200-350 ng/μL.

qRT-PCR

The RNA of each plant tissues sample was prepared for reverse transcription with specific primers designed from NCBI primer design tool. Reverse transcription was then performed by following the instructions of the TaqMan® MicroRNA Reverse Transcription Kit (Applied Biosystems, CA, USA). This kit included Multiscribe™ Reverse transcriptase (50 U/μL), dNTPs with dTTP (100 nm), Reverse transcription buffer (10x), RNase inhibitor (20 μ/μL), nuclease-free water and total RNAs; the volume for each reaction was 15 μL. Post reaction completion, 150 μL of nuclease-free water was added to the products and stored at -20 ℃ until ready for qRT-PCR. qRT-PCR was performed on a 96 well plate within the 7300 Fast Real-Time PCR System (Applied Biosystems, CA, USA). The samples ran for each replicate went as followed: all ItfARF genes and 2 reference genes (EF1α & UBC). SYBR Green was used to analyze gene expression within the qRT-PCR system. A total of 6 reactions were ran, three biological replicates for each root type, and each biological replicate had three technical replicates. The reaction temperature program settings were as follows, 10 minutes at 95 ℃, with 40 cycles of 15 seconds at 95 ℃, and 60 seconds at 60 ℃.

Data and statistical analysis

Statistical analysis was performed after obtaining sample triplicate Ct values. Elongation factor 1α (EF1α) and ubiquitin-c (UBC) served as the reference genes, and the average of each reference gene ΔCt values were combined and subtracted from all other ItfARF genes ΔCt values to obtain a second normalization. Differentially expressed genes were discovered using statistical calculation, $p < 0.05$. Fold change was then calculated for each gene using the formula: $2^{-(\Delta\Delta Ct)}$. A

hierarchical clustering analysis was performed using Multi Experiment Viewer (MeV) to create a heat-map of gene expression for all the ItfARF genes.

Chapter 3: Results

Identification and sequence analysis of ARF genes in *I. trifida*

To identify the ARF transcription factor genes in *I. trifida*, the ARF protein domain (PF00025) was used to blast search against the *I. trifida* genome. The first search led to a total of 282 possible protein sequences in *I. trifida*. After eliminating the redundant sequences only 26 ItfARF sequences remained. These 26 sequences were then compared and categorized into 13 different ARF gene subfamilies, including ItfARF1, ItfARF2, ItfARF3, ItfARF4, ItfARF5, ItfARF6, ItfARF8, ItfARF9, ItfARF10 ItfARF11, ItfARF16, ItfARF18, and ItfARF19 named so after their similarities to *Arabidopsis*. These 26 sequences are listed in Table 1. The average number of exons in the gene sequences was 10, having a similar distribution to the *Arabidopsis* ARF genes. The length of the CDS varied from 897bp (ItfARF1a) all the way up to 3375bp (ItfARF19a). The 26 possible genes all produced adequate proteins ranging from 298 to 1124 amino acids in length. The predicted MW of the ItfARF were as low as 33.4 kDa and as high as 124.164 kDa. The PI value of the ItfARF genes were from 4.99 to 9.44, most of these values fell in the range of 5-7 suggesting they encode weak acid proteins, while those few that ranged from 7.56-9.44 encode for weak basic proteins.

The 26 identified ARF genes were located among the 15 chromosomal pairs with the exclusion of Chr. 8, 13, and 14 as no ARF genes were mapped there. A majority of the genes stacked onto a select few chromosomes (Fig. 1), having five genes on chromosome 10 (19.2%), four genes on chromosome 6, three genes each on chromosomes 2, and 4, two genes each on chromosomes 1, 7, 9, and 11, and then one gene on chromosomes 3, 5, 12, and 15. The ARF genes, ItfARF8a & ItfARF8b and ItfARF1b & ItfARF1c are the only two sets of duplicate genes that met the 80% sequence similarity for each of their respective nucleotide sequences.

Conserved domains and motif analysis of ItfARF proteins

For the prediction of protein function, the analysis of both domains and subdomains was apparent [2]. Out of the twenty-six ItfARF proteins, there were six among them that lacked at least one of the three typical domains, five of those six were lacking the Aux/IAA CTD domain (ItfARF1a, 3a, 3b, 11, and 16c). Only one sequence was missing a B3 DNA binding domain ItfARF2a, and a single different sequence, ItfARF1a was lacking an ARF domain (Table 1). Regarding the domains, the MD gave insight into the sequences ability to either be a transcriptional activator or a repressor. The protein sequences that were rich in (Q) have been regarded as activators ItfARF5, 6a, 6b, 8a, 8b, 19a and 19b, while the other protein sequences whose MD were rich in (P, S, and T) are regarded as transcriptional repressors. The conserved motif analysis allowed for further confirmation of these notions. As shown in Fig. 2, motifs 1, and 2 correspond to a DNA-binding domain; motifs 8, 9, 10, and 11 correspond to an ARF domain, while motifs 17, 18, 19, and 20 correspond to an Aux/IAA domain. lastly, all seven ItfARF transcriptional activators lack motif 16 while still maintaining motifs 18, 19, and 20, suggesting that this combination of motifs signify a transcriptional activator. Cross comparing Table 1 and Figure 2 we can see that these motifs correlate well with each other. Each protein sequence had a variable number of motifs, but the motifs for each domain were conserved with minimal variation preceding and succeeding them.

Phylogenetic analysis of ItfARFs

To understand the evolutionary and phylogenetic relationship of the ItfARF genes at the molecular level, the phylogenetic analysis began with the alignment of the potential ItfARF

protein sequences using Clustal-W in MEGA-X. A single ItfARF protein sequence, ItfARF2a had to be removed before the analysis was complete, due to its missing domain. The lack of the b3-DNA binding domain caused them ItfARF2a to root to multiple ARF genes across all four species. For classification specifics, Class III held each ItfARF gene that was considered an activator (Fig. 3). Within the neighbor-joint phylogenetic tree there was a total of nine sister pairs, and two sister triplets. Those ItfARF genes that were in a sister pair or triplet, displayed similar gene structure in terms of exon/intron distribution and protein length. Those proteins that were not sister paired and thus had a different exon/intron distribution, suggesting functional diversity among the genes.

