BMC Plant Biology

RESEARCH

Open Access

Genome-wide identification and functional analysis of the ARF gene family in tetraploid potato reveal its potential role in anthocyanin biosynthesis

Xia Zhang[1], Rong Fan[1], Zhuo Yu[1], Xinyue Huang[1], Huiting Wang[1], Wenfeng Xu[1] and Xiaoxia Yu[1*]

Abstract

Background Auxin response factors (ARFs) are plant-specific transcription factors that are crucial for flower development, lateral root formation, leaf senescence, and fruit ripening. Information on the ARF family genes in tetraploid potato remains unidentified.

Results In this study, we identified 92 StARF genes including alleles in the tetraploid potato genome (C88.v1), classified into four subfamilies, and unevenly distributed across 48 chromosomes. The promoter regions contained numerous light, plant hormones, and stress response elements, including those for low-temperature, drought, and anaerobic-induction *cis*-elements. Collinearity analysis suggested that StARF family members amplification results from whole genome and segmental duplications. Tissue-specific expression patterns manifested in most *StARF* family genes. RNA-seq data and WGCNA analysis of two tetraploid potato varieties with different-colored tuber flesh identified 11 differentially expressed *StARF* genes correlated with key anthocyanin synthesis genes. Protein-protein interaction predictions highlighted StARF23-1 as a potential key regulator of the anthocyanin biosynthesis pathway, warranting further investigation.

Conclusions Overall, our study comprehensively analyzes the StARF gene family in tetraploid potato and identifies candidate genes linked to anthocyanin synthesis, providing a foundation for future research on the regulatory role of StARF transcription factors in colored potato anthocyanin biosynthesis.

Keywords Tetraploid potato, Auxin response factor (ARF), WGCNA analysis, Flesh, Anthocyanin

*Correspondence:
Xiaoxia Yu
yuxiaoxia1985@sina.com
[1]Agricultural College, Inner Mongolia Agricultural University, Hohhot, Inner Mongolia, China

Introduction

Anthocyanins, a class of water-soluble natural pigments present in the vacuoles of plants, belong to flavonoid secondary metabolites and have a wide range of biological functions in plants. Besides imparting vibrant hues to plant organs, these compounds confer resilience to low temperatures, ultraviolet radiation, and pathogens [1]. From a human health perspective, anthocyanins, which are edible natural pigments, are considered safe and nontoxic and possess strong antioxidant properties. Furthermore, anthocyanins possess a multitude of physiological functions, including prevention of cardiovascular disease, anti-aging, anti-tumor, inhibition of cancer cell growth, and protection of vision [2, 3]. Anthocyanin biosynthesis is precisely and dynamically regulated, and it depends crucially on the synergistic action of two gene types: structural genes and regulatory genes. The structural genes encode various enzymes required in the anthocyanin biosynthetic pathway, such as PAL, CHS, CHI, F3H and DFR. The transcription factors (TFs) encoded by regulatory genes serve as the switches for the expression of these enzyme genes, directly influencing the expression patterns of these structural genes [4]. At present, MYB, bHLH, and WD40 are three known transcription factor families that play a major role in anthocyanin biosynthesis [5]. They can either act independently to regulate the expression of structural genes involved in anthocyanin biosynthesis or combine to form an MBW protein complex that regulates the transcription of structural genes. Besides the aforementioned three TFs, TFs such as bZIP [6], ERF [7], SPL [8], and WRKY [9] also regulate anthocyanin biosynthesis. Additionally, recent evidence has shown that ARFs also play important roles in regulating anthocyanin biosynthesis in plants [10–13].

Auxin response factors (ARFs), as key TFs in plant growth, can specifically bind to TGTCTC auxin response elements (AuxREs) within the promoter regions of target genes, activating or inhibiting the expression of downstream target genes and regulating plant growth and development [14, 15]. The typical ARF protein features three conserved domains [16, 17]. The leading one is the B3-type DNA-binding domain (DBD), sitting at the N-terminus and tasked with identifying AuxREs. The second region is the non-conserved middle region (MR), which is located in the middle of the protein and is responsible for the activation or inhibition of auxin-responsive gene expression. This region is divided into an activation domain (AD) and repression domain (RD). AD is characterized by a high proportion of glutamine (Q), serine (S), and leucine (L) residues, whereas RD predominantly comprises proline (P), S, and glycine (G) residues. In addition, the C-terminal dimerization domain (CTD) encompasses two domains (III and IV), which can form homodimers or heterodimers with ARF and Aux/IAA proteins, facilitating protein-protein interactions [18, 19]. The first ARF transcription factor was identified in *Arabidopsis thaliana* by Uimasov et al. [20], using the yeast one-hybrid method, and was designated as AtARF1. Subsequently, a total of 22 AtARF members were identified in genetic and molecular experiments [21–24]. To date, researchers have carried out genome-wide identification and analysis on ARF gene families across a variety of crops, including rice [25], tomato [26], maize [27], apple [28], eggplant [29], castor bean [30], and soybean [14].

