Genome-Wide Identification Analysis of The Auxin Response Factor Gene Family (ARF) in Zizania Latifolia and its Role in Swelling Stem Formation After Ustilago Esculenta Infection

Jie Li
Anhui Agricultural University
Jinfeng Hou
Anhui Agricultural University
Zhiyuan Lu
Anhui Agricultural University
Yang Yang
Anhui Agricultural University
Hao Wang
Anhui Agricultural University
Ying Wu
Anhui Agricultural University
Lingyun Yuan
Anhui Agricultural University
Guohu Chen
Anhui Agricultural University
Chenggang Wang
Anhui Agricultural University
Defang Gan
Anhui Agricultural University
Shidong Zhu (sdzhuaau@sina.cn)
Anhui Agricultural University https://orcid.org/0000-0001-6046-9060

Research article

Keywords: ZIARF, Z. latifolia, Phylogenetic analysis, auxin-activated signalling pathway, Swelling stem formation, Transcriptomic analysis, U. esculenta infection

Posted Date: January 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-151191/v1

Abstract

Background: Auxin signalling plays a crucial role in plant growth and development. Although the auxin response factor (*ARF*) gene family has been studied in some plant species, its structural features, molecular evolution, and expression profile in *Zizania latifolia* (*Z. latifolia*) are still not clear.

Results: Our study identified 33 putative YSL genes from the whole *Z. latifolia* genome. Furthermore, a comprehensive overview of the *ZlARFs* was undertaken, including phylogenetic relationship, gene structures, conserved domains, synteny, Ka/Ks, motifs, and subcellular locations of the gene product. Synteny analyses and the calculation of Ka/Ks values suggested that all 57 orthologous/paralogous gene pairs between *Z. latifolia* and *Z. latifolia*, *Z. latifolia* and *Oryza sativa* have experienced strong purifying selection. The phylogenetic analysis of ARFs indicated that the *ZlARFs* can be divided into 6 classes and that most *ZlARFs* from *Z. latifolia* have closer relationships with *Oryza sativa* than with *Arabidopsis*. RNA-Seq data and qRT-PCR analyses showed that *ZlARF* genes were expressed in TDF treatment and *U. esculenta* infection, while some *ZlARFs* exhibited high expression levels only in *U. esculenta* infection. Meanwhile, the interaction networks and gene ontology (GO) term of the *ZlARF* genes were constructed and 23 *ZlARF* co-expressed genes were identified, most of which were down-regulated involve auxin-activated signalling pathway in after swelling stem formation. Transcriptome analysis results verified the relevant functions of ARF genes, and most *ZlARF* genes regulated physiological processes in response to differential cell expansion.

Conclusion: Comprehensive bioinformation analysis of the auxin response factor gene family (ARF) in *Z. latifolia* and its association with swelling stem formation after *U. esculenta* infection. The bioinformatic and RNA-Seq analyses provided valuable information for further study on the regulation of the growth and development of swelling stem formation by *ZlARFs* in *Z. latifolia*.

Highlights

1. Identifying the *ARF* gene family members in *Z. latifolia* is important.
2. *ZlARF* and *OsARF* were found to have high similarity, and *Z. latifolia* and *Oryza sativa* have ancient genetic affinity.
3. The information of genomic scaffolds was analyzed by collinear analysis for the first time.
4. Transcriptome analysis results indicated the relevant role in swelling stem formation in TDF treatment and *U. esculenta* infection.

Background

The genus *Z. latifolia* belongs to the wild rice tribe (*Oryza*) of the Gramineae family [1]. *Z. latifolia* is grown as a significant agricultural crop across Asia and has been cultivated for more than 1500 years. In China, the planting area was second only to lotus root as a perennial aquatic vegetable, with both the swelling stem and grain being edible. Formation of swelling stem was pretty interesting usually parasitized by a smut fungus, *Ustilago esculenta* Henn., which stimulates the enlargement of the swollen culm under suitable conditions [2]. Interestingly, stem growth is spontaneous and due to plant infection with *Ustilago esculenta* (*U. esculenta*) in the environment, leading to the formation of a swelling stem [3]. After forming, *Z. latifolia* comes in three different forms “normal *Z. latifolia*”, “grey *Z. latifolia*” and “male *Z. latifolia*”. Grey *Z. latifolia* could not be reversibly developed to white Jiaobai. Rather, it eventually becomes “male *Z. latifolia*”, which did not form stems, shows normal flowering, and does not contain *U. esculenta* [4]. However, the mechanism was still unclear. Yan applied an ultra-high performance liquid chromatography coupled to triple quadrupole mass spectrometry (UHPLC-QqQMS) - based metabolomics and identified 672 metabolites from *Z. latifolia* [5]. As it is rich in bioactive compounds, *Z. latifolia* exhibits antioxidant activity, alleviates insulin resistance and lipotoxicity, protects against cardiovascular disease, and possesses other bioactivities and health benefits [6, 7]. Therefore, *Z. latifolia* is both a health-promoting food and a high-quality nutrition source. It may also serve as an excellent food for patients with diabetes, obesity, and hyperlipidaemia. In our previous study, the utilization of exogenous IAA and TIBA in present study proved IAA played a dominant role in regulating *Z. latifolia* swelling stem growth [8]. Most classical plant hormones (IAA) are also produced by pathogenic and symbiotic fungi [9]. We found that auxin plays an important role in stem formation of *Z. latifolia*. Therefore, further research has important research value of auxin response to stem formation of *Z. latifolia*.

The phytohormone indole-3-acetic acid (IAA) plays a central role in the coordinated growth, development and stem formation in many plants [10, 11]. Auxin is a class of plant hormones and growth regulators with some morphogen-like characteristics. The phytohormone auxin represented by IAA affects a variety of complex biological processes, including embryogenesis, lateral root initiation, apical dominance, growth and stem tip formation [12, 13]. Auxin signalling enhances some response genes, including that of Gretchen Hagen 3 (*GH3)*, *Aux/IAA*, and small auxin up RNA (SAUR) gene family members [14, 15]. It regulates plant development and biological processes by interacting with the plant hormone signal transcription factors that have auxin response elements (AuxREs) in some affected genes. ARFs contain three unique domains: a conserved N-terminal DNA-binding domain (DBD) [16, 17], a variable middle transcriptional regulatory region (MR) [17, 18] that functions as an activation domain (AD) or repression domain (RD), and a C-terminal domain (CTD) [19]. A massive number of speculative AuxREs have been found in upstream promoter elements of early auxin responsive genes, such as a large number of duplicated conserved motif TGTCTC or certain variations of it (TGTCAC and TGTCCC), which confers auxin responsiveness [20, 21]. Plant transcription factors that specifically bind to TGTCTC structural motifs and control auxin responses are named as *ARFs*. ARF genes can specifically bind to TGTCTC including in AuxREs, which can regulate responses upon synthesis of the plant hormone auxin [22].

AtARF1 gene was primarily discovered in wild-type *Arabidopsis* through an adapted exploratory yeast one hybrid (Y1H) with a core TGTCTC motif used as a decoy sequence [23]. Further, a recent report highlighted the functional importance of the MR. This report indicated that the intrinsically disordered middle region and the PB1 interaction domain of the ARFs drive protein assembly formation [18]. The DBD is a highly conserved structural domain found exclusively in transcription factors that enables proteins of higher plants to interact with domains in other proteins [24]. The MR between DBD and CTD activates or suppresses transcription according to its amino acid composition [25].

It is reported that the ARF proteins are encoded by a many family and are conserved throughout the evolution of the plant kingdom, and since the expansion of the family seems to be related to the evolution and diversity of plants, the function of the ARFs has been thoroughly studied [26, 27] In recent years, microarray assays has shown that the *Arabidopsis AtARF1* and *AtARF5* domains are have greater similarity to domains with a TGTCGG sequence than to those with the AuxRE TGTCTC element [16]. Many functions of the ARF gene interaction mechanisms and their regulation of biological process, including upregulated growth levels, in *Z. latifolia* need to be explored Taking advantage of the genomic identification of the *Arabidopsis thaliana ARF* gene (*AtARF*) family, many researchers have found that the *Arabidopsis ARF1* and *ARF2* molecules function as transcriptional inhibitors connected with the regulation of plant development, organogenesis and growth. [28, 29]. Meanwhile, the ARF family has been found in different species (*Zea mays*, *Oryza sativa* and so on), the identity of which has been discovered and continuously verified. Sequences derived from massive numbers of sequencing projects are conducive to functional research, offering a resource for discovering transcription factor families. A differential genome analysis of species showed that *Oryza sativa*, *Arabidopsis*, *Hordeum vulgare, Solanum lycopersicum*, *Setaria italica,* tobacco*, Vitis vinifera*, and *Zea mays* have 25, 23, 17, 17, 24, 50, 20 and 31 *ARF* family members, respectively [26, 30-36].

Auxin response factors (*ARFs*) are plant-specific transcription factors that couple the perception of the hormone auxin to gene expression programmes essential to all land plants [37]. The complete genome sequencing of *Z. latifolia* was published in 2015 and has been widely publicized, laying a solid foundation for research on *Z. latifolia* genes [3]. However, the *ARF* gene family has not been identified in *Z. latifolia*. The identification and name of an *ARF* gene in this species will provide a solid foundation for future research. The recent development of the RNA-Seq method has led to not only greater information on genome-wide gene expression but also to analyses with advantages such as higher sensitivity and greater and more dynamic ranges of gene expression patterns and functions to research [38]. The complete genomic sequence of *Z. latifolia* provides a precious resource for analysing and comparing members of the *ARF* gene family in terms of evolution at different stages. The genome-wide analysis indicated that the dominant auxin-responsive gene family in plants has various nuclear localization signals and dual origins [39]. Because *ARF* genes are critical factors in plant growth and development, exploring these genes will aid to information for the comprehensive understanding of the auxin changes in *Z. latifolia*.

In this study, we performed a systematic analysis of the ARF family in *Z. latifolia*. The biophysical properties of *ZlARF* genes were determined. Furthermore, RNA-Seq data from three stem formation were analysed. A comprehensive bioinformatic analysis of *ZlARF* genes was employed, including an analysis of the conserved motifs, exon-intron structure, phylogenetic tree, conserved domains, synteny, Ka/Ks values and divergence times, expression patterns of 33 *ZlARFs*. Meanwhile, the homology of each *ZlARFs* gene family member and relative gene expression levels during swelling stem formation were determined. Samples with before and after stem formation were selected according to their auxin level for transcriptome analysis. By transcriptome sequencing, we explored the relevant expression and functional profiles before and after swollen stem formation, TDF (Triadimefon)-treatment and *U. esculenta* infection in *Z. latifolia*. Our results suggested that *ZlARFs* genes provide valuable information for further studies on the mechanism of swelling stem formation in *Z. latifolia.*

Methods

Plant materials and Treatments

To explore the changes in ARF before and after stem formation, we identified the ARF gene family and analysed the relationship between the evolution of the gene family and gene expression. The specific analysis was follows: The single-season cultivar (*Z. latifolia* Turcz., cv. 'Dabieshan No. 1') was selected as the experimental material (Yuexi County Longjing Ecological Agriculture Development Co., Ltd., Anhui, China) (Fig. 1).