For better understanding the functional and evolutionary relationship of the ARF gene family, an unrooted phylogenetic tree was created by multiple sequence alignment in MEGA-X for all selected plant species, including sweet potato. This analysis used the 25 ItfARF protein sequences and all the ARF proteins from *Arabidopsis*, rice, and cotton. Based on the neighbor-joining phylogenetic tree, the 89 ARF transcription factors fall into five major classifications ranging from I to IV, the classes were derived from their phylogenetic relationship. Out of all the classes, Class III was the largest having 27 ARF members holding 29.2% of the total ARF genes analyzed. Class III was comprised of three subgroups: IIIa, IIIb, and IIIc, these contained 4, 9, and 12 ARF genes, respectively. Classes IIIb and IIIc contained ItfARFs with a Q-rich MD, which included ItfARF6a, 6b, 8a, 8b for IIIc and 19a, and 19b for IIIb. ItfARFs were in each class except for class V, suggesting that the ItfARF gene family arose before the lineage spilt.

Expression profiling of ItfARF genes in root tissues

qRT-PCR was employed to better understand the expression profiles of all ItfARF genes within sweet potato normal roots and storage roots. Through the gene expression analysis the differeing functions of all ItfARF genes was apparent. There were some ItfARF genes with similar gene expression patterns between the two root types, while other genes showed clear differences, with higher expression in one root type over the other (Fig. 5). For example, ItfARF1a, ItfARF8a and ItfARF8b were highly expressed in both normal roots and storage roots. ItfARF2b, ItfARF3a, ItfARF6b, and ItfARF10b, are all moderately expressed across both tissue types, suggesting that they may be primarly expressed under specific conditions. Only two of the 26 tested ItfARF genes were more highly expressed in the normal root than that in the storage root, those genes were: ItfARF6a and ItfARF10a. There were some genes that had little to no expression at all such as: ItfARF2a, ItfARF2c, ItfARF3b, ItfARF9b, ItfARF16a, ItfARF18, and ItfARF19a, these suggest The rest of the ItfARF genes were expressed more highly in the storage roots as opposed to the normal roots, those genes being: ItfARF1b, ItfARF1c, ItfARF4a, ItfARF4b, ItfARF5, ItfARF9a, ItfARF16b, ItfARF16c, and ItfARF19b. Out of those nine genes, ItfARF4a, ItfARF5, ItfARF9a, ItfARF16b, ItfARF16c, and ItfARF19b had the highest level of expression in the storage roots.

Chapter 4: Discussion

ARF genes in *I. trifida*

ARF are responsible for the regulation of multiple plant processes, including lateral root formation, fruit initiation, apical dominance, and cellular senescence [32]. ARF genes have been continuously reported throughout many different plant species, but they have not been reported within sweet potatoes, an emerging biofuel crop. In the present study, 25 ARF transcription factor genes belonging to 13 ARF gene families were identified and characterized in *I. trifida*. Overall, the number of ItfARF genes surpassed those in other plant species, such as *Arabidopsis* and rice; however, when categorized into gene families, *I. trifida* had the smallest amount of ARF gene family representation at 13 [2]. This gives rise to the question: what happened in the evolutionary process between *Arabidopsis*, rice, and *I. trifida* that caused a disparity in total gene number? The gene family presence might have been decreased throughout the evolution of the species. This can be seen in (Fig. 4) where there are no ItfARF, OsARF, or GrARF genes in classification V, possibly meaning that those AtARF genes were derived from a singular AtARF gene [22]. There is also the notion that *I. trifida* has gone through genomic rearrangements due to gene duplication, giving rise to a higher quantity of ItfARF genes. Genes can also lose function due to genomic rearrangements, such as ItfARF1a as it does not contain an ARF domain or a CTD domain. With the lack of these two domains, ItfARF1a cannot be considered an ARF gene. However, ItfARF2a lacks a B3-DNA binding domain, and both ItfARF1a and ItfARF2a are located directly next to each other on chromosome 6 suggesting that these two ItfARF genes form a singular gene, decreasing the total ItfARF gene count to 25. An important note on the reduction of gene family representation in *I. trifida* regards the nomenclature system in place. All

the ItfARF genes were named according to the most accurate sequence similarity of ARF proteins in *Arabidopsis*.

Phylogenetic characterization and function of ARF gene family

The ARF transcription factor orthologous relationships can be seen throughout the phylogenetic analysis. Those relationships between *Arabidopsis* and *I.* trifida, are almost constant for each ItfARF gene, save ItfARF1a, and ItfARF11. This implies that the orthologous classes I-IV are conserved, but it does not imply that the function of the ARF genes in *I. trifida* are directly related to the *Arabidopsis* ARF gene functions. The ARF genes across species may be functionally different, either because of plant specific evolutions or because of gene duplication. Regarding previous studies, ARF double mutant genes typically have a more pronounced phenotype than their single mutant counterpart, implying closely related ItfARF genes may be functionally redundant for plant growth [36, 40]. As seen in *Arabidopsis*, ARF2 regulatory functions played multiple roles relating to plant aging such as, flower initiation, rosette leaf senescence, and floral organ abscissions. ARF1 mutations were then seen to enhance those ARF2 regulatory phenotypes, suggesting that ARF1 is partially redundant with ARF2 [30]. ARF7 and ARF19 mutations had functional redundancies between both genes, that impacted and impaired, lateral root formation and abnormal gravitropism in both the hypocotyl and the root [33, 35, 36]. ARF4 double mutants enhance ARF3 mutants regarding ARF3's role in gynoecium abaxial identity [37]. ARF6 and ARF8 both regulate stamen and gynoecium maturation. In single mutations of ARF6 and ARF8, flower maturation was delayed, reducing fertility, while the double mutants were completely infertile [38]. ARF8 has also been seen to affect hypocotyl elongation and root growth habit regarding light sensitivity, indicating a possible dominant role

in the development of *I. trifida* [34]. ARF10 and ARF16 double mutants display abnormal root cap gravitropism and root cap developmental defects [31]. Each of these mutants in the phylogenetic tree, all fall together under their classes: ARF1 & ARF2, ARF3 & ARF4, ARF6 & ARF8, and ARF10 & ARF16 are all in classes, I, II, III, and IV respectively. This mutation classification relationship suggests that functionally redundant genes have close evolutionary relationships.

For the phylogenetic analysis in this study, there were nine sister pairs and two sister triplets of ItfARF genes identified. All the nine sister pairs were each distributed across different chromosomes. Additionally, both sets of sister triplets were located across different chromosomes (Fig. 1). For the ItfARF chromosomal distribution, it should be noted that no ARF genes were located on chromosomes, 8, 13, or 14, whereas in *O. sativa* there is an ARF gene on chromosome 8, suggesting that the ARF gene might have been lost or transferred to a different chromosome during the species evolutionary history.