Numerous studies have shown that in plants, ARFs are involved in a variety of biological processes, including auxin signaling, growth and development, abiotic stress responses, and secondary metabolism. For instance, *AtARF6*, *AtARF8* [31], *OsARF19* [32], *PgARF14* and *PgARF53* [33] have been proven to be involved in root development in *Arabidopsis*, rice and *Panax ginseng*. *SlARF5*, *SlARF7*, *SlARF8A* and *SlARF8B* have been verified to mediate fruit development in tomato [34]. *MdARF17* [35], *GmARF16* [36], *OsARF17* and *OsARF25* [37] have been demonstrated to play regulatory roles in the responses of apple to drought stress, soybean to salt stress, and rice to low-temperature stress, respectively. Recent studies have indicated that members of the ARF family may play pivotal roles in regulating anthocyanin biosynthesis. Jiang et al. [10] observed that a loss-of-function mutant of *AtARF2* resulted in a notable reduction in the content of flavonols and proanthocyanidins in seedlings and seeds. Moreover, the expression of key flavonoid biosynthetic genes exhibited a notable decline. However, overexpression of *AtARF2* showed the opposite trend, indicating that *AtARF2* functions as a positive regulator of the flavonoid synthesis pathway. In apple, the transcription factor MdARF13 has been demonstrated to bind to the AuxRE element in the *MdDFR* gene promoter region, inhibiting *MdDFR* expression and reducing anthocyanin accumulation [11]. Overexpression of *MdARF2* has been demonstrated to result in a notable reduction in the expression levels of structural genes involved in anthocyanin biosynthesis, such as *MdDFR*, *MdCHS*, and *MdUFGT*, impeding the accumulation of anthocyanins [12]. Li et al. [13] demonstrated that MdERF3 is a positive regulator of the anthocyanin biosynthesis pathway and that MdARF5-1 can inhibit anthocyanin accumulation by antagonizing MdERF3.

Cultivated potato (*Solanum tuberosum* L.), an autotetraploid (2n = 4x = 48) dicotyledonous plant of the *Solanaceae* family, is one of the important food crops around the world due to its excellent characteristics such as rich nutrition, high yield and wide adaptability [38]. Colored potato represents a distinct category of cultivated potato. The underground storage organs are abundant in the anthocyanins and have higher nutritional value than common varieties with white or yellow flesh. Colored

potato rich in anthocyanins has the potential for development and edible value in the extraction of natural edible pigments, development of safe and efficient antioxidants, and enhancement of human immune function [39]. Therefore, it is highly significant to conduct research on the regulatory mechanism of anthocyanin synthesis in colored potato. However, tetraploid potato cultivars exhibit a highly heterogeneous genome and a more intricate genetic background, which presents substantial challenges in genome assembly and considerably impedes the advancement of genetic analysis and molecular breeding improvement [38]. To the best of our knowledge, genome-wide identification and analysis of ARF family members in tetraploid potato and their involvement in anthocyanin biosynthesis in colored potato are yet to be reported. In 2022, Bao et al. [40] published the complete genome of an autotetraploid potato (C88.v1), comprising 48 high-quality chromosomes. The decoding of this genome has not only significantly advanced scientific research pertaining to tetraploid potato genomics, but also provided a crucial theoretical foundation and data foundation for genetically analyzing and improving anthocyanin traits in colored potato.

In this study, bioinformatics tools were employed to identify StARF family members within the C88.v1 tetraploid potato genome for the first time, followed by a comprehensive analysis of their physicochemical properties, phylogeny, gene structure, gene duplication events, and tissue expression patterns. Furthermore, this study used RNA-seq data from purple and yellow potato flesh to identify differentially expressed *StARF* genes. Candidate *StARF* genes that may be involved in anthocyanin biosynthesis in colored potato were further analyzed using WGCNA. This study offers a theoretical foundation for understanding the mechanism of anthocyanin biosynthesis mediated by the StARF gene family in colored potato and provides candidate genes for the molecular breeding of colored potato.

Results

Identification and protein characterization of StARF family members

After conducting a local BLASTP alignment and HMMER search in the potato genome database, and then verifying conserved domains using NCBI-CDD and SMART, a total of 92 *StARF* genes including alleles were identified. Chromosomal mapping was carried out for these genes. The results showed that the 92 *StARF* genes were distributed across 48 chromosomes in the tetraploid potato (Fig. 1). Based on chromosomal location, the 92 *StARF* gene alleles were designated as *StARF1-StARF24*, with additional -1 to -4 added to the gene name for their alleles. Among them, 23 *StARF* genes had multiple alleles, including 22 genes with 4 alleles, and 1 gene with 3 alleles. The highest number of *StARF* genes was observed on chromosomes 2_2, 2_3, and 2_4, with 4 *StARF* genes each. The remaining chromosomes exhibited a range of 1–3 *StARF* genes. Moreover, the number of StARF family members on each chromosome was not directly proportional to the chromosome length.

The physicochemical properties of StARF proteins were analyzed using ExPASy online software. The results showed that the StARF protein sequences varied from 602 to 1246 aa in length (Table S1). The molecular weight (MW) of the proteins ranged from 66.07 to 141.92 kD. The isoelectric point (pI) ranged from 5.22 to 7.94, with 89% of genes having a pI < 7. It indicates that most of the proteins in question possess acidic characteristics, and therefore are likely to function optimally in an acidic environment. The instability index (II) of all StARF proteins was greater than 40, whereas the grand average of hydropathicity (GRAVY) was negative, indicating that 92 members of the StARF family are unstable hydrophilic proteins. The subcellular localization prediction results indicated that most StARF proteins were localized in the nucleus, with only StARF15-1 and StARF15-3 located on the plasma membrane. It suggests that the StARF family members primarily exert regulatory functions within the nucleus.