For subsequent experiments, stem samples were collected at 153 and 159 days after transplantation, with stem diameters about 5cm, respectively. Meanwhile, we used two- method to TDF -treatment (80 mg L^{-1}) and *U. esculenta* infection in *Z. latifolia.* Triadimefon (TDF, CAS: 43121-43-3) is widely used in agriculture and medicine as fungicide, as it can prevent fungal colonization and inhibit the formation of stems and hyphae. *Z. latifolia* was formed by *U. esculenta* infection, we could stimulate the expansion of the stem formation through injection of male *Z. latifolia*. All swelling stems were harvested at appropriate time point, and each sampling was selected in triplicate using freeze tube storage. All samples were placed in cryopreservation tubes and immediately followed by liquid nitrogen storage in a refrigerator at -80°C for RNA extraction and some index determination. All experiments were performed in three replicates.

Genome-wide identification of *ARFs* gene family members in *Z. latifolia*

The genome annotations and sequences of *Z. latifolia* were downloaded using the *Z. latifolia* genome database (*Z. latifolia*, http://ibi.zju.edu.cn/ricerelativesgd), and a local BLASTP database was built using the BioEdit tool. The comprehensive identification of *Z. latifoliaARF* gene family members were achieved using all *AtARFs* from the *Arabidopsis* genome database. The *ZlARF* gene family members were identified by a BLASTP search [60]. Each new query parameter for the *AtARF* sequence was set to an E-value $\leq 10^{-10}$ to avoid the loss and addition of orthologues. The domains of obtained ZlARF proteins were also further verified using NCBI-Conserved Domain database (https://www.ncbi.nlm.nih.gov/Structure/) search program and SMART databases (http://smart.emblheidelberg.de/) [61]. Those proteins which lack ARF domain were removed from further analysis. In addition, protein sequences that were found with obvious errors in their gene. All the candidate genes that might contain the ARF domain were further verified. Furthermore, the query sequence was used for further analysis of the structural domain using the Pfam program (http://pfam.xfam.org). Finally, 33 ARF gene models were identified in the *Z. latifolia* genome for further analysis.

Related annotation information about the full-length sequences, chromosome location, deduced polypeptides, and CDS sequences were acquired from the *Z. latifolia* genomic database using the UltraEdit tool. The sequences containing DBD and MR domains were screened. The PI/Mw was calculated with an ExPASy server to predict the molecular weight of the polypeptide (Mw), instability index, theoretical isoelectric point (PI) and amino acid number (aa) of each *ZlARF* protein sequence (https://web.expasy.org/protparam/). Finally, we identified and isolated 33 *Z. latifoliaARF* genes (*ZlARFs*). These *ARF* genes are

named by chromosomal location. The N-terminal signal peptide of the 33 *ZlARFs* protein sequences was predicted using the WoLF PSORT [62] server (https://wolfpsort.hgc.jp/), which predicts genes subcellular locations.

Interspecific phylogenetic tree analysis, classification of *ZlARF* genes

The phylogenetic tree was constructed from the 33 *ZlARFs* (*Z. latifolia*), 23 *AtARFs* (*Arabidopsis thaliana*) and 25 *OsARF* (*Oryza sativa*) protein sequences described above using the latest MEGA-X software and the neighbor-joining (NJ) method, and the 1000 replicates bootstrap test was adopted using the ARF protein sequences [63], save the NWK format file. To beautify the evolutionary tree and put the composition of amino acids into the evolutionary tree by the online tool iTOL [64]. The classification of the *ZlARFs* gene were classified by comparing them to the ARF of *ZlARFs*, *AtARFs* and *OsARFs*. Meanwhile, 33 *ZlARFs* evolutionary tree also was constructed using the same method.

Analysis of motif composition structural domain of *ZlARFs* and cis-elements predictions

The genomic sequence and GFF3 file of the 33 *ARF* genes were downloaded from the *Z. latifolia* genomic database, and the intro-exon number patterns were detected through the GSDS 2.0 website as previously described [65]. These sequences were used for further alignment of the structural domains, namely, the DBD, ARF and CTD domains, at the Pfam website (http://pfam.xfam.org) [66]. MEME Suite Version 5.0.5 (http://meme-suite.org/) was used to analyse the conserved motifs of the *ZlARFs* gene family, and the parameters are set as follows: motif site distribution was zero or one per sequence (ZOOPS), the maximum number of motifs that that MEME should find was set to 30, sites per motif was set to be between 2 and 33 and motif width was between 6 and 50. A condensed phylogenetic tree was constructed by the construct/test neighbour-joining (NJ) tree method using the MEGA-X software (Version 10.0.5) and a 1000 replicate bootstrap test. Tbtools (Version:1.047) (Chen C, 2020) was used for visual analysis. The promoter regions (3.0 kb upstream sequences) of all YSL genes were used to predict cis-acting elements in the PlantCARE database [67]. Tbtools (v1.047) was used for all visual analysis [68].

Collinearity analyses and inference of divergence time

Synteny analysis among the *Z. latifolia*, *Arabidopsis thaliana* and *Oryza sativa* genomes were conducted via all against all BLASTP comparisons. *Z. latifolia* genome was scaffold information and no specific chromosome. To represent the co-progressivity relationship, combine all chromosome locations into one marked scaffold. We compared gene pair to *Z. latifolia- Oryza sativa* and *Z. latifolia- Z. latifolia*. The all protein sequences were used as queries to search the corresponding protein sequence data ($E < 1*e^{-5}$, top 5 matches). Synteny pairs were extracted using MCScanX [69] to identify syntenic blocks and duplications within the *ARFs*. For synteny analysis, Circos tool [70] and TBtools was used to mark the identified positions on the chromosomes. The rates of synonymous (Ks) and nonsynonymous (Ka) substitutions were calculated for duplicated ARF genes with the Ka/Ks calculator. Ks could be used as a proxy for time to estimate the dates of the segmental duplication events. Ks values > 2.0 were discarded to avoid the risk of substitution saturation. The Ks value was calculated for each of the gene pairs and then used to calculate the approximate date of the divergence time ($T = Ks/2\lambda$), divergence rates $\lambda = 1.5*10^{-8}$ for *Z. latifolia* species [71].

Transcriptome (RNA-Seq) analysis, cDNA library construction

Total RNA from the 12 stem samples taken before and after stem formation, TDF-treatment and *U. esculenta* infection were extracted using the mirVana miRNA isolation kit (mirVana™ miRNA isolation kit, Ambion-1561) following the manufacturer's protocol. RNA integrity was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The samples with an RNA integrity number (RIN) ≥ 7 were subjected to the subsequent analysis. The libraries were constructed using the TruSeq Stranded mRNA LT sample prep kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. Then, these cDNA libraries were sequenced on an Illumina sequencing platform (HiSeqTM 2500 or an Illumina HiSeq X Ten), and 125 bp/150 bp paired-end reads were generated. In addition, the p-values < 0.05 & |log2 (fold change)| ≥ 0.5 was set as the threshold for significantly differential expression. The gene expression of *Z. latifolia* was profiled by RNA-Seq datasets from the NCBI Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA). Those datasets could be retrieved from the SRA database by searching BioProject id PRJNA551367.

Heatmap and microscopic observations

A hierarchical clustered heatmap for ARF genes was plotted with the heatmap package [72], wherein "Euclidean" distance and "Complete" cluster were used for both row-based and column-based clustering. To study the interaction relationships of the above experimental groups, their respective swollen stem samples were stained using aniline blue (CAS: 28631-66-5, Solarbio), according to an improved method [73]. Fresh stem tip samples of *Z. latifolia* tissue was sliced and fixed with a fluid containing the Carnot fixative (glacial acetic acid and alcohol, volume ratio of 1:3) and put into a 2-mL centrifuge tube, stored in refrigerator at 4°C for 24 h. To it was added the right amount 10% KOH solution, after which it was subjected to 85°C high-temperature treatment lasting 1 h; washed several times during the KOH solution, after processing the used for subsequent dyeing observation. The slices were transferred into a petri dish containing aniline blue dye, and stained on a horizontal shaker for 5–10 min. After decolorization with 75% alcohol, the slices were observed and photographed under a fluorescence microscope (Olympus, Japan, SZX10).

Functional annotation, KEGG enrichment pathway analysis and interacting network

Differentially expressed genes (DEGs) were subjected to GO and KEGG enrichment (http://www.genome.jp/kegg/) analyses to characterize their biological functions and significantly enriched metabolic pathways or signal transduction pathways. The DEGs were submitted to the GO analysis of enrichment using the GO-Seq R package based on Wallenius' noncentral hyper-geometric distribution [74]. The KOBAS 3.0 [75] website (http://kobas.cbi.pku.edu.cn/index.php) was used to test statistically the enrichment of the DEGs in KEGG pathways. The interaction network associated with *ZlARF* genes was constructed using the STRING database [76] and Cytoscape software [77]. The STRING tool (http://string-db.org/, version 11.0) was analysed 33ARF interacting network. We constructed an interactive network of *Z. latifolia* by comparing *Oryza sativa* genome used red to marked in the figure.

RNA extraction, cDNA synthesis and quantitative real-time PCR analysis

To explore the expression patterns of 33 *ARF* genes were performed before and after swelling stem formation in *Z. latifolia* using real-time qPCR. Total RNA samples were isolated from different *Z. latifolia* stages using an RNA-prep pure plant kit (Tiangen, DP432). The DNase-treated RNA was reverse transcribed for cDNA using the Prime Script™ RT reagent kit (TaKaRa, Japan). qRT-PCR was then performed with the TB Green™ *Premix Ex Taq*™ II (TaKaRa, RR820A). The *ZlARFs* gene primers were designed using Oligo 7.60 software. qRT-PCR amplification was performed with a Bio-Rad CFX96™ system according to the manufacturer's protocol (Bio-Rad, CA, USA). The relative expression level of the *Zl-Actin* gene was used as an internal reference for data standardization. Three biological replicates of each sample were subjected to qRT-PCR, and the relative gene expression levels were calculated with the $2^{-\Delta\Delta Ct}$ method [78]. The real-time qPCR primer sequences are shown in the additional information (Table S4). The result is displayed as the mean ± SD (n = 3 biological replicates).

Statistical analysis

The statistical analysis of the relative expression data was performed using SPSS software. Every sample was based on three biological replicates. The SPSS was used for the analysis of significance (Duncan test significant: * $P < 0.05$; highly significant: ** $P < 0.01$).