Gene expression of ItfARF genes in root tissues

Protein data such as motifs and domains are useful for the prediction of novel gene function. Through sequence analysis of all the ItfARF protein sequences it was found that only seven of the 25 genes contained a middle region rich in Q, meaning these seven genes most likely act as activators. These genes those genes being ItfARF5, ItfARF6a, ItfARF6b**,** ItfARF8a, ItfARF8b, ItfARF19a, and ItfARF19b. The remaining 18 genes rich in SPT and not in Q, are most likely acting as transcriptional repressors. In *Arabidopsis* the ARF genes that act as activators are ARF5, ARF6, ARF7, ARF8, and ARF19. All the activators, for both *Arabidopsis* and *I. trifida* fall into the same phylogenetic classification of class III (Fig.4), suggesting that the

activation genes and domains are conserved through the evolutionary process. The study also showed the root-specific expression profiles of ARF transcription factors in sweet potato. The results showed that the majority of ARF genes were more active in the storage root, with only nine genes in total that had moderate to high expression levels within the normal roots. Out of those nine ARF genes, there were four ItfARF activators, those being ItfARF6a, ItfARF6b, ItfARF8a, and ItfARF8b. Of those four ARF genes, ItfARF6a was the only transcriptional activator that was solely expressed in the normal root; the other three were expressed equally across both root types, this suggests that ItfARF6a is a main activator of the normal root growth. ItfARF8a and ItfARF8b were highly expressed across both root tissue types, suggesting that these two ItfARF genes are the most prevalent activators for the entire sweet potato root system. As for the other three ItfARF gene activators, ItfARF5 and ItfARF19b both have moderately high expression levels in the storage root alone, while ItfARF19a has almost no expression at all, meaning ItfARF19a might be express more in the hypocotyl than the entire root itself, and ItfARF5 and ItfARF19b are activators for storage root development. Relating expression levels to evolutionary history with other species, both ItfARF10 & ItfARF16 gene sets have relative expression regarding the root system of *I. trifida.* ItfARF10a, is almost exclusively expressed in the normal roots, ItfARF10b is shown to have moderate expression across both roots, ItfARF16b and ItfARF16c are expressed in both roots; however, the storage root expression levels are much higher. This relates back to the ARF10 and ARF16 genes in *Arabidopsis* that affect root cap gravitropism and root cap development [31]. Furthermore, all ItfARF10 genes and ItfARF16 genes were classified together in class IV of both phylogenetic trees, meaning not only are these genes functionally similar, they are evolutionarily similar. These results suggest that ItfARF genes play important roles in both normal and storage root development. Further research should

be conducted to investigate the ItfARF genes for further functional verification of their role in sweet potato root systems, namely those seven ItfARF genes that specifically express in the storage root.

Tables and Figures

Table 1. The characteristics of identified ARF gene families in *I. trifida*

Gene Name	Gene Symbol	CDS length (bp)	Domains	Deduced protein			Chr	Genome location	Exon Number
				Length (aa)	MW (kDA)	PI			
ItfARF1a	itf06g08240	897	B3	298	33.4	7.72	6	10610618 - 10608728	7
ItfARF1b	itf09g04490	1977	B3, ARF, AUX/IAA	658	73.862	5.86	9	2229243 - 2223211	14
ItfARF1c	itf10g21760	1980	B3, ARF, AUX/IAA	659	74.102	5.97	10	22768345 - 22773913	14
ItfARF2a	itf06g08230	1755	ARF, AUX/IAA	584	65.324	5.72	6	10608275 - 10605949	7
ItfARF2b	itf10g13730	2559	B3, ARF, AUX/IAA	852	94.263	6.38	10	16445652 - 16440359	14
ItfARF2c	itf11g10870	2556	B3, ARF, AUX/IAA	851	94.731	6.09	11	6559617 - 6564877	14
ItfARF3a	itf09g07030	2169	B3, ARF	722	79.299	6.57	9	3677006 - 3671593	10
ItfARF3b	itf10g19820	1998	B3, ARF	665	73.034	7.72	10	21532579 - 21536723	10
ItfARF4a	itf04g30540	2466	B3, ARF, AUX/IAA	821	90.894	6.75	4	29721501 - 29715473	12
ItfARF4b	itf07g11450	2412	B3, ARF, AUX/IAA	803	89.081	5.56	7	9997889 - 9991527	12
ItfARF5	itf01g35200	2844	B3, ARF, AUX/IAA	947	105.019	5.2	1	31640623 - 31635815	13
ItfARF6a	itf06g20390	2853	B3, ARF, AUX/IAA	950	105.591	6.33	6	22207636 - 22201198	14
ItfARF6b	itf10g04850	2706	B3, ARF, AUX/IAA	901	99.541	6.01	10	4527118 - 4520349	14
ItfARF8a	itf03g05320	2442	B3, ARF, AUX/IAA	813	90.204	5.62	3	3300707 - 3306419	14
ItfARF8b	itf05g23120	2538	B3, ARF, AUX/IAA	845	93.964	5.82	5	23445845 - 23452495	14
ItfARF9a	itf10g16260	1938	B3, ARF, AUX/IAA	645	72.391	6.69	10	18600817 - 18605381	14
ItfARF9b	itf12g17310	1959	B3, ARF, AUX/IAA	652	73.094	6.31	12	16906880 - 16902499	14
ItfARF10a	itf04g30820	2115	B3, ARF, AUX/IAA	704	77.748	7.83	4	29891980 - 29887654	4
ItfARF10b	itf07g11640	2040	B3, ARF, AUX/IAA	679	74.738	7.56	7	10271997 - 10268425	4
ItfARF11	itf02g08750	1731	B3, ARF	576	64.539	4.99	2	8156302 - 8152920	11
ItfARF16a	itf02g08580	2043	B3, ARF, AUX/IAA	680	75.246	6.53	2	8048900 - 8051867	3
ItfARF16b	itf04g06590	2076	B3, ARF, AUX/IAA	691	75.873	6.56	4	3914771 - 3910977	3
ItfARF16c	itf06g19540	2031	B3, ARF	676	74.815	5.75	6	21612765 - 21615756	2
ItfARF18	itf15g10150	2070	B3, ARF, AUX/IAA	689	76.415	6.22	15	6848850 - 6843635	14
ItfARF19a	itf01g29370	3375	B3, ARF, AUX/IAA	1124	124.164	6.62	1	27916897 - 27923035	14
ItfARF19b	itf02g07020	3180	B3, ARF, AUX/IAA	1059	117.923	6.05	2	7111545 - 7105401	13