Construction of phylogenetic tree and subfamily classification of StARF gene family

Constructing a phylogenetic tree can help us understand the evolutionary relationships and history of gene families. Thus, we used 92 StARF proteins and 23 AtARF proteins to build a phylogenetic tree. The classification of StARF proteins into subfamilies was based on the classification standard of the AtARF proteins in *Arabidopsis thaliana*. These results showed that the 115 ARF proteins could be distinctly classified into five subfamilies (subfamilies I, II, III, IV, and V) (Fig. 2). Subfamily I (AtARF1/2-like) contained 25 StARF proteins and 5 AtARF proteins. Subfamily II (AtARF5/6/7/8/19-like) contained 39 StARF proteins and 5 AtARF proteins. Subfamily III (AtARF3/4-like) contained 8 StARF proteins and 2 AtARF proteins. Subfamily IV (AtARF10/16/17-like) contained 20 StARF proteins and 3 AtARF proteins, while subfamily V was devoid of StARF proteins and represented a distinct branch of eight AtARF proteins.

To delve deeper into the phylogenetic relationship and history of the ARF gene family, we further established a phylogenetic tree with the identified StARF and ARF protein sequences from four reported plants. Analysis of the phylogenetic tree (Fig. S1) revealed that most StARF proteins clustered closely with the ARF proteins of dicotyledonous plants (*Arabidopsis* and tomato). In contrast, most ARF family members from monocotyledonous plants (rice and maize) could not be grouped

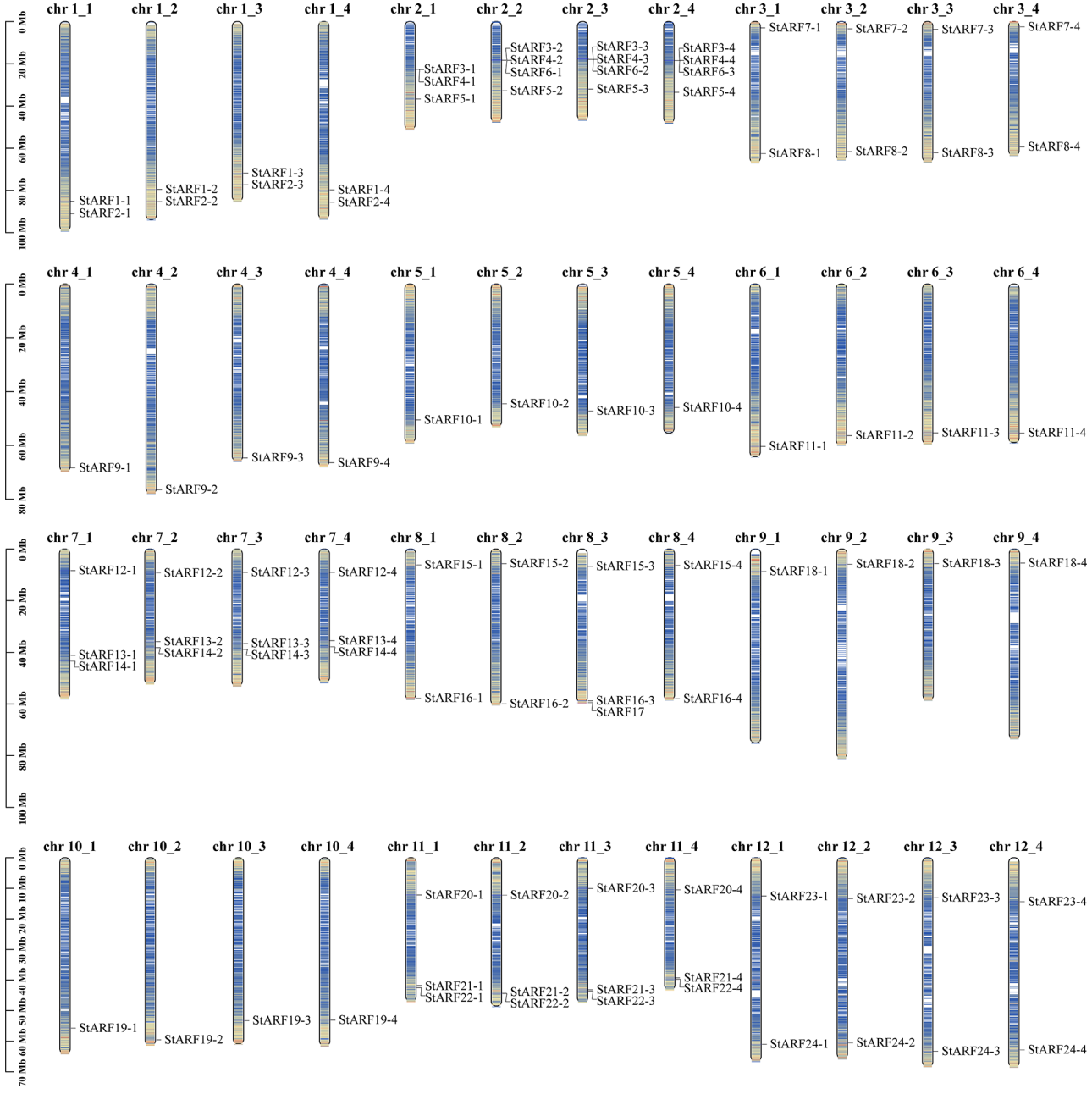

Fig. 1 Chromosomal distribution of the StARF gene family members. The blue rectangular bars represent the chromosomes of tetraploid potato, and the number and length of the chromosomes (Mb) are marked

under the same branch as ARF family members from dicotyledonous plants. These findings further suggest that the ARF proteins of monocotyledonous and dicotyledonous plants evolved in relatively independent directions according to their specific requirements for growth, development, or response to environmental factors during plant evolution.