Results

Genome-wide identification of the ZlARFs gene family in Z. latifolia

To identify the *ZlARFs* gene family of *Z. latifolia*, we performed BLASTP program searches of the *Z. latifolia* genome databases by ARF sequence summarized in the *Oryza sativa* and *Arabidopsis thaliana* data sets (Fig. S1). Using each predicted protein sequence as a query to search, 33 potential *ARF* proteins sequence were discovered. All *ZlARFs* family members contain the DBD and MR conserved structural domains by Pfam and SMART website to identification. On the basis of our analysis, a generic *ZlARF* naming system used, from *ZlARF1* to *ZlARF33* by chromosome location, which distinguished each *ARF* gene according to the *Z. latifolia* scaffold information. All *ZlARFs* with the DBD and MR domain are summarized in Table 1. Overall, the analysis showed that there were 33 *ZlARFs* members in the genome of *Z. latifolia*.

Table 1
The ARF gene family members in *Z. latifolia*.

Gene name[A]	Gene_ID[B]	Domain[C]	Chromosome location[D]	ORF length (bp)[E]	Deduced polypeptide[F]				Direction[G] (R/F)	No. Of intron[H]	*Ara* hor
					Length (aa)	MW (KDa)	PI	Instability index			
ZlARF1	*Zlat_10046274*	DBD; MR; CTD	scaffold_21:842350:855269	2442	813	90.40	6.49	58.37	R	17	AT1 (At
ZlARF2	*Zlat_10045507*	DBD; MR; CTD	scaffold_23:1002962:1014822	2433	810	89.09	5.54	53.63	F	14	AT4 (At
ZlARF3	*Zlat_10042822*	DBD; MR	scaffold_44:189023:210245	2706	901	100.51	5.86	73.27	F	13	AT1 (At
ZlARF4	*Zlat_10039405*	DBD; MR; CTD	scaffold_84:856365:861456	2529	842	92.99	6.74	59.32	R	13	AT5 (At
ZlARF5	*Zlat_10036653*	DBD; MR; CTD	scaffold_119:422479:427812	2511	836	92.53	6.69	56.90	R	15	AT5 (At
ZlARF6	*Zlat_10033866*	DBD; MR; CTD	scaffold_166:63647:71937	3399	1132	125.74	6.06	62.29	F	13	AT1 (At
ZlARF7	*Zlat_10033398*	DBD; MR; CTD	scaffold_169:614405:620911	2442	813	90.50	5.95	62.86	R	13	AT5 (At
ZlARF8	*Zlat_10032874*	DBD; MR	scaffold_173:486166:491783	1227	408	45.30	6.25	49.24	F	11	AT2 (At
ZlARF9	*Zlat_10031865*	DBD; MR; CTD	scaffold_190:298677:304803	2454	817	91.06	5.98	63.85	R	13	AT5 (At
ZlARF10	*Zlat_10031816*	DBD; MR	scaffold_190:27070:31806	2757	918	101.26	5.86	52.41	R	12	AT1 (At
ZlARF11	*Zlat_10031564*	DBD; MR; CTD	scaffold_211:13403:18162	2676	891	99.25	5.85	63.99	R	13	AT1 (At
ZlARF12	*Zlat_10028683*	DBD; MR	scaffold_223:24568:27701	1182	393	43.67	6.27	68.73	F	12	AT1 (At
ZlARF13	*Zlat_10028341*	DBD; MR	scaffold_239:609211:611540	2088	695	74.90	7.11	48.00	R	2	AT4 (At
ZlARF14	*Zlat_10026465*	DBD; MR	scaffold_266:621432:622703	1140	379	41.89	9.05	39.17	F	1	AT4 (At
ZlARF15	*Zlat_10022744*	DBD; MR; CTD	scaffold_324:216217:223745	3423	1140	126.44	6.41	65.26	R	13	AT1 (At
ZlARF16	*Zlat_10023255*	DBD; MR; CTD	scaffold_336:152463:158522	3183	1060	117.69	7.59	72.76	F	16	AT1 (At
ZlARF17	*Zlat_10022955*	DBD; MR	scaffold_347:482128:484518	2142	713	76.37	6.15	46.55	R	3	AT4 (At
ZlARF18	*Zlat_10022003*	DBD; MR; CTD	scaffold_358:524224:528820	2181	726	80.79	5.68	50.55	R	14	AT4 (At
ZlARF19	*Zlat_10021233*	DBD; MR; CTD	scaffold_378:444906:451433	1623	540	60.18	7.68	53.93	F	12	AT1 (At

[A] Names of *ZlARF* genes in *Z. latifolia*.

[B] Gene ID: annotated in *Z. latifolia* genome.

[C] DBD: B3 DNA-binding domain; MR: Middle transcriptional regulatory region; CTD: C-terminal domain.

[D] Scaffold numbers assembled in *Z. latifolia* genome.

[E] ORF length: Length of open reading frame in base pairs.

[F] Length: The number of amino acids, PI: theoretical isoelectric point, Mw: molecular weight of polypeptide, Instability index.

[G] Direction (R/F): *ZlARFs*.

[H] No. Of Extron: Exon number of ARF in *Z. latifolia*.

[I] Homology: the homology with *Arabidopsis ARFs* (*AtARFs*).

Gene name[A]	Gene_ID[B]	Domain[C]	Chromosome location[D]	ORF length (bp)[E]	Deduced polypeptide[F]				Direction[G] (R/F)	No. Of intron[H]	*Ara* hor
					Length (aa)	MW (KDa)	PI	Instability index			
ZlARF20	*Zlat_10019056*	DBD; MR; CTD	scaffold_432:225158:229813	2640	879	98.07	5.78	62.27	F	13	AT1 (At
ZlARF21	*Zlat_10018204*	DBD; MR; CTD	scaffold_457:428399:433130	2049	682	75.79	6.18	59.80	F	13	AT1 (At
ZlARF22	*Zlat_10017981*	DBD; MR; CTD	scaffold_460:101457:108766	2979	992	110.89	5.83	69.65	F	14	AT1 (At
ZlARF23	*Zlat_10010482*	DBD; MR	scaffold_735:135203:141844	2067	688	74.55	6.44	49.78	F	9	AT2 (At
ZlARF24	*Zlat_10009593*	DBD; MR	scaffold_820:62920:66333	2061	686	74.94	6.57	47.96	F	2	AT4 (At
ZlARF25	*Zlat_10008345*	DBD; MR	scaffold_855:199807:202584	2073	690	75.08	8.18	47.81	R	2	AT4 (At
ZlARF26	*Zlat_10008072*	DBD; MR	scaffold_872:196040:201540	2115	704	77.63	7.60	55.53	F	10	AT2 (At
ZlARF27	*Zlat_10005792*	DBD; MR	scaffold_1005:123221:128458	2751	916	100.74	6.02	54.76	R	12	AT1 (At
ZlARF28	*Zlat_10005803*	DBD; MR; CTD	scaffold_1041:97033:110057	2700	899	98.93	6.14	52.53	F	16	AT5 (At
ZlARF29	*Zlat_10005411*	DBD; MR	scaffold_1089:69:4914	1065	354	39.24	6.11	56.48	R	9	AT2 (At
ZlARF30	*Zlat_10002430*	DBD; MR	scaffold_1824:15174:22474	2445	814	88.00	6.77	52.03	R	10	AT2 (At
ZlARF31	*Zlat_10002256*	DBD; MR; CTD	scaffold_1883:1595:7058	2493	830	92.45	6.24	62.45	R	14	AT5 (At
ZlARF32	*Zlat_10000945*	DBD; MR	scaffold_2488:6240:8497	2115	704	75.94	6.30	51.76	F	2	AT4 (At
ZlARF33	*Zlat_10000277*	DBD; MR	scaffold_3192:2200:4370	1551	516	55.33	5.27	45.46	R	1	AT1 (At

[A] Names of *ZlARF* genes in *Z. latifolia*.

[B] Gene ID: annotated in *Z. latifolia* genome.

[C] DBD: B3 DNA-binding domain; MR: Middle transcriptional regulatory region; CTD: C-terminal domain.

[D] Scaffold numbers assembled in *Z. latifolia* genome.

[E] ORF length: Length of open reading frame in base pairs.

[F] Length: The number of amino acids, PI: theoretical isoelectric point, Mw: molecular weight of polypeptide, Instability index.

[G] Direction (R/F): *ZlARFs*.

[H] No. Of Extron: Exon number of ARF in *Z. latifolia*.

[I] Homology: the homology with *Arabidopsis ARFs* (*AtARFs*).

The deduced polypeptides are characterized by three types of information: length (number of amino acids, aa), molecular weight and theoretical PI. The ZlARF proteins showed wide variation in their length, molecular weight (MW) and isoelectric point (pI) (Table S3). *ZlARF* ORF lengths ranged from 1065 (*ZlARF29*) to 3399 bp (*ZlARF6*) and the molecular weights ranged from 39.24 (*ZlARF19*) to 126.44 kDa (*ZlARF15*) (Table 1). The PI range from 5.54 (*Z1ARF2*) to 9.05 (*Z1ARF14*), among which 6 ARF protein PI more than 7 were alkaline, and the remaining 26 ARF genes less than 7 were acidic.

Exon-intron, structural domains, conserved motifs and cis**-elements of the** ZlARF **members**

The functional motifs and gene exon-intron positions were analysed based on evolutionary tree relationships (Fig. 2, Fig. 3). The full-length protein sequences alignments from all the *ZlARFs* gene products generated unrooted phylogenetic trees, which indicated that all the *ZlARFs* can be divided into four major categories, which are presented by colour (purple, orange, red, and green) by branching the evolutionary tree (Fig. 2A). Thirty-three *ZlARFs* formed 12 orthologous gene pairs, and all pairs were solidly supported by the bootstrap tests(N > = 99%). The results indicated that all ARF genes were interrupted by introns, and the number of introns varied from 1 to 17 (Fig. 2B). The exon-intron organization suggested that all *ZlARF* genes showed conserved patterns of

gene structure were associated with the DBD and MR domain. In the same evolutionary tree categories, the intron exon data were basically consistent. Overall, the highly similar gene structures and motif distributions of the *ZlARF* members were consistent with their phylogenetic relationships (Fig. 2, Fig. 3). The main structural domains were detected by Pfam. In addition, all sequences contained the MR and DBD domains, among 17 sequences (*ZlARF1/3/4/5/6/7/9/11/15/16/18/19/20/21/22/28/31*) contained the CTD domains (Fig. 2C). Furthermore, the sequence alignment of the homologous domain sequences of the ZlARF proteins revealed that the domain sequences were highly conserved (Fig. S3).

We then used MEME suite version 5.0.4 to analyse the sequence characteristics of the *ZlARF* protein sequences. The results showed that there were 20 independent motifs in the 33 *ZlARFs* gene. We named these motifs 1–20 and used different colours to distinguish the motifs (Fig. 3). We found that the amino acid sequences of all the motifs were highly conserved, a finding consistent detected *Arabidopsis thaliana* and *Oryza sativa* ARF families [26, 40, 41]. The DBD was composed of motifs 1, 2 and 10; the MR domain comprised motifs 4, 6, 9 and 11; and the CTD domain comprised motifs 7, 8 and 13. All the predicted *ZlARF* protein sequences had common motifs 1, 2, 4, 5, 6, 9, 10 and 11 (Fig. 3A, Fig. S2). cis-element analysis showed that several promoters contained defines and stress responsiveness, auxin-responsive element, zein metabolism regulation, meristem expression, abscisic acid responsiveness and gibberellin-responsiveness (Table S1, Fig. 3B), indicating the roles of several *ZlARFs* genes in auxin-responsive element and other plant hormone-responsive.