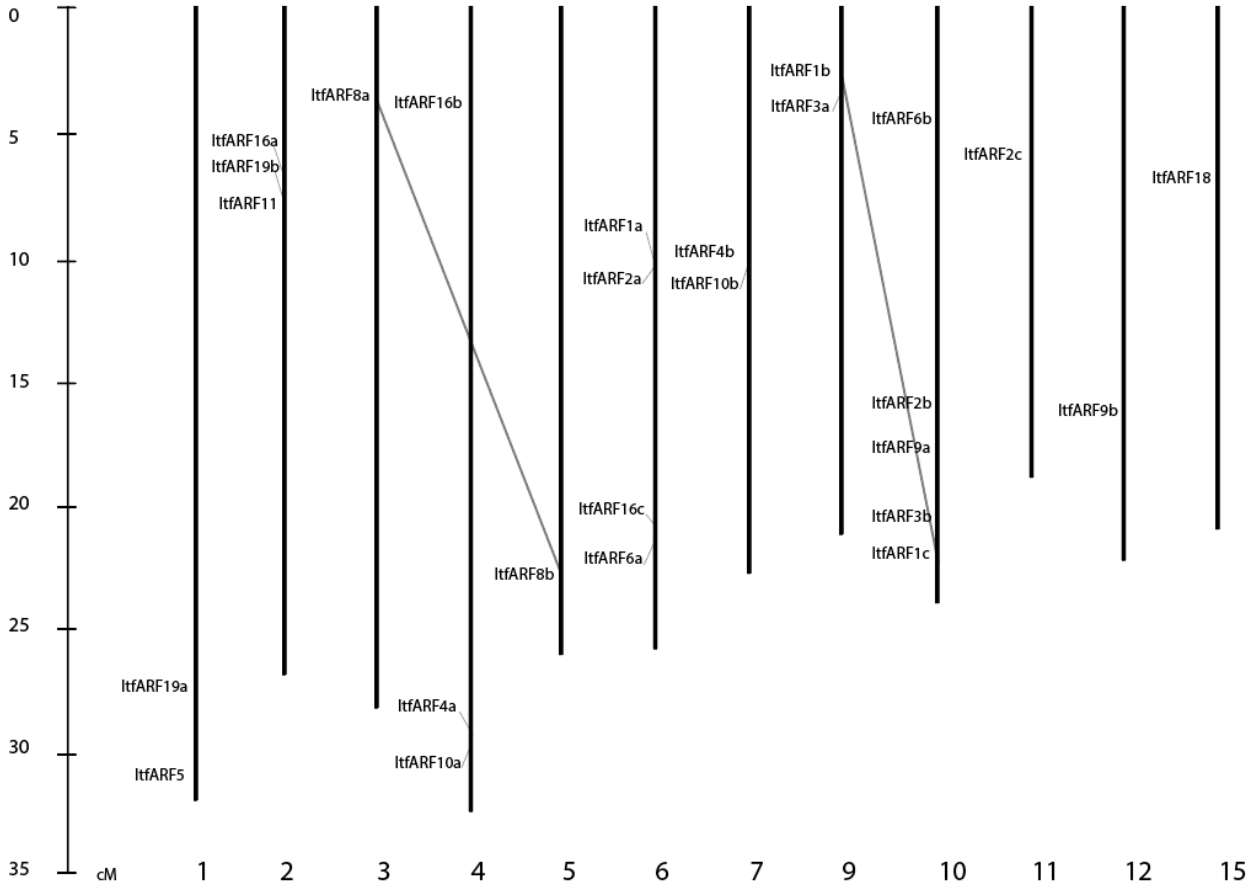

Figure 1. Chromosomal distribution of ARF genes in *I. trifida*, spanning 12 of the 15 chromosomes, gene duplication analysis of ItfARF were presented with a grey line.