Gene structure, protein domains and conserved motifs

To explore the structural composition and classification characteristics of the StARF gene family, we conducted an in-depth analysis of the gene structure, conserved motifs, and protein domains of StARF family members. As shown in Fig. S2, most StARF family members had complex and diverse gene structures, consisting of 2 to 18 exons. Phylogenetic tree analysis revealed that StARF members within the same subfamily had highly similar gene structures, while significant differences existed between different subfamilies. For instance, StARF members of subfamily IV exhibited the lowest exon count. Except for *StARF15-1* and *StARF15-3*, all StARF members in subfamily I contained 14 exons. Furthermore, the

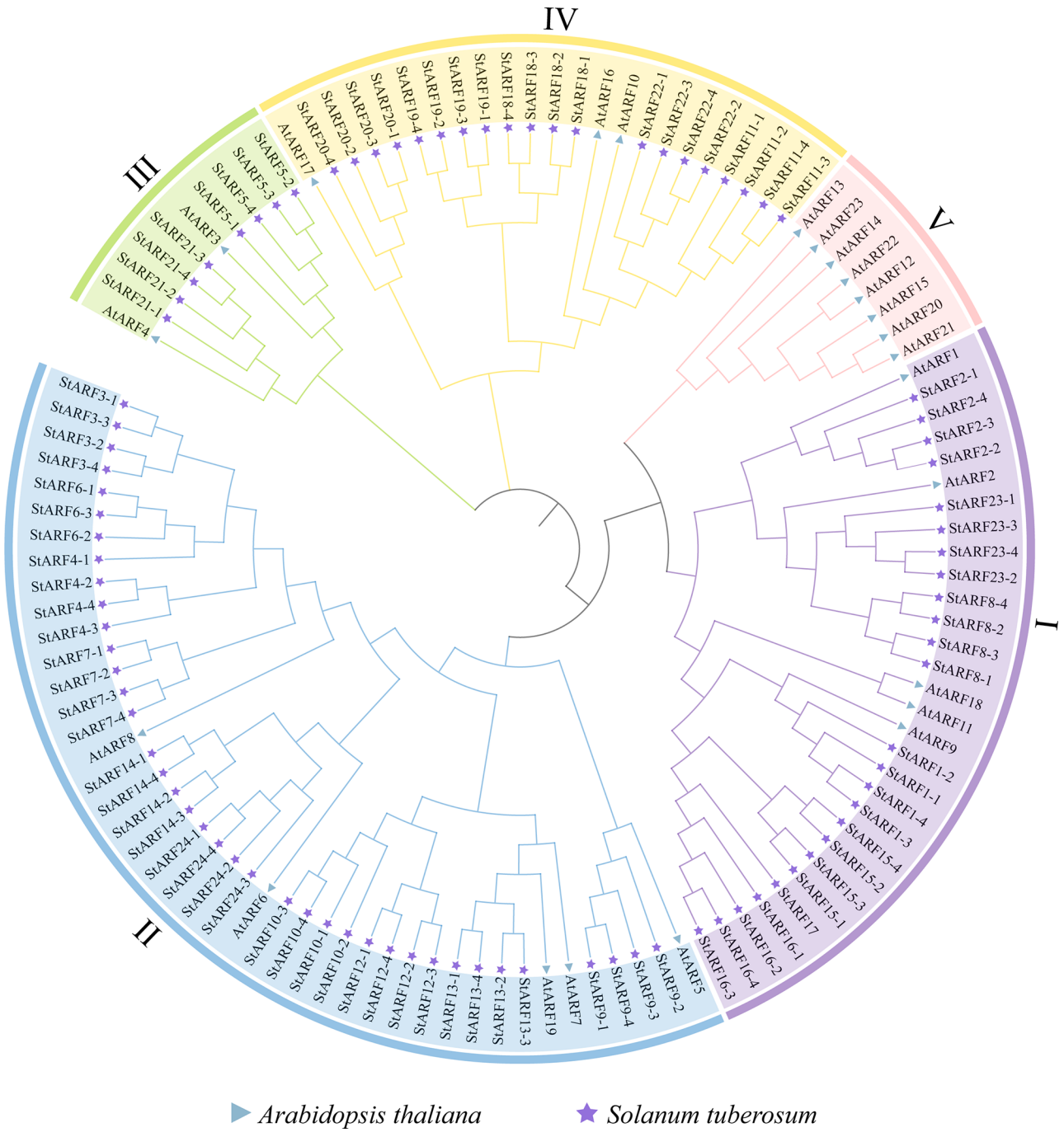

Fig. 2 Phylogenetic tree of ARF gene family in *Arabidopsis thaliana* and tetraploid potato. Different subfamilies are represented by arcs of different colors. Blue triangles and purple pentagrams represent ARF proteins in *Arabidopsis* and potato, respectively

exons in subfamily I were more closely arranged than those in other subfamilies. In addition, some genes, such as *StARF4-1*, *StARF19-1*, and *StARF20*, exhibited relatively long intron sequences or were mis-annotated as a single gene. This may be due to biases in the genomic annotation information and awaits further analysis and research.