Comparison of phylogenetic tree branches and the relationships of AtARFs, OsARFs **and** ZlARFs

To investigate the evolutionary relationships between *ZlARFs* gene in *Z. latifolia* and we compared 33 *ZlARFs*, 23 *AtARF* and 25 *OsARF* genes and constructed phylogenetic trees (Fig. 4, Table S2). The phylogenetic trees showed that the *ARF* gene families in *Z. latifolia, Oryza sativa* and *Arabidopsis* could be divided into 6 groups, designated group -. (Fig. 4). Group and Group constituted the largest clades, containing a total of 39 ARFs, and accounted for 19.75 and 28.40% of the sequences, respectively (Fig. 4B). Interestingly, the ZlARF protein numbers in Groups , , , and were almost the same (Fig. 4B), indicating that these ARF genes from the three species plants may have come from a common ancestor. Furthermore, we also found that the number of ARF in *Oryza sativa* was almost the same as that of *Z. latifolia*. However, Arabidopsis produced an aggregation rate of up to 75.57% in the group (Table S2, Fig. 4B).

In the Fig. 4B, the phylogenetic tree indicated that there were 6 *ZlARFs* gene in group (43.75%, 6 of 16), while only 3 *AtARFs* gene and 6 *OsARFs* gene (24.00%, 6 of 25) were assigned to this group. The other groups were determined in the same manner: in group , there were 2 *AtARFs* (8.70%, 2 of 23) and 4 *OsARFs* (16.00%, 4 of 25); in group , there were 2 *AtARFs* (8.70%, 2 of 23) and 5 *OsARFs* (20.00%, 5 of 25); in group , there were 11 *AtARFs* (47.83%, 11 of 23) and 1 *OsARF* (4.00%, 1 of 25); in group , there was 1 *AtARF* (4.35%, 1 out of 23) and 1 *OsARF* (4.00%, 1 of 25); in group , there were 4 *AtARFs* (17.39%, 4 of 23) and 8 *OsARFs* (20.00%, 5 of 25) (Fig. 4, Table S2). The results also indicate that most of the *ZlARFs* share high homology with *OsARFs* family members than they do with *AtARFs* family members, as shown by the evolutionary tree (Fig. 4). For instance, *ZlARF13, 24, 25, 33* were clustered with *OsARF8, 13, 18, 22* in group , while *AtARF10, 16, 17* were clustered into a separate clade. The outermost histogram shows the number of amino acids. The amino acid peptide is 354 (*ZlARF29*)-1140 *(AtARF15)*. The ZlARF proteins showed wide variation in their length (Fig. 4A). However, the total number of amino acids was concentrated between 500 and 1000, ratio of 78.79%. Further analysis of the composition results by orthologous gene pairs. Every group was further subdivided into 16, 11, 13, 14, 4, 23 members that formed 6, 3, 5, 4, 1, 8 orthologous gene pairs. Finally, 81 ARF members formed 27 orthologous gene pairs, among 17 ZlARF-OsARF orthologous gene pairs. This indicates that ZlARFs was more closely related to OsARFs.

Analysis of subcellular localization, synteny, Ka/Ks values and divergence times of the ZlARFs

ARF protein sequences subcellular localization based on WoLF PSORT. In the current study, the N-terminal signal peptide prediction for the 33 *ZlARFs* was performed using the WoLF PSORT signal peptide prediction program. According to the prediction results, we found that the *ZlARFs* family is located mainly in the nucleus with a proportion of 76% (Fig. 5A). *ZlARF1* was located in mitochondria. *ZlARF6*, *ZlARF14*, *ZlARF22*, and *ZlARF30* were located in chloroplasts. *ZlARF20*, *ZlARF29*, and *ZlARF33* were located in the cytoplasm, and the other 25 *ZlARFs* were located in the nucleus.

The mapping of the *ZlARF* genes loci showed that an inconsistent distribution of the genes with only scaffold information, chromosomal was incomplete (Table 1). We rebuild all scaffold information into one class for collinear analysis. Further analyses of ARF gene evolution and divergence times among *Z. latifolia, Arabidopsis* and *Oryza sativa* showed that a total of 57 orthologous gene pairs exhibited a collinear relationship (10 *Z. latifolia - Z. latifolia*, 47 *Z. latifolia - Oryza sativa*; Fig. 5, Fig. 6, Table S3). These results demonstrated that the *ARF* genes of *Z. latifolia* and *Oryza sativa* appeared to be derived from a common ancestor and that the function of these *ARF* genes of *Z. latifolia* plants might be the same as those of *Oryza sativa*. In addition, among the orthologous gene pairs, each *OsARF* gene presented 1–3 *ZlARF* orthologous genes (Fig. 6, Table S3), suggesting that a few *ZlARF* genes had been duplicated by genome. The Ka/Ks values of these gene pairs were all less than 0.45 except for three pair (*ZlARF28-LOC_Os11g32070.1, ZlARF17-OsARF10*, *Zlat_10014903-OsARF13*, Ka/Ks = 0.65, 0.54, 0.67), and the average divergence times were estimated to be 12.17 million years ago (Mya, *Z. latifolia - Z. latifolia*) and 4.16 Mya (*Z. latifolia - Oryza sativa*) (Table S3, Fig. 5B, C). These results demonstrated that the ARF gene pairs shared between *Z. latifolia* and *Z. latifolia*, *Z. latifolia* and *Oryza sativa* had undergone strong purifying selection with limited functional divergence after whole-genome duplication.

Expression patterns of the ZlARF **genes in swollen stem formation**

To investigate the physiological roles of the *ZlARF* genes, the real-time PCR technique was used to detect the spatial expression of individual members of the gene family. The accumulation of the transcriptional products of 33 *ZlARF* genes in the before and after stem formation were evaluated (Fig. 7). Most *ZlARFs* gene were up/down-regulated, indicating that they might play a central role in swelling stem formation.

The results showed that the transcriptional significant of the *ZlARF* genes varied greatly in before and after stem formation, suggesting that the *ZlARF* genes had multiple functions in *Z. latifolia* stem formation and development. 16 *ZlARF* genes were remarkable expressed after stem formation. Meanwhile, the expression of 8 ARFs were significantly up-regulated and 8 down-regulated. The *ZlARFs* gene is widely expressed throughout the process of *Z. latifolia* stem formation, where it may play a certain role in many aspects of swelling stem formation.

idopsis and *Oryza sativa*.

GO, KEGG and interacting network analysis of ZIARFs

All ARF gene members were annotated for identification of GO term and encyclopedia of genes and genomes (KEGG) analysis (Fig. 8). GO analysis indicated that 33 ARF genes were enriched in DNA binding, nucleus, regulation of transcription, DNA-templated, auxin-activated signalling pathway (GO: 0003677, GO: 0005634, GO: 0006355, GO: 0009734); KEGG analysis showed that four genes were annotated plant hormone signal transduction (ko04075). Although only four genes were involved in signalling pathways for plant hormone signal transduction. However, the GO analysis found that all the genes were annotated with auxin-activated signalling pathway (GO:0009734) (Fig. 8A, Table S5). The results showed that *ZIARFs* may participate in the activation of auxin signal to further stimulate the expansion and growth of stem in *Z. latifolia*. Cellular component found that the ARF gene is located in the nucleus, which is consistent with the previous subcellular prediction. To gain further insight into the roles of these *ZIARF* genes in stem formation, interaction networks of the *ZIARF* genes were constructed. From the STRING database, 23 co-expressed genes were identified by compare to *Oryza sativa* genome (Fig. 8B). Most of the *ZIARF* genes interact with IAA genes (Aux/IAA family). These genes may together to regulate stem formation and expand.

Comparative transcriptome analysis of ZIARF genes in Z. latifolia stem formation

To determine the characteristics of all *ZIARFs* gene according to their expression before and swollen stem formation in *Z. latifolia*, we investigated the expression patterns of the genes using second-generation RNA-Seq technology. The percentages of the reads aligned to the genome were, on average, greater than or equal to 80%, which signifies both the quality of the libraries and the relative completeness of the *Z. latifolia* genome. The RNA-Seq data from stem formation, TDF-treatment and *U. esculenta* infection of *Z. latifolia* showed that the ZIARF members exhibited variable expression patterns; several genes were broadly expressed in all different treatment, while some members were exhibited a more specific pattern (Fig. 9). The qRT-PCR results from stem formation confirmed the RNA-Seq data. In our study, 10 up-regulation in stem formation among 4 log2FC value greater than 0.5; 13 down-regulation in TDF-treatment among 3 log2FC value greater than 0.5; 16 down-regulation in *U. esculenta* infection among 3 log2FC value greater than 0.5. The length of stem formation of *Z. latifolia* was measured (Fig. 9A). The stem had not yet formed after TDF (triadimefon) treatment. The stem length of normal *Z. latifolia* was faster than that of *U. esculenta* infected male *Z. latifolia*. By microscopic observation, TDF treatment can obviously inhibit the formation of *U. esculenta*. *U. esculenta* infection can make male *Z. latifolia* swollen stem formation (Fig. 9B, C, D). Together, these results suggested that some ZIARF genes may play important roles in stem formation after *U. esculenta* infection. ARF genes family may be key to stem formation of *Z. latifolia*. Functional verification of these genes will be the focus of our follow-up research.

Discussion

Auxin response factor plays an important role in swelling stem formation

It is well known that *Z. latifolia* has high levels of nutrient and protein content, primarily due to cell division and swelling stem elongation [42]. Auxin involve in fungal infection and host defines responses are crucial events involved in plant–fungus interactions [43]. Auxin is an important plant hormone and because control of cell expansion and proliferation is closely related to the *ARF* family of plant hormone signalling pathways [44]. We need to understand the role of these hormones in the gene regulation network during auxin-regulated development, and these genes can play valuable roles during genetic evolution and improvement [45]. Exogenous auxin can increase the yield and photosynthesis capacity in stem formation. lead to the increase of endogenous IAA content, so it is of great significance to study the expression of *ARF* family genes [8]. Base on qRT-PCR, we found that the relative expression levels of many genes of *ARF* genes were significantly expression (Fig. 7). To elucidate the role of the *ZIARF* gene in growth regulation, auxin signal transduction and plant hormone response in *Z. latifolia*. We described the main structural characteristics and relative expression model of the *ARF* gene families by bioinformation.