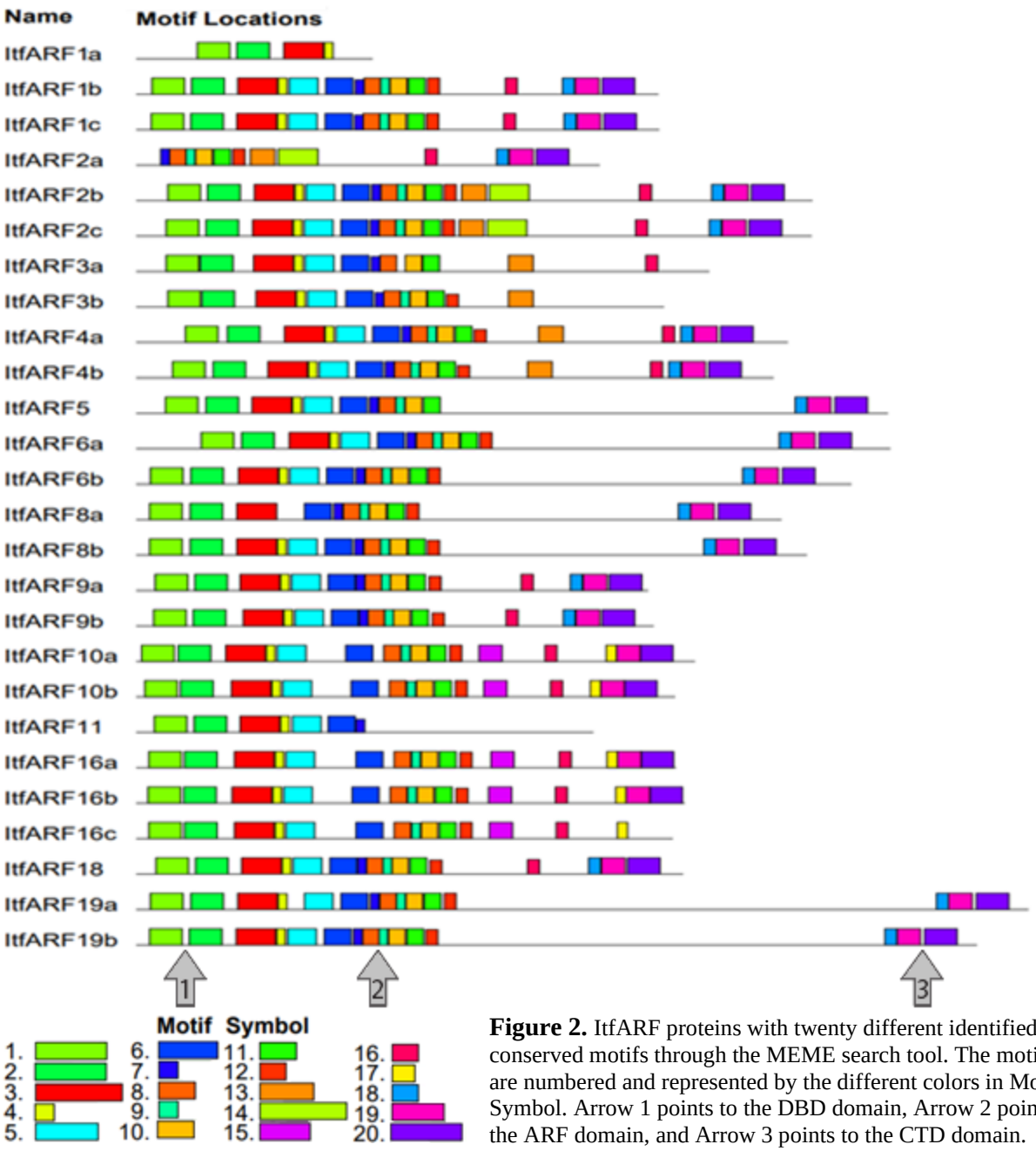

Figure 2. ItfARF proteins with twenty different identified conserved motifs through the MEME search tool. The motifs are numbered and represented by the different colors in Motif Symbol. Arrow 1 points to the DBD domain, Arrow 2 points to the ARF domain, and Arrow 3 points to the CTD domain.

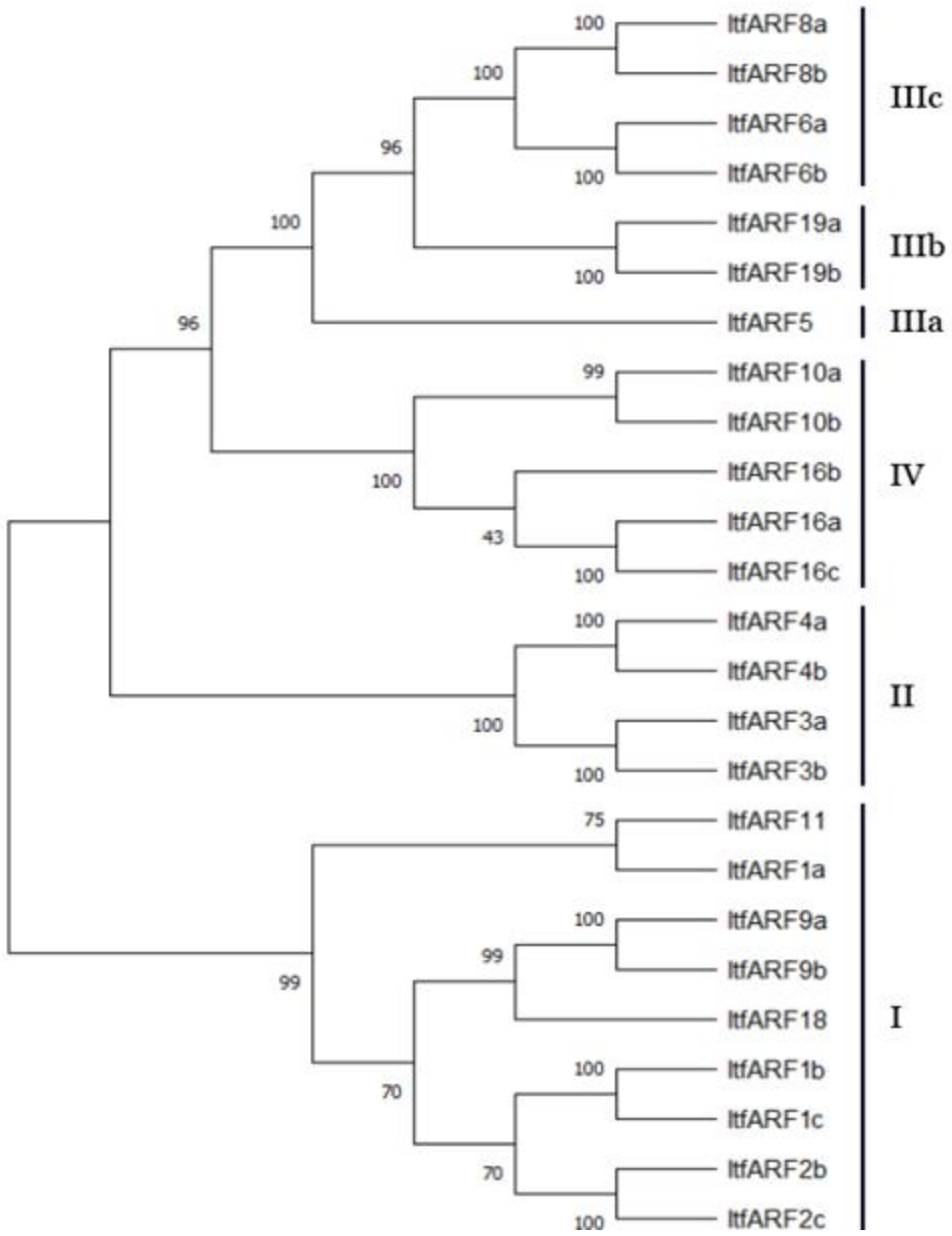

Figure 3. Phylogenetic relationship among the *I.* trifida ARF proteins was generated using MEGA-X. The Tree was made the specifications of the neighbor-joining method with the JTT model and bootstrapped 1000 times with pairwise gap deletions.