We analyzed the conserved domains of the StARF members (Fig. S3C). In total, 55 StARF family members contained the B3, Auxin_resp, and AUX_IAA domains, which included 25 StARF members of subfamily I, 28 StARF members of subfamily II, and 2 StARF members of subfamily IV. Meanwhile, 37 StARF members lacked the AUX_IAA domain while retaining the B3 and Auxin_resp domains, which was observed in 11 members of

Subfamily II, 18 members of Subfamily IV, and all members of Subfamily III. We further conducted a multiple sequence alignment of StARF proteins using DNAMAN 9.0 (Fig. S4). The results showed these three protein domains contained numerous highly conserved amino acid residues. For example, the Auxin_resp domain, within the central functional region, contained amino acids phenylalanine (F), proline (P), glycine (G), serine (S), and tryptophan (W). These highly conserved amino acid sites are essential for the StARF protein function. Furthermore, the amino acid composition and content of this domain exhibited notable differences among distinct StARF subfamilies, indicating potential differential gene functions among various StARF subfamily members.

A total of 10 motifs (Motif 1-Motif 10) were identified in 92 StARF family members through conserved motifs analysis (Fig. S3B, and Table S2). All members of the StARF family exhibited conserved motifs 1, 2, 3, 5, 6, 7, 9, and 10, indicating these eight motifs demonstrated a relatively substantial conservation during the evolutionary history of *StARF* genes. In conclusion, the gene structures and protein domains distributions observed in members of the same subfamily were similar, providing further support for the reliability of the StARF subfamily classification.

Analysis of *cis*-acting elements

Analyzing *cis*-acting elements in gene promoters can help understand the regulation of gene expression and predict the potential functions of genes. We performed a *cis*-acting element analysis on the 2 kb upstream sequences of 92 *StARF* genes, and visualized 36 *cis*-elements. The results showed the presence of a considerable number of light-responsive elements and a multitude of hormone-responsive elements within the promoter regions of the *StARF* genes (Fig. S5), such as G-box, I-box, GARE-motif, CGTCA-motif. Most of the promoter regions of *StARF* genes also contained stress-related responsive elements, including low-temperature responsive elements (LTR), drought-inducible elements (MBS), and anaerobic-induction *cis*-elements (ARE). Furthermore, most of the promoter regions of *StARF* genes exhibited MYB and MYC binding sites, indicating the potential regulation of StARF family member expression by MYB and bHLH transcription factors.

Gene duplication events and collinearity analysis

To gain insight into the amplification of *StARF* genes in tetraploid potato, we analyzed gene duplication events among StARF family members. The results showed that gene duplication events were present in all 92 StARF family members, and 147 collinear pairs were identified, with 17 pairs of nonalleles and 130 pairs of alleles (Fig. 3, and Table S3). No instances of tandem duplications were observed among the identified duplication events. These gene duplication events significantly contributed to the amplification and functional diversification of the StARF gene family members in tetraploid potato. Whole genome duplication (WGD) and segmental duplication events were pivotal driving forces. Furthermore, the Ka and Ks values of 147 collinear pairs were calculated, revealing that most duplicated gene pairs exhibited Ka/Ks ratios of <1 (Table S4). This suggests that purifying selection played a pivotal role in driving the evolution of the StARF gene family.

To acquire a deeper understanding about the phylogenetic affiliations of ARF genes across diverse plant species, we constructed a collinearity map between potato and four representative plants: two dicotyledonous plants (*Arabidopsis* and tomato) and two monocotyledonous plants (rice and maize). The results showed varying degrees of collinearity between potato and other plants (Fig. 4, and Table S5). Among these, the number of homologous gene pairs between *SlARF* genes and *StARF* genes was the highest (144 pairs), indicating a close evolutionary relationship between the tomato and potato. A total of 73, 33, and 19 pairs of homologous genes were detected in potato and *Arabidopsis*, potato and rice, and potato and maize, respectively. The number of homologous gene pairs between the potato and monocotyledonous plants exhibited a marked decline. The number of homologous gene pairs between potato and dicotyledonous plants was markedly higher than between potato and monocotyledonous plants. These results indicate that a greater number of gene duplication events may have occurred in the *ARF* genes of dicotyledonous plants during evolutionary history.

Analysis of the expression patterns of *StARF* genes

Gene expression patterns typically correlate with gene functions. In this study, the FPKM values of 12 potato tissues were extracted from the transcriptome data obtained from the National Genomics Data Center (NGDC), and the expression patterns of *StARF* genes were subsequently analyzed. The results showed that the expression patterns of the StARF family members in diverse tissues and the expression trends of individual genes exhibited notable divergence (Fig. S6). A substantial proportion of *StARF* genes exhibited pronounced tissue-specific expression, with the majority displaying elevated levels of expression in the underground stolons. For example, the *StARF24-1* gene exhibited a distinctive expression profile, displaying high levels of expression exclusively in medium tubers (5–10 cm), and relatively low expression in other tissues. In contrast, the *StARF8-4* gene showed an inverse expression pattern. In addition, the expression of the StARF family members during flower development was analyzed. These results demonstrated that the