Comprehensive analysis of ARF genes in stem formation of *Z. latifolia*

Through our research, a search for *ARF* genes in the *Z. latifolia* genome resulted in the identification of 33 members (Table 1). Among the 33 *ZIAEFs*, 16 members not included CTD domain, which generally lacked partial sequences in flanking region. All of these proteins are nucleus proteins (Fig. 5A) and exhibit similar patterns of the main motif distribution base on phylogenetic analysis (Fig. 3A). Subcellular localization is the main content of our future validation. We found that the nucleus is one of the sites with the greatest concentration of *ZIARF*. Conserved C-terminal region includes a CTD domain that can form ARF/Aux-IAA heterodimers and ARF/ARF heterodimers and ARF/ARF homodimers through a phytohormone auxin reaction [46, 47]. This finding is in accordance with the results of an *Oryza sativa* [26] analysis, suggesting that these *ZIARFs* may be regulated in an auxin-independent manner. Meanwhile, we compared the evolutionary relationship, synteny, Ka/Ks values and divergence times of *ARF* genes in *Z. latifolia, Oryza sativa* and *Arabidopsis,* and found that the genes in *Z. latifolia* and *Oryza sativa* were the most similar (Fig. 4). Based on phylogenetic analysis, the intron and exon distributions were found to be basically the same in in the *ZIARFs* gene of the same group (Fig. 2). In other words, the distribution of a gene family within a gene structure can also reflect the relationship between phylogenetically related genes [30, 48].

Relevant functions and expression patterns analysis

Our analysis of the genetic structure and motifs in *ZIARF* will help with understanding its different roles in swelling stem formation. The feature of these motif and *cis*-elements was important for predicting the transcription factor function[17]. In addition, the motif distributions of some closely related sequences are similar, suggesting that these *ZIARF* families may have similar functions (Fig. 3). In general, the similarity of most *ZIARFs* gene structures and motif compositions in each group supports the findings from the phylogenetic analysis.

All predicted members of the *ZIARF* families were divided into six different colour-coded groups (group 1-group 6) (Fig. 4), in a manner similar to the classification scheme used for *Arabidopsis* [49] and *Oryza sativa* [26]. Interestingly, the ZIARF protein numbers in Groups □, □, □ and □ were almost the same

(Fig. 4B), indicating that these ARF genes from the three species plants may have come from a common ancestor [50]. These results revealed that the ARF gene family is affected by polyploidy and has experienced segmental and genome duplication in *Z. latifolia* species.

In the evolutionary analysis, we found that each group had different genes from *Oryza sativa* and *Arabidopsis*, indicating that the *ARF* gene was mutated before the three species evolved and diversified. Further evolutionary analysis showed that there were 27 orthologous gene pairs combined in *Z. latifolia* and *Oryza sativa*. In the synteny analysis, there were 57 orthologous gene pairs, including the 47 *ZlARF-OsARF* pairs and 10 *ZlARF-ZlARF* (Fig. 5, Fig. 6). This result shows that the relationship between *Z. latifolia* and *Oryza sativa* is closer than that between *Z. latifolia* and *Arabidopsis*, and the *ARF* family was better preserved during the evolution of these three plants. In addition, we found that the *OsARFs* family was evenly distributed, and in *Arabidopsis*, they were mainly found in group Ⅱ gene clusters. This suggests that the genes in this population may have been acquired by *Z. latifolia* or may have been lost from the *Oryza sativa* lineage [51] after differentiation of the common ancestor of *Oryza sativa* and *Z. latifolia* [52]. Additionally, we estimated the divergence times of the orthologous ARF pairs between *Z. latifolia- Z. latifolia* and *Z. latifolia- Oryza sativa,* which showed Ka/Ks ratios less than 0.45 and indicating the occurrence of strong purifying selection (Fig. 5, Table S3), reflecting the strong control exerted over these genes in evolution [53]. Moreover, the average divergence times of *Z. latifolia- Z. latifolia* and *Z. latifolia- Oryza sativa* were estimated to be 12.70 and 4.16 Mya, respectively, indicating that *Z. latifolia* and *Oryza sativa* shared a common ancestor and exhibit different differentiation times [54]. These results imply that the ARF gene family might be a good candidate molecular reference for analysis of plant evolution.

Our study found that most duplicated *ZlARFs* were expressed in different period and treatment, suggesting that these genes had specific or cellular functions. According to GO and analysis, most of them are located in transcription, DNA-templated, auxin-activated signalling pathway, nucleus and plant hormone signal transaction [14, 15] (Fig. 8). PPI showed that ARF and IAA could together to improve plant development [55] and stem formation [56]. In the current study, we used qRT-PCR and RNA-Seq to explore the expression profile of the *ZlARF* gene before and after swelling stem formation in *Z. latifolia*. (Fig. 7, Fig. 9). The expression of the *ZlARF* gene during *Z. latifolia* development was rapidly changed compared to that before stem formation in *Z. latifolia*. Among stem formation, the gene expression levels of different 33 *ZlARF* genes clearly differed, and some of these genes were very highly and specifically expressed. These results were similar to those obtained in peach [57] and sweet orange [58], suggesting that some *ARF* genes play important roles in stem developmental processes. The expression profile of a gene is often related to its function. Since ARFs regulate auxin biological processes based on gene expression, we hypothesized that the *ZlARF* gene responds to the formation of swollen stem.

Then, we assayed 6 RNA-Seq from stem formation, TDF treatment and *U. esculenta* infection, and found that 33 ZlARF genes were expressed. Among them, the down-regulated expressions of *ZlARF9, ZlARF19* and *ZlARF30* were found stem formation and *U. esculenta* infection, while the up-regulated expressions were found after TDF treatment (Fig. 9). The results showed that the symbiosis between *Z. latifolia* and *U. esculenta* could mediate the expression of ARF gene and influence the stem formation [59]. However, further research is required to investigate the detailed regulatory mechanisms of *ZlARF9, ZlARF19* and *ZlARF30* related to induced *U. esculenta* stimulating stem formation of *Z. latifolia*. In summary, 33 *ZlARF* gene family was identified based on genome-wide, and related functions were predicted by bioinformatics analysis. These findings provide a solid foundation for an insight into the potential function of the *ZlARF* gene. Comprehensive analysis was helpful to explore *ARF* genes for further functional identification and the yield provides a theoretical basis of *Z. latifolia*.

Conclusions

In this study, we identified 33 *ZlARF* gene family members in *Z. latifolia* species. Base on the *Z. latifolia* genomic database, we have gained new insight and comprehensively evaluate about the gene structures, *cis*-elements, synteny and evolutionary relationships of the *Z. latifolia ARF* genes. Additionally, expression pattern assays based on RNA-Seq and qRT-PCR data revealed that *ZlARFs* present after stem formation patterns and are involved in stem growth and expand, especially auxin-activated signalling pathway. Combined analyses of RNA-Seq datasets from different of *Z. latifolia* demonstrated that 33 *ZlAREF* genes were up/downregulated in the *U. esculenta* infection, possibly indicating crucial roles in auxin regulation and transport. Furthermore, the co-expression networks assay exhibited that only *ZlARFs* and *IAAs* can be detected the co-expressed genes among these *ARF* genes. These results provided basic information for studying the mechanism of ARF function in the stem formation in *Z. latifolia*. In the future, more experimental and bioinformatics work is needed to fully understand the functions of these important candidate genes and the regulatory mechanisms of some important stem formation action-related genes in this particular species.

Abbreviations

ORF: Open Reading Frame; DBD: B3 DNA-binding domain; DEGs: Different expression genes; ARF: auxin response factor; CTD: C-terminal dimerization domain; aa: amino acid; PI: theoretical isoelectric point; Mw: molecular weight of polypeptide; GH3: Gretchen Hagen 3; SAUR: small auxin up RNA; IAA: Indole-3-acetic acid; ZOOPS: zero or one per sequence; MSA: Multiple sequence alignment, NJ: neighbor joining; qRT-PCR: Quantitative Real-time PCR; PPI: Protein-protein interaction; AuxREs: auxin response elements; KEGG: Encyclopedia of Genes and Genomes; GO: Gene Ontology.

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: All authors agree to publish.

Availability of data and materials: The raw RNA-Seq data used in this study have been deposited in the Nation Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database under the accession number SRP212030 (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?study=SRP212030).

Consent for publication: Not applicable.

Funding: This work was supported by Anhui Province Key Research and Development Program (201904a06020057), Anhui Province Science and Technology Project (1304032040), Natural Science Foundation of Anhui Province (2008085MC80) and Anhui Provincial Education Department Natural Science Research Key Project (KJ2019A0188). This funding body has no role in the design of the study; collection, analysis and interpretation of data; and inwriting the manuscript

Author ' contributions: SD and JL designed the research, analysed the data, and wrote the manuscript. JF, ZJ, YY and LY performed the integration analyses of the data. CG and GH provided some resources. YW, DF and HW assisted in completing the experiments. All authors discussed the data and reviewed and commented on the manuscript.

Acknowledgements: We are grateful to the Researcher Chenggang Wang, Dr. Shidong Zhu (Anhui agricultural university) for providing assistance in accomplishing this work. We are grateful to Zhejiang University for providing the genome-wide data platform.