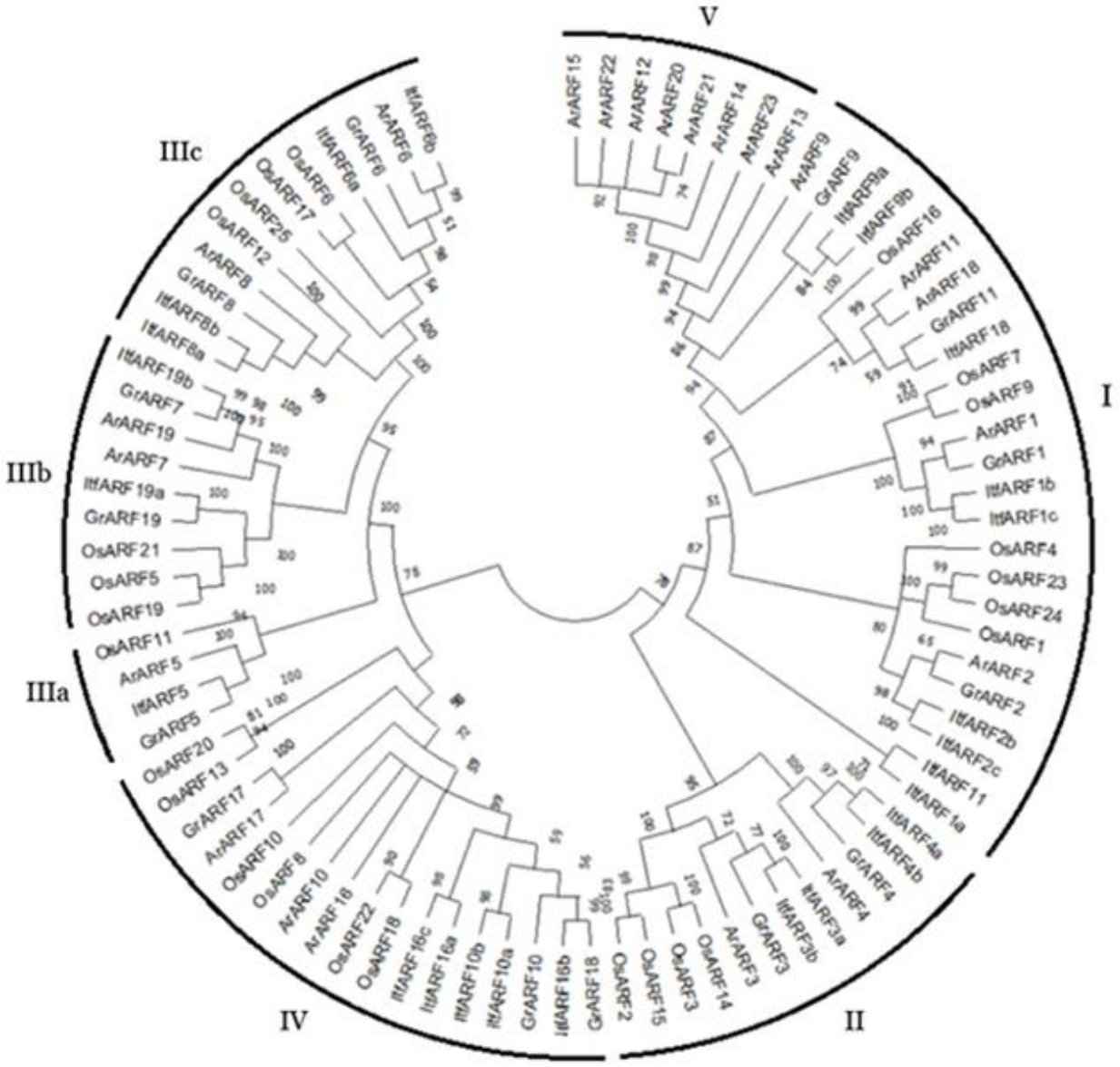

Figure 4. Phylogenetic tree comprised of ARF proteins found in *Arabidopsis*, cotton, *I. trifida*, and rice. This tree was constructed using MEGA-X software using the neighbor-joining method with the JTT model. The parameters were 1000 bootstraps and pairwise gap deletions. The 89 ARF proteins were classified into five classes: I, II, III, IV, and V, and class III was further divided into IIIa, IIIb, and IIIc.

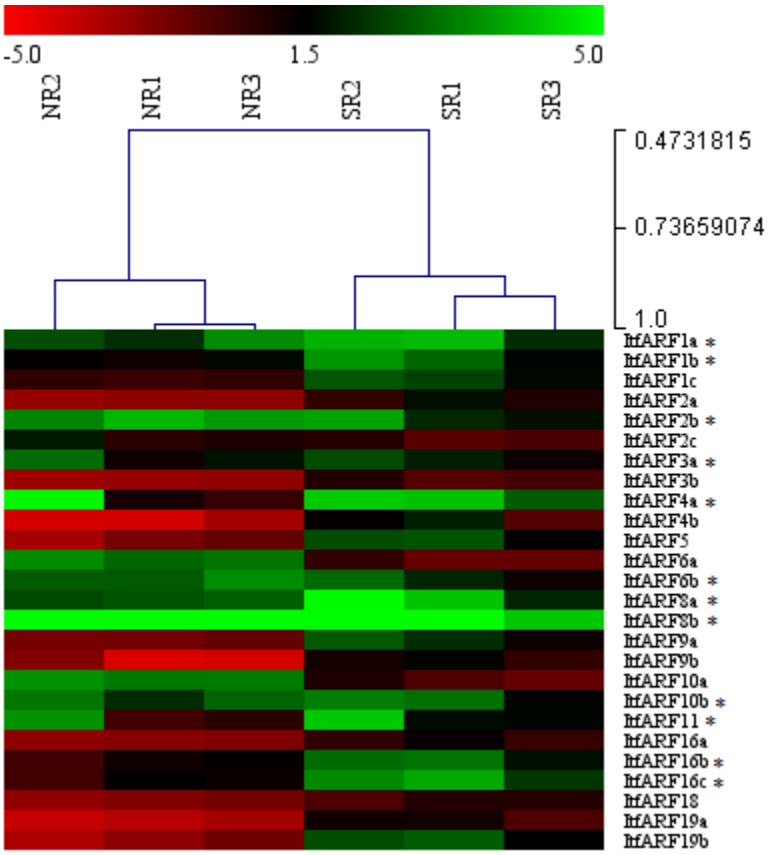

Figure 5. Heatmap of ARF gene expression within all 6 biological samples of both the normal root and the storage root. The tissue samples are named from left to right at the top of the figure, and the names of the 26 genes are directly to the right of the figure. The expression rankings were shown using a range of color: higher gene expression is represented by green; lower expression, red; and median expression, black. *'s denotes which genes are statistically different, * $p < 0.05$.