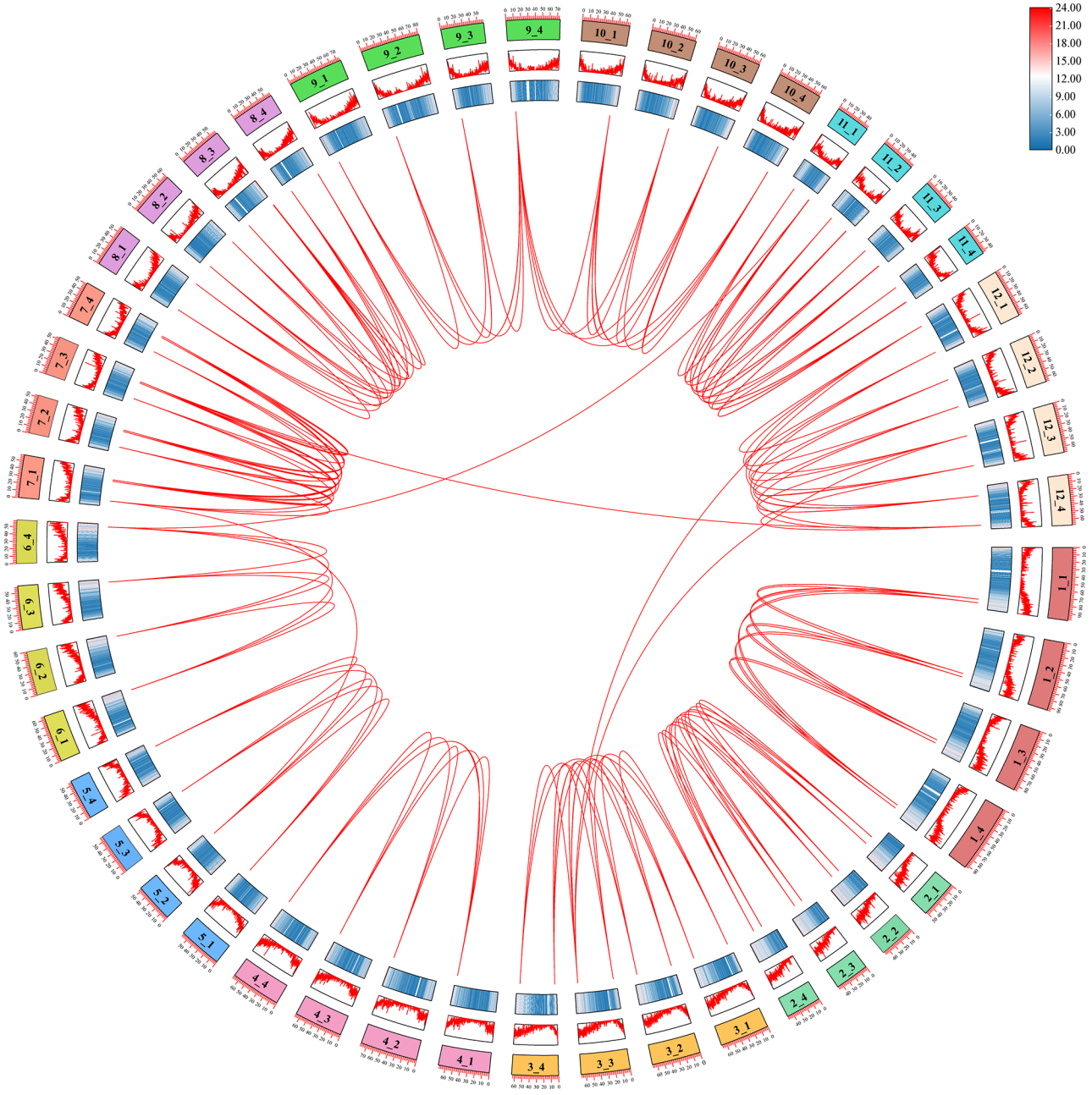

Fig. 3 Duplicated events in the StARF gene family. Red lines indicate the *StARF* duplicated gene pairs. Rectangular boxes of different colors represent different chromosomes, and the chromosome numbers are displayed on each chromosome. The blue heat map and broken red-line map represent gene densities on the chromosomes

expression patterns of *StARF* genes could be classified into multiple categories. For instance, the gene *StARF19* exhibited the highest expression levels in the flower buds. The genes *StARF13-1* and *StARF13-4* were highly expressed in stamens and petals. The gene *StARF17* was expressed in flower buds but not in the other two floral organs. These findings suggest that *StARF* genes are involved in several crucial processes during potato growth and development. The expression of certain genes was not detected in the 12 tissues. It is postulated that these genes may exhibit specific spatiotemporal expression patterns, necessitating further investigation.

To study *StARF* genes related to anthocyanin synthesis, we analyzed the expression profiles of *StARF* genes in the transcriptome data of Zicai 3 (rich in anthocyanins) and Longshu 7 (without anthocyanins) at different tuber development stages. The results demonstrated that with the maturation of potato, the anthocyanin content in the tuber flesh exhibited an initial increase followed by a subsequent decrease (Fig. S7). During this process,