References

1. Kelley LA, Sternberg MJE. **Protein structure prediction on the Web: a case study using the Phyre server**. Nature Protocols. 2009; **4**(3):363-371.
2. Bruce SA, Saville BJ, Emery RJN. **Ustilago maydis Produces Cytokinins and Abscisic Acid for Potential Regulation of Tumor Formation in Maize**. J Plant Growth Regul. 2011; **30**(1):51-63.
3. Guo L, Qiu J, Han Z, Ye Z, Chen C, Liu C, Xin X, Ye CY, Wang YY, Xie H. **A host plant genome (Zizania latifolia) after a century-long endophyte infection**. Plant Journal. 2015; **83**(4):600-609.
4. Guo HB, Li SM, Peng J, Ke WD. **Zizania latifolia Turcz. cultivated in China**. Genetic Resources and Crop Evolution. 2007; **54**(6):1211-1217.
5. Yan N, Du YM, Liu XM, Chu MJ, Shi J, Zhang HB, Liu YH, Zhang ZF. **A comparative UHPLC-QqQ-MS-based metabolomics approach for evaluating Chinese and North American wild rice**. Food Chemistry. 2019; **275**:618-627.
6. Han SF, Zhang H, Zhai CK. **Protective potentials of wild rice (Zizania latifolia (Griseb) Turcz) against obesity and lipotoxicity induced by a high-fat/cholesterol diet in rats**. Food Chem Toxicol. 2012; **50**(7):2263-2269.
7. Yan N, Du Y, Liu X, Chu C, Shi J, Zhang H, Liu Y, Zhang Z. **Morphological Characteristics, Nutrients, and Bioactive Compounds of Zizania latifolia, and Health Benefits of Its Seeds**. Molecules. 2018; **23**(7):1561.
8. Li J, Guan Y, Yuan L, Hou J, Wang C, Liu F, Yang Y, Lu Z, Chen G, Zhu S. **Effects of exogenous IAA in regulating photosynthetic capacity, carbohydrate metabolism and yield of Zizania latifolia**. Scientia Horticulturae. 2019; **253**:276-285.
9. Chanclud E, Morel JB. **Plant hormones: a fungal point of view**. Mol Plant Pathol. 2016; **17**(8):1289-1297.
10. Nguyen CT, Tran GB, Nguyen NH. **Homeostasis of histone acetylation is critical for auxin signaling and root morphogenesis**. Plant Mol Biol. 2020; **103**(1-2):1-7.
11. Chung KR, Tzeng DD. **Biosynthesis of indole-3-acetic acid by the gall-inducing fungus *Ustilago esculenta***. J Biol Sci. 2004; **4**(6):744-750.
12. Woodward AW, Bartel B. **Auxin: Regulation, Action, and Interaction**. Ann Bot. 2005; **95**(5):707-735.
13. Chandler JW. **Auxin response factors**. Plant Cell Environ. 2016; **39**(5):1014-1028.
14. Weijers D, Nemhauser J, Yang Z. **Auxin: small molecule, big impact**. Journal of Experimental Botany. 2018; **69**(2):133-136.
15. Roosjen M, Paque S, Weijers D. **Auxin Response Factors: output control in auxin biology**. Journal of Experimental Botany. 2018; **69**(2):179-188.
16. Boer DR, Freire-Rios A, van den Berg WA, Saaki T, Manfield IW, Kepinski S, Lopez-Vidrieo I, Franco-Zorrilla JM, de Vries SC, Solano R *et al.* **Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors**. Cell. 2014; **156**(3):577-589.
17. Guilfoyle TJ, Hagen G. **Auxin response factors**. Curr Opin Plant Biol. 2007; **10**(5):453-460.
18. Powers SK, Holehouse AS, Korasick DA, Schreiber KH, Clark NM, Jing H, Emenecker R, Han S, Tycksen E, Hwang I *et al.* **Nucleo-cytoplasmic Partitioning of ARF Proteins Controls Auxin Responses in Arabidopsis thaliana**. Mol Cell. 2019; **76**(1):177-190 e175.
19. Korasick DA, Westfall CS, Lee SG, Nanao MH, Dumas R, Hagen G, Guilfoyle TJ, Jez JM, Strader LC. **Molecular basis for auxin response factor protein interaction and the control of auxin response repression**. Proc Natl Acad Sci U S A. 2014; **111**(14):5427-5432.
20. Li Y, Liu ZB, Shi X, Hagen G, Guilfoyle TJ. **An auxin-inducible element in soybean SAUR promoters**. Plant Physiol. 1994; **106**(1):37-43.
21. Ulmasov T, Liu ZB, Hagen G, Guilfoyle TJ. **Composite structure of auxin response elements**. Plant Cell. 1995; **7**(10):1611-1623.
22. Hagen G, Guilfoyle T. **Auxin-responsive gene expression: genes, promoters and regulatory factors**. Plant Mol Biol. 2002; **49**(3-4):373-385.
23. Ulmasov T, Hagen G, Guilfoyle TJ. **ARF1, a transcription factor that binds to auxin response elements**. Science. 1997a; **276**(5320):1865-1868.
24. Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y. **Solution structure of the B3 DNA binding domain of the Arabidopsis cold-responsive transcription factor RAV1**. The Plant Cell. 2004; **16**(12):3448-3459.
25. Tiwari SB, Hagen G, Guilfoyle T. **The roles of auxin response factor domains in auxin-responsive transcription**. Plant Cell. 2003; **15**(2):533-543.
26. 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.
27. Okushima Y, Overvoorde PJ, Arima K, Alonso JM, Chan A, Chang C, Ecker JR, Hughes B, Lui A, Nguyen D *et al.* **Functional genomic analysis of the auxin response factor gene family members in Arabidopsis thaliana: unique and overlapping functions of ARF7 and ARF19**. Plant Cell. 2005; **17**(2):444-463.

28. Ellis CM, Punita N, Young JC, Gretchen H, Guilfoyle TJ, Reed JW. **AUXIN RESPONSE FACTOR1 and AUXIN RESPONSE FACTOR2 regulate senescence and floral organ abscission in Arabidopsis thaliana**. Development. 2005; **132**(20):4563-4574.
29. Schruff MC, Melissa S, Sushma T, Sally A, Nick F, Scott RJ. **The AUXIN RESPONSE FACTOR 2 gene of Arabidopsis links auxin signalling, cell division, and the size of seeds and other organs**. Development. 2006; **133**(2):251-261.
30. Rahul K, Tyagi AK, Sharma AK. **Genome-wide analysis of auxin response factor (ARF) gene family from tomato and analysis of their role in flower and fruit development**. Mol Genet Genomics. 2011; **285**(3):245-260.
31. Tombuloglu H. **Genome-wide analysis of the auxin response factors (ARF) gene family in barley (Hordeum vulgare L.)**. J Plant Biochem Biot. 2018;(1):1-11.
32. Wan S, Li W, Zhu Y, Liu Z, Huang W, Zhan J. **Genome-wide identification, characterization and expression analysis of the auxin response factor gene family in Vitis vinifera**. Plant Cell Rep. 2014; **33**(8):1365-1375.
33. Wu J, Wang F, Cheng L, Kong F, Peng Z, Liu S, Yu X, Lu G. **Identification, isolation and expression analysis of auxin response factor (ARF) genes in Solanum lycopersicum**. Plant cell rep. 2011; **30**(11):2059.
34. 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.
35. Yoko O, Overvoorde PJ, Kazunari A, Alonso JM, April C, Charlie C, Ecker JR, Beth H, Amy L, Diana N. **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):444-463.
36. Zhao Y, Weng QY, Hai-Lian MA, Song JH, Yuan JC, Wang LY, Dong ZP, Liu YH. **Genome-Wide Identification and Bioinformatics Analysis of ARF Gene Family in Foxtail millet Setaria italica**. J Plant Genetic Resources. 2016.
37. Galli M, Khakhar A, Lu ZF, Chen ZL, Sen S, Joshi T, Nemhauser JL, Schmitz RJ, Gallavotti A. **The DNA binding landscape of the maize auxin response factor family**. Nature Communications. 2018; **9**:14.
38. Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE *et al*. **RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome**. Bmc Plant Biology. 2010; **10**:16.
39. Wu W, Liu Y, Wang Y, Li H, Liu J, Tan J, He J, Bai J, Ma H. **Evolution analysis of the *Aux/IAA* gene family in plants shows dual origins and variable nuclear localization signals**. Int J Mol Sci. 2017; **18**(10):2107.
40. 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):742.
41. Xiao L, Mei S, Rui X, Huai S, Jia W, Shi, Z. **Genomewide identification and expression analysis of the ARF gene family in apple**. Journal Genet. 2014; **93**(3):785-797.
42. Luo J, Zhou JJ, Zhang JZ. **Aux/IAA gene family in plants: molecular structure, regulation, and function**. Int J Mol Sci. 2018; **19**(1):259.
43. Breakspear A, Liu C, Roy S, Stacey N, Rogers C, Trick M, Morieri G, Mysore KS, Wen J, Oldroyd GE *et al*. **The root hair "infectome" of Medicago truncatula uncovers changes in cell cycle genes and reveals a requirement for Auxin signaling in rhizobial infection**. Plant Cell. 2014; **26**(12):4680-4701.
44. Li SB, Xie ZZ, Hu CG, Zhang JZ. **A Review of Auxin Response Factors (ARFs) in Plants**. Front Plant Sci. 2016; **7**(742):47.
45. Wojcik AM, Wojcikowska B, Gaj MD. **Current Perspectives on the Auxin-Mediated Genetic Network that Controls the Induction of Somatic Embryogenesis in Plants**. Int J Mol Sci. 2020; **21**(4):1333.
46. Ouellet F, Overvoorde PJ, Theologis A. **IAA17/AXR3: biochemical insight into an auxin mutant phenotype**. The Plant Cell. 2001; **13**(4):829-841.
47. Ulmasov T, Hagen G, Guilfoyle TJ. **Dimerization and DNA binding of auxin response factors**. Plant J. 1999; **19**(3):309-319.
48. Li X, Duan X, Jiang H, Sun Y, Tang Y, Yuan Z, Guo J, Liang W, Chen L, Yin J. **Genome-wide analysis of basic/helix-loop-helix transcription factor family in rice and Arabidopsis**. Plant physiol. 2006; **141**(4):1167-1184.
49. Remington DL, Vision TJ, Guilfoyle TJ, Reed JW. **Contrasting modes of diversification in the Aux/IAA and ARF gene families**. Plant Physiol. 2004; **135**(3):1738-1752.
50. Gao B, Wang L, Oliver M, Chen M, Zhang J. **Phylogenomic synteny network analyses reveal ancestral transpositions of auxin response factor genes in plants**. Plant Methods. 2020; **16**(1):70.
51. Kellogg EA. **The evolutionary history of ehrhartoideae, Oryzeae, and Oryza**. Rice. 2009; **2**(1):1-14.
52. Tapan K. Mondal RJH: **The wild Oryza genomes**: Springer, Cham; 2018.
53. Chen G, Wang J, Wang H, Wang C, Tang X, Li J, Zhang L, Song J, Hou J, Yuan L. **Genome-wide analysis of proline-rich extension-like receptor protein kinase (PERK) in Brassica rapa and its association with the pollen development**. BMC Genomics. 2020; **21**(1):401.
54. Liu M, Ma Z, Wang A, Zheng T, Huang L, Sun W, Zhang Y, Jin W, Zhan J, Cai Y *et al*. **Genome-Wide Investigation of the Auxin Response Factor Gene Family in Tartary Buckwheat (Fagopyrum tataricum)**. Int J Mol Sci. 2018; **19**(11):3526.
55. Jodder J. **miRNA-mediated regulation of auxin signaling pathway during plant development and stress responses**. J Biosci. 2020; **45**(1).
56. Li SB, Xie ZZ, Hu CG, Zhang JZ. **A Review of Auxin Response Factors (ARFs) in Plants**. Front Plant Sci. 2016; **7**:47.
57. Diao D, Hu X, Guan D, Wang W, Liu Y. **Genome-wide identification of the ARF (auxin response factor) gene family in peach and their expression analysis**. Molecular Biology Reports. 2020; **47**(5).
58. Li SB, OuYang WZ, Hou XJ, Xie LL, Hu CG, Zhang JZ. **Genome-wide identification, isolation and expression analysis of auxin response factor (ARF) gene family in sweet orange (Citrus sinensis)**. Front Plant Sci. 2015; **6**:119.
59. Kunkel BN, Harper CP. **The roles of auxin during interactions between bacterial plant pathogens and their hosts**. J Exp Bot. 2018; **69**(2):245-254.