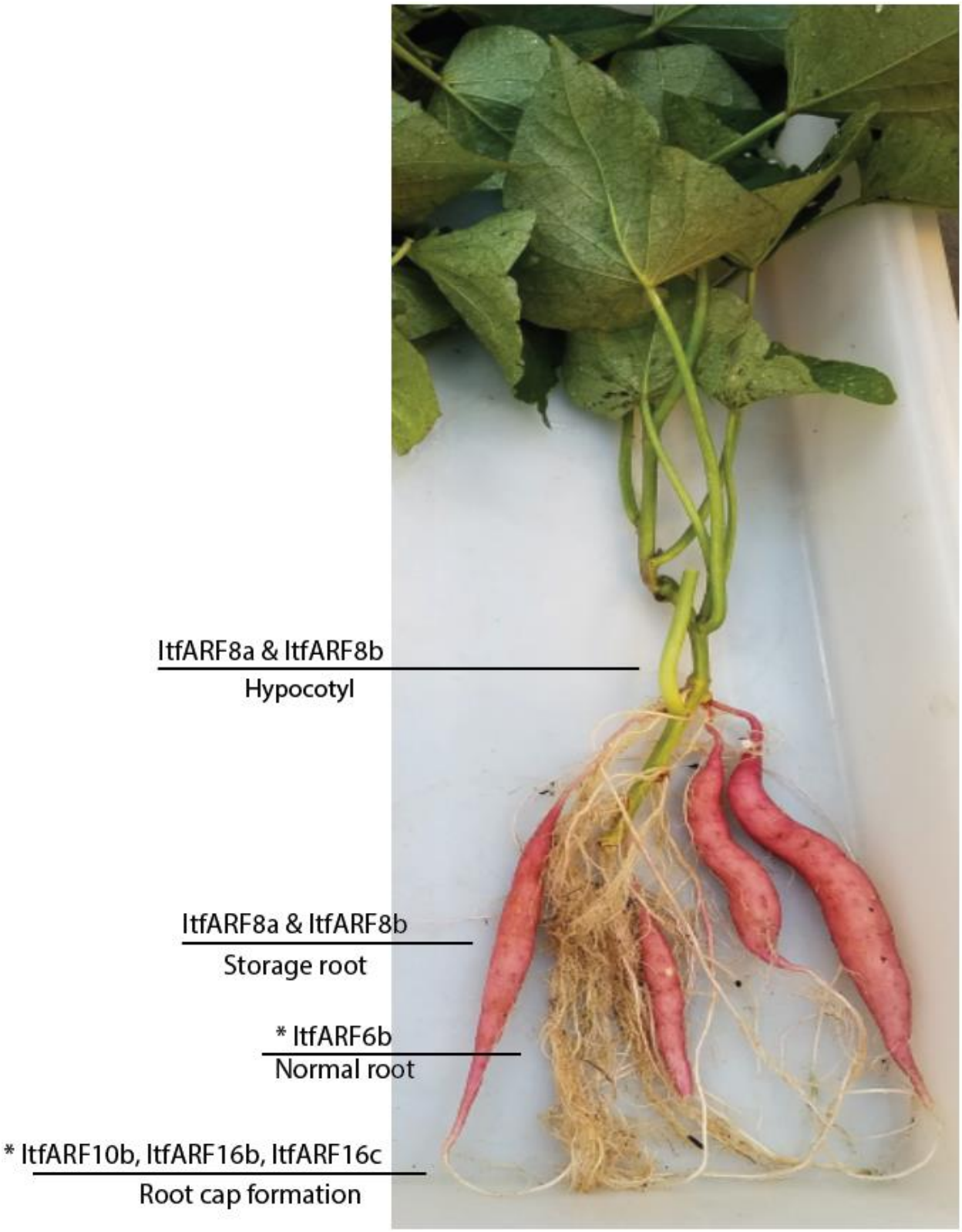

Figure 6. 5-week-old *I. batatas* normal root and storage root samples. Areas labeled correspond to where those genes are most likely expressed. An * denotes ItfARF8a & ItfARF8b, as their function is predicted to affect overall root growth and development.

References

[1] Li, Si-Bei, et al. "A review of auxin response factors (ARFs) in plants." *Frontiers in plant science* 7 (2016): 47.

[2] Guilfoyle, Tom J., and Gretchen Hagen. "Auxin response factors." Current opinion in plant biology 10.5 (2007): 453-460.

[3] Kumar, Rahul, Akhilesh K. Tyagi, and Arun K. Sharma. "Genome-wide analysis of auxin response factor (ARF) gene family from tomato and analysis of their role in flower and fruit development." *Molecular Genetics and Genomics* 285.3 (2011): 245-260.

[4] Barry G. Hall, Building Phylogenetic Trees from Molecular Data with MEGA, *Molecular Biology and Evolution*, Volume 30, Issue 5, May 2013, Pages 1229–1235.

[5] 5, F. W., & Thimann, K. V. (1937). Phytohormones. *Phytohormones.*

[6] Ulmasov, T., Murfett, J., Hagen, G., & Guilfoyle, T. J. (1997). Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *The Plant Cell*, *9*(11), 1963-1971.

[7] Ulmasov, T., Hagen, G., & Guilfoyle, T. J. (1999). Dimerization and DNA binding of auxin response factors. *The plant journal*, *19*(3), 309-319.

[8] Ulmasov, T., Hagen, G., & Guilfoyle, T. J. (1997). ARF1, a transcription factor that binds to auxin response elements. *Science*, *276*(5320), 1865-1868.

[9] Tiwari SB, Hagen G, Guilfoyle TJ. The roles of auxin response factor domains in auxin-responsive transcription, The Plant Cell, 2003, vol. 15 (pg. 533-543)

[10] Tiwari SB, Wang XJ, Hagen G, Guilfoyle TJ. Aux/IAA proteins are active repressors and their stability and activity are modulated by auxin, The Plant Cell, 2001, vol. 13 (pg. 2809-2822)

[11] Wang S, Tiwari SB, Hagen G, Guilfoyle TJ. AUXIN RESPONSE FACTOR7 restores the expression of auxin-responsive genes in mutant Arabidopsis leaf mesophyll protoplasts, The Plant Cell, 2005, vol. 17 (pg. 1979-1993)

[12] Wang, D., Pei, K., Fu, Y., Sun, Z., Li, S., Liu, H., ... & Tao, Y. (2007). Genome-wide analysis of the auxin response factors (ARF) gene family in rice (Oryza sativa). *Gene*, *394*(1-2), 13-24.

[13] Liu, Y., Jiang, H., Chen, W., Qian, Y., Ma, Q., Cheng, B., & Zhu, S. (2011). Genome-wide analysis of the auxin response factor (ARF) gene family in maize (Zea mays). *Plant Growth Regulation*, *63*(3), 225-234.

[14] Li, S. B., OuYang, W. Z., Hou, X. J., Xie, L. L., Hu, C. G., & Zhang, J. Z. (2015). Genome-wide identification, isolation and expression analysis of auxin response factor (ARF) gene family in sweet orange (Citrus sinensis). Frontiers in Plant Science, 6, 119.