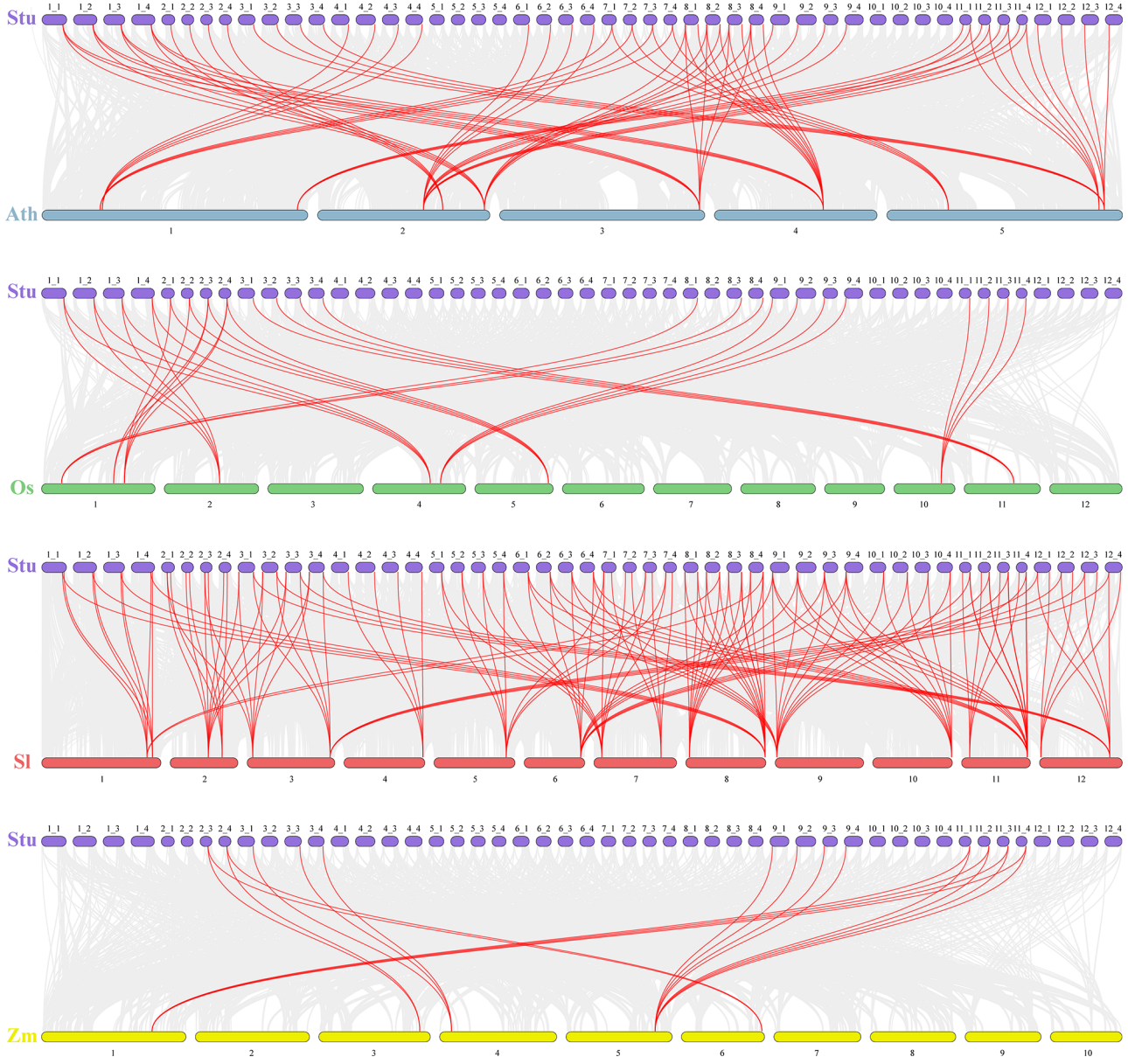

Fig. 4 Collinearity analysis of *ARF* genes between tetraploid potato and four other plants. The gray lines in the background represent the collinear blocks between the genomes of potato and other species, whereas the red lines represent the collinear *ARF* gene pairs. Different plant genomes use the following abbreviations: Stu, *Solanum tuberosum*; Ath, *Arabidopsis thaliana;* Os, *Oryza sativa;* Sl, *Solanum lycopersicum;* Zm, *Zea mays*

we identified a total of 41 differentially expressed StARF genes (Fig. 5B, and Table S6). Compared with 'Longshu 7', 27 *StARF* genes were upregulated, whereas 14 were downregulated in the purple flesh. Among these genes, *StARF5-4*, *StARF7-1*, *StARF14-1*, *StARF15-4*, *StARF19-3*, and *StARF24-2* were consistently upregulated in the four tuber developmental stages of purple flesh, whereas *StARF23-1* was consistently downregulated. In addition, correlation analysis was conducted between these differentially expressed *StARF* genes and the anthocyanin content in tuber flesh, as well as between these genes and the expression levels of key structural genes in anthocyanin biosynthesis. We identified 27 key structural genes involved in anthocyanin biosynthesis by analyzing the RNA-seq data. (Fig. S8). Correlation analysis indicated that 15 genes were positively correlated with the anthocyanin content ($r = 0.708–0.991$, $p < 0.05$). Eight genes were negatively correlated with the anthocyanin content, with r ranging from -0.719 to -0.830 (Fig. 5C). Among these genes, the gene *StARF2-4* showed a relatively significant positive correlation with the remaining 23 structural genes ($r = 0.708–0.913$, $p < 0.05$), except for one *PAL* gene (*C88_C09H2G035680.1*), one *4CL* gene (*C88_C03H2G064360.1*), and two *C4H* genes