60. Die JV, Gil J, Millan T. **Genome-wide identification of the auxin response factor gene family in Cicer arietinum**. BMC Genomics. 2018; **19**(1):301.
61. Letunic I, Doerks T, Bork P. **SMART 7: recent updates to the protein domain annotation resource**. Nucleic Acids Res. 2012; **40**(Database issue):D302-305.
62. Crooks GE, Hon G, Chandonia JM, Brenner SE. **WebLogo: a sequence logo generator**. Genome Res. 2004; **14**(6):1188-1190.
63. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. **MEGA X: molecular evolutionary genetics analysis across computing platforms**. Mol Biol Evol. 2018; **35**(6):1547-1549.
64. Letunic I, Bork P. **Interactive Tree Of Life (iTOL) v4: recent updates and new developments**. Nucleic Acids Res. 2019; **47**(W1):W256-W259.
65. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH. **JBrowse: a next-generation genome browser**. Genome res. 2009; **19**(9):1630-1638.
66. Yi T, Qinglong D, Zhirui J, Fumei C, Peihua C, Zongshan Z. **Genome-wide identification and analysis of the MADS-box gene family in apple**. Gene. 2015; **555**(2):277-290.
67. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S. **PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences**. Nucleic Acids Res. 2002; **30**(1):325-327.
68. 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):1194-1202.
69. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H *et al*. **MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity**. Nucleic Acids Res. 2012; **40**(7):e49.
70. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. **Circos: an information aesthetic for comparative genomics**. Genome Res. 2009; **19**(9):1639-1645.
71. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. **KaKs Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies**. Genomics Proteomics Bioinformatics. 2010; **8**(1):77-80.
72. Barter RL, Yu B. **Superheat: an R package for creating beautiful and extendable heatmaps for visualizing complex data**. Stats. 2017; **27**(12):1-30.
73. Faoro F, Maffi D, Cantu D, Iriti M. **Chemical-induced resistance against powdery mildew in barley: the effects of chitosan and benzothiadiazole**. Biocontrol. 2008; **53**(2):387-401.
74. Young MD, Wakefield MJ, Smyth GK, Oshlack A. **Gene ontology analysis for RNA-seq: accounting for selection bias**. Genome Biology. 2010; **11**(2):12.
75. Mao XZ, Cai T, Olyarchuk JG, Wei LP. **Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary**. Bioinformatics. 2005; **21**(19):3787-3793.
76. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P *et al*. **STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets**. Nucleic Acids Res. 2019; **47**(D1):D607-D613.
77. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. Genome Res. 2003; **13**(11):2498-2504.
78. Livak KJ, Schmittgen TD. **Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method**. Methods. 2001; **25**(4):402-408.

Figures

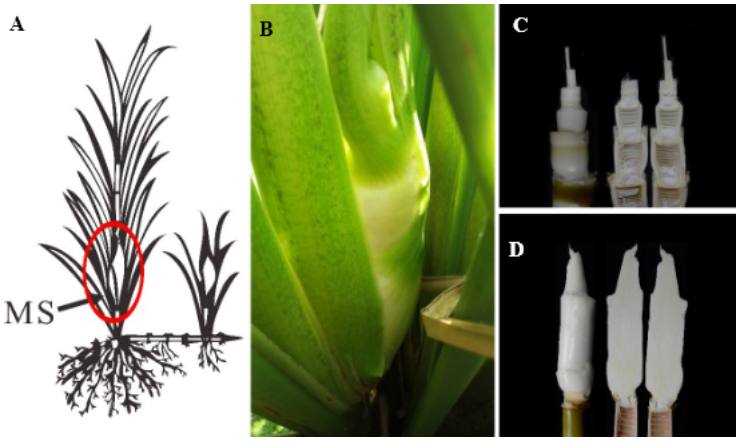

Figure 1

Z. latifolia was selected as the experimental material. (A) Schematic representation of swelling gall formation (GF) on Z. latifolia. (B) The swelling gall image was taken with a camera (Nikon D7100). (C) Z. latifolia before the swollen gall formation (JB_A). (D) Z. latifolia after the swollen gall formation (JB_B).

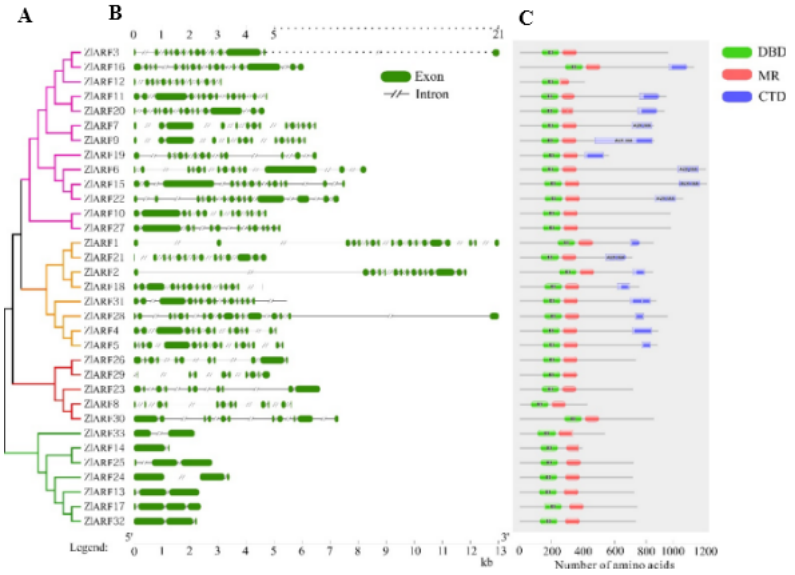

Figure 2

Exon-intron structures and structural domains of the ZIARFs family. (A) The left panel shows the phylogenetic relationships of ZIARFs gene and the intron and exon regions, respectively. (B) Exons-introns are presented as dark-green boxes and thin lines, respectively. (C) The right panel shows the composition and location of the main structural domains. Green represents the DBD domain, red represents the MR domain, and purple represents the CTD domain.

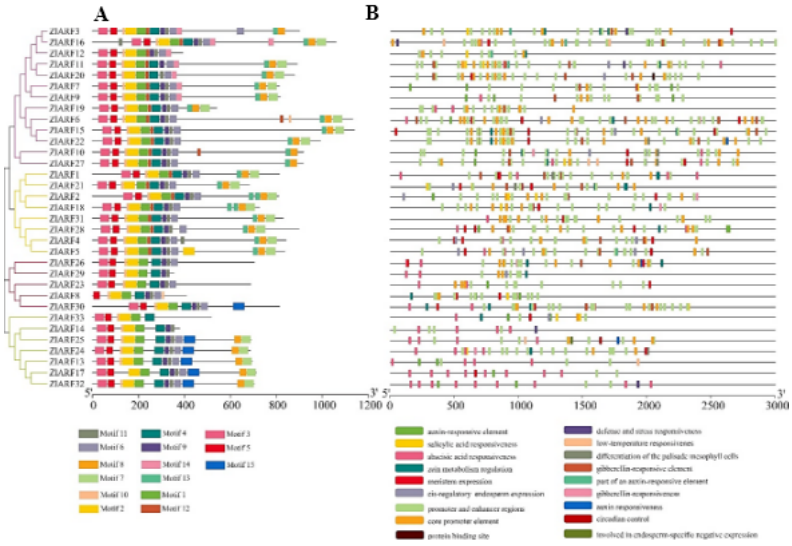

Figure 3

Identification of the conserved sequences and cis-acting regulatory element analysis in 33 ZIARFs. (A) The unrooted tree was built using MEGA-X software. Motifs were determined using MEME suite version 5.0.4. Each motif is shown by a colourful box, with the corresponding motif numbers listed in the lower panel. (B) Different cis-acting regulatory elements are indicated by different colored boxes. All the cis-acting regulatory elements were calculated and their function were annotated.

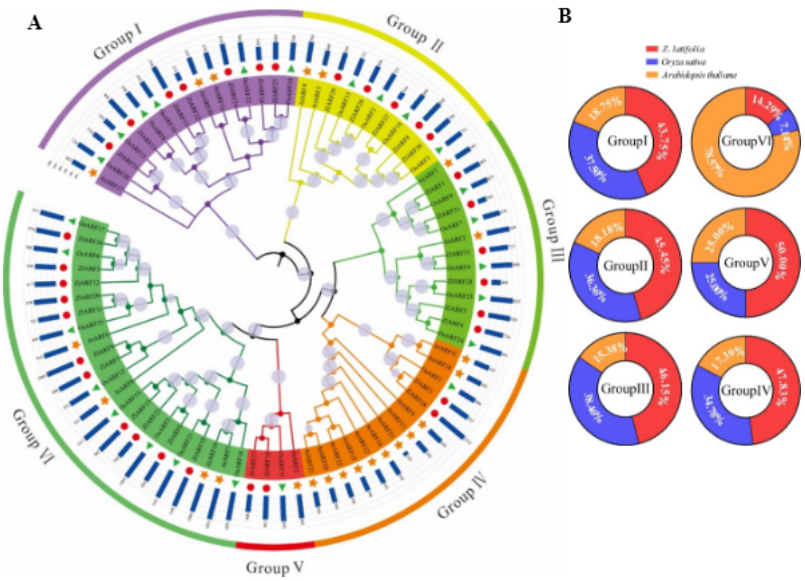

Figure 4

Phylogenetic relationships of all members of the Z. latifolia, Arabidopsis and Oryza sativa ARF families. (A) Using the MEGA-X program, the unrooted tree was constructed by the NJ tree method. Bootstrap tests were used for 1000 replicates. Yellow stars represent Arabidopsis thaliana, green triangles represent Oryza sativa, and red circles represent Z. latifolia, and the histograms showed amino acid sequence length. (B) Number and percentage of ARF proteins across the four plant species within each subfamily.

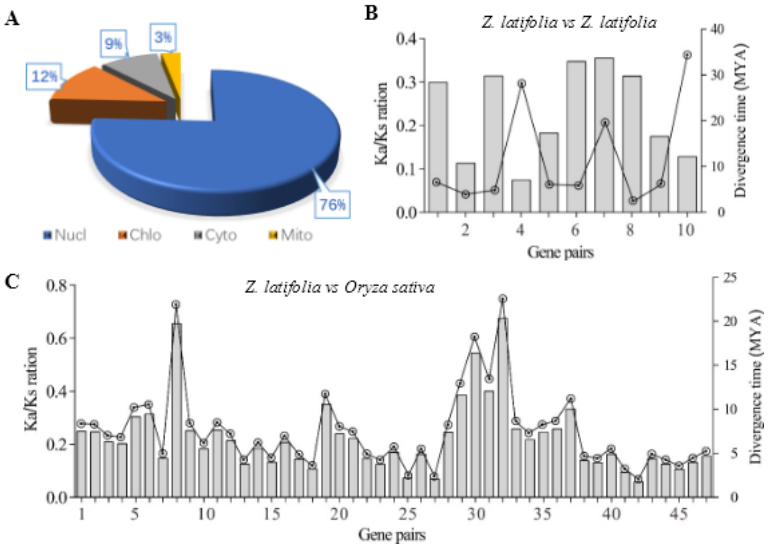

Figure 5

ARF protein sequence subcellular localization, Ka/Ks values and divergence times was predicted. (A) ARF protein sequence subcellular localization as predicted by WoLF PSORT. Nucl: nuclear; Chlo: chloroplast; Cyto: cytoplasmic; Mito: mitochondrial. (B) Ka/Ks values and divergence times of orthologous / paralogous gene pairs. Compare Z. latifolia- Z. latifolia and Z. latifolia -Oryza sativa to find the relevant Ka/Ks values and divergence times.