[15] Shen, C., Yue, R., Sun, T., Zhang, L., Xu, L., Tie, S., ... & Yang, Y. (2015). Genome-wide identification and expression analysis of auxin response factor gene family in Medicago truncatula. *Frontiers in Plant Science*, *6*, 73.

[16] Wang, S., Bai, Y., Shen, C., Wu, Y., Zhang, S., Jiang, D., ... & Qi, Y. (2010). Auxin-related gene families in abiotic stress response in Sorghum bicolor. Functional & integrative genomics, 10(4), 533-546.

[17] Le, B., Nawaz, M. A., Rehman, H. M., Le, T., Yang, S. H., Golokhvast, K. S., ... & Chung, G. (2016). Genome-wide characterization and expression pattern of auxin response factor (ARF) gene family in soybean and common bean. *Genes & Genomics*, *38*(12), 1165-1178.

[18] Van Ha, C., Le, D. T., Nishiyama, R., Watanabe, Y. A. S. U. K. O., Sulieman, S., Tran, U. T., ... & Tran, L. S. P. (2013). The auxin response factor transcription factor family in soybean: genome-wide identification and expression analyses during development and water stress. *DNA research*, *20*(5), 511-524.

[19] Wang, Y., Deng, D., Shi, Y., Miao, N., Bian, Y., & Yin, Z. (2012). Diversification, phylogeny and evolution of auxin response factor (ARF) family: insights gained from analyzing maize ARF genes. *Molecular biology reports*, *39*(3), 2401-2415.

[20] Xing, H., Pudake, R. N., Guo, G., Xing, G., Hu, Z., Zhang, Y., ... & Ni, Z. (2011). Genome-wide identification and expression profiling of auxin response factor (ARF) gene family in maize. *BMC genomics*, *12*(1), 1-13.

[21] Song, X., Liu, G., Duan, W., Liu, T., Huang, Z., Ren, J., ... & Hou, X. (2014). Genome-wide identification, classification, and expression analysis of the heat shock transcription factor family in Chinese cabbage. *Molecular genetics and genomics*, *289*(4), 541-551.

[22] Sun, R., Wang, K., Guo, T., Jones, D. C., Cobb, J., Zhang, B., & Wang, Q. (2015). Genome-wide identification of auxin response factor (ARF) genes and its tissue-specific prominent expression in Gossypium raimondii. *Functional & integrative genomics*, *15*(4), 481-493.

[23] Zhang, B., & Pan, X. (2009). Expression of microRNAs in cotton. *Molecular biotechnology*, *42*(3), 269-274.

[24] Xie, F., Jones, D.C., Wang, Q., Sun, R. and Zhang, B. (2015), Small RNA sequencing identifies miRNA roles in ovule and fibre development. Plant Biotechnol J, 13: 355-369

[25] Barry G. Hall, Building Phylogenetic Trees from Molecular Data with MEGA, *Molecular Biology and Evolution*, Volume 30, Issue 5, May 2013, Pages 1229–1235

[26] Kriegner, A., Cervantes, J. C., Burg, K., Mwanga, R. O., & Zhang, D. (2003). A genetic linkage map of sweetpotato [Ipomoea batatas (L.) Lam.] based on AFLP markers. *Molecular Breeding*, *11*(3), 169-185.

[27] Li, M., Yang, S., Xu, W., Pu, Z., Feng, J., Wang, Z., ... & Tan, W. (2019). The wild sweetpotato (Ipomoea trifida) genome provides insights into storage root development. *BMC plant biology*, *19*(1), 119.

[28] Rajapakse, S., Nilmalgoda, S. D., Molnar, M., Ballard, R. E., Austin, D. F., & Bohac, J. R. (2004). Phylogenetic relationships of the sweetpotato in Ipomoea series Batatas (Convolvulaceae) based on nuclear β-amylase gene sequences. *Molecular phylogenetics and evolution*, *30*(3), 623-632.

[29] Huang, J. C., & Sun, M. (2000). Genetic diversity and relationships of sweetpotato and its wild relatives in Ipomoea series Batatas (Convolvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. *Theoretical and Applied Genetics*, *100*(7), 1050-1060.

[30] Ellis, C. M., Nagpal, P., Young, J. C., Hagen, G., Guilfoyle, T. J., & Reed, J. W. (2005). AUXIN RESPONSE FACTOR1 and AUXIN RESPONSE FACTOR2 regulate senescence and floral organ abscission in Arabidopsis thaliana. *Development*, *132*(20), 4563-4574.

[31] Wang, J. W., Wang, L. J., Mao, Y. B., Cai, W. J., Xue, H. W., & Chen, X. Y. (2005). Control of root cap formation by microRNA-targeted auxin response factors in Arabidopsis. *The Plant Cell*, *17*(8), 2204-2216.

[32] Goh, Tatsuaki, et al. "Multiple AUX/IAA–ARF modules regulate lateral root formation: the role of Arabidopsis SHY2/IAA3-mediated auxin signalling." *Philosophical Transactions of the Royal Society B: Biological Sciences* 367.1595 (2012): 1461-1468.

[33] Narise, Takafumi, et al. "Involvement of auxin signaling mediated by IAA14 and ARF7/19 in membrane lipid remodeling during phosphate starvation." *Plant molecular biology* 72.4-5 (2010): 533-544.

[34] Tian, Chang-en, et al. "Disruption and overexpression of auxin response factor 8 gene of Arabidopsis affect hypocotyl elongation and root growth habit, indicating its possible involvement in auxin homeostasis in light condition." *The Plant Journal* 40.3 (2004): 333-343.

[35] Sieburth, Leslie E., et al. "SCARFACE encodes an ARF-GAP that is required for normal auxin efflux and vein patterning in Arabidopsis." The Plant Cell 18.6 (2006): 1396-1411.

[36] Okushima, Yoko, et al. "Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in Arabidopsis thaliana: unique and overlapping functions of ARF7 and ARF19." *The Plant Cell* 17.2 (2005): 444-463.

[37] Pekker, Irena, John Paul Alvarez, and Yuval Eshed. "Auxin response factors mediate Arabidopsis organ asymmetry via modulation of KANADI activity." *The Plant Cell* 17.11 (2005): 2899-2910.

[38] Nagpal, Punita, et al. "Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation." *Development* 132.18 (2005): 4107-4118.

[39] Wu, Shan, et al. "Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement." Nature communications 9.1 (2018): 1-12.

[40] Okushima, Yoko, et al. "ARF7 and ARF19 regulate lateral root formation via direct activation of LBD/ASL genes in Arabidopsis." The Plant Cell 19.1 (2007): 118-130.