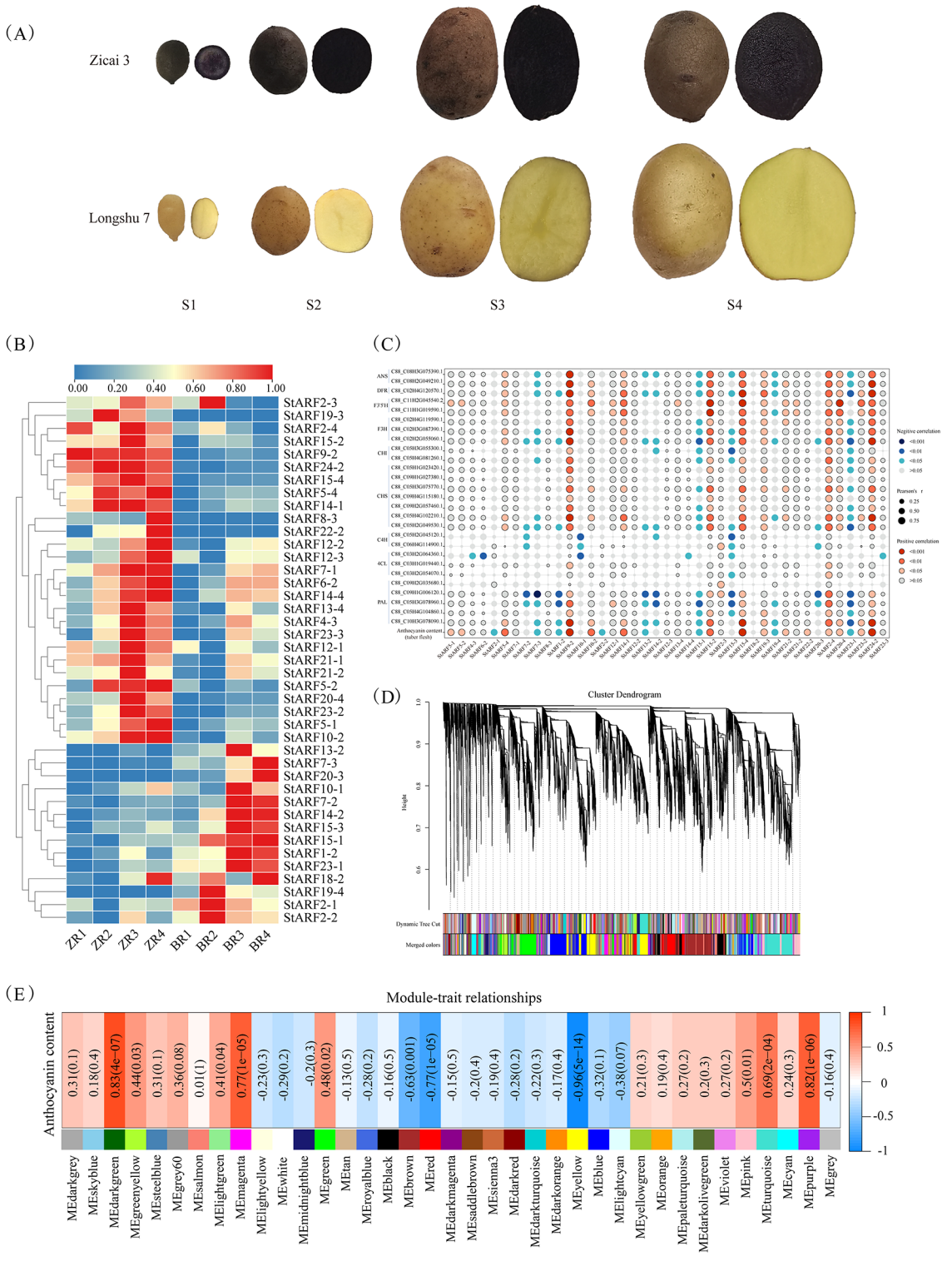

Fig. 5 (See legend on next page.)

(*C88_C06H4G114900.1* and *C88_C05H2G045120.1*). The gene *StARF23-1* exhibited a significant negative correlation with 17 structural genes (r = -0.722 to -0.906, $p < 0.05$). This indicates that the StARF family members are likely to be involved in the biosynthesis of anthocyanins in potato tuber flesh.

(See figure on previous page.)

Fig. 5 Expression and WGCNA analysis of *StARF* genes in different-colored flesh. **(A)** Cross-sections of tubers from two potato varieties with different flesh colors. S1 represents the tuber formation stage. S2 represents the tuber bulking stage. S3 represents the starch accumulation stage. S4 represents the tuber maturation stage. **(B)** Expression profiles of 41 *StARF* genes at different tuber development stages of 'Zicai 3' and 'Longshu 7'. ZR1: 'Zicai 3' of tuber formation stage. ZR2: 'Zicai 3' of tuber bulking stage. ZR3: 'Zicai 3' of starch accumulation stage. ZR4: 'Zicai 3' of tuber maturation stage. BR1: 'Longshu 7' of tuber formation stage. BR2: 'Longshu 7' of the tuber bulking stage. BR3: 'Longshu 7' of starch accumulation stage. BR4: 'Longshu 7' of the tuber maturation stage. **(C)** Correlation analysis between differentially expressed genes and anthocyanin content, as well as between differentially expressed genes and key structural genes involved in anthocyanin biosynthesis. **(D)** Gene clustering tree and module cutting. **(E)** Correlation heat map of each module and purple potato flesh. Red represents positive correlation, whereas blue represents negative correlation. The values inside and outside the brackets in the figure are P-values and correlation coefficients *r*, respectively

Co-expression network analysis of *StARF* genes

TFs are crucial regulatory proteins that play pivotal roles in diverse biological processes. To elucidate the interactions between *StARF* genes and their co-expressed TFs, we generated a weighted gene co-expression network from the RNA-seq data of 'Zicai 3' and 'Longshu 7'. Based on the correlation between gene expression levels, a clustering tree was constructed and divided into 37 modules according to the criteria of the hybrid dynamic cut, with different colors representing different modules (Fig. 5D). The MEturquoise module was the most extensive