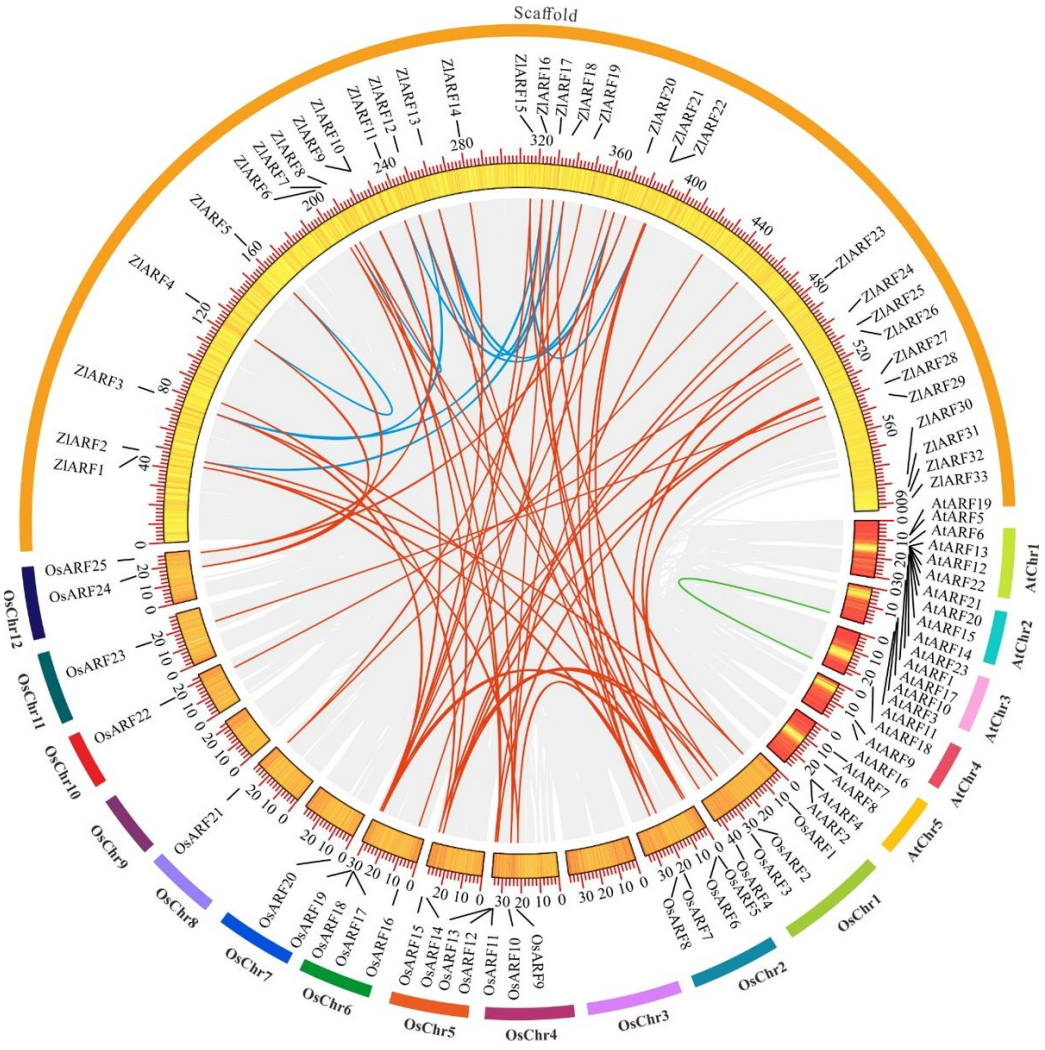

Figure 6

Syntenic analysis and gene expression of ARF genes family in Z. latifolia, Arabidopsis and Oryza sativa species. Interspecific collinearity analysis in Z. latifolia, Arabidopsis and Oryza sativa.

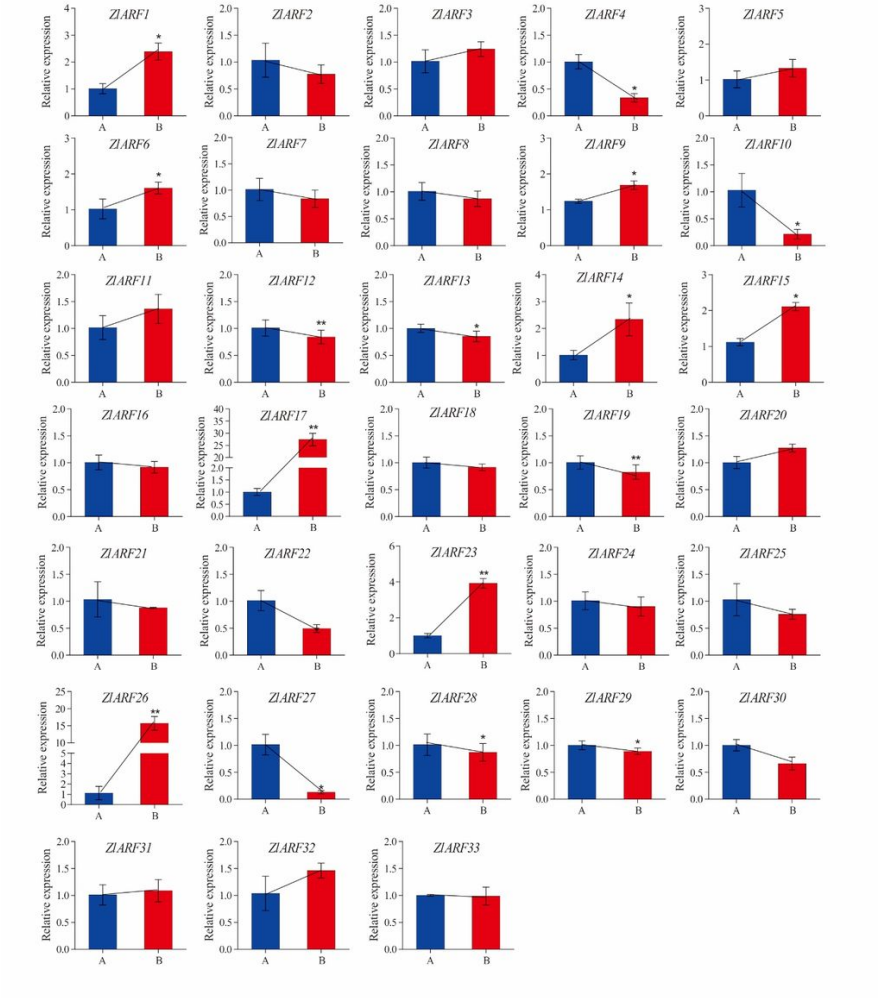

Figure 7

Expression analysis of ZlARFs gene before and after formation of swollen galls in Z. latifolia. A represents the gene level before swollen gall formation; B represents the gene level after swollen gall formation. The experiment included three biological replicates, and SPSS was used for the analysis of significance (Duncan test significant: * $P < 0.05$; very significant: ** $P < 0.01$).

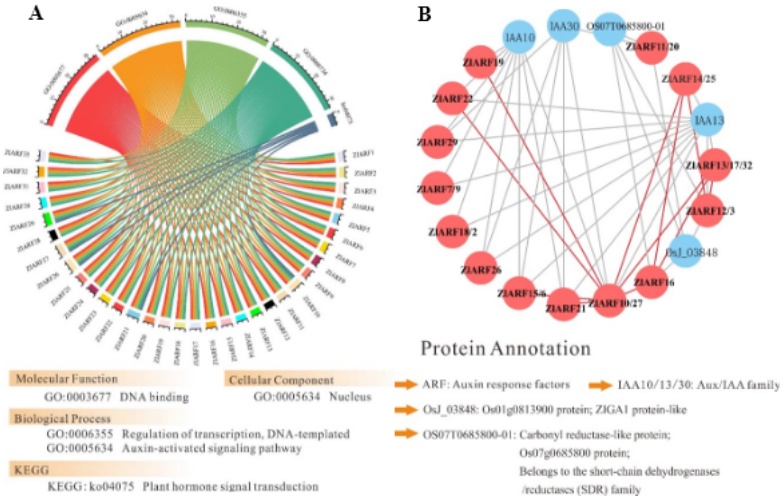

Figure 8

GO term, KEGG enrichment and interacting network was analyzed 33 ARF genes family in Z. latifolia. (A) The numbers of 4 GO term and KEGG enrichment were counted in 33 ZlARFs. (B) Interacting network was analyzed by STRING tools. The identified 33 ZlARFs was shown in red.

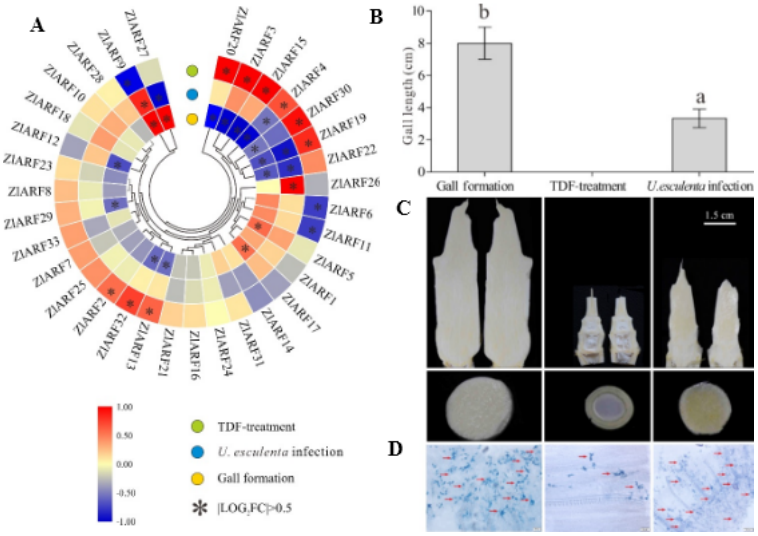

Figure 9

RNA-seq analyses and stain observation of ZIARF genes in gall formation, TDF-treatment and U.esculenta infection of Z. latifolia. (A) Heat map analysis of ZIARFs genes in different treatment; the color bar on the top represents log2 (fold change) values. (B) Different treatment gall length was measured. (C) Observations on the growth of Z. latifolia plants in different treatment. (D) The U. esculenta was observed by aniline blue stain in different samples. The arrows indicate the presence of clustered U. esculenta. Bar = 100 μm.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.docx
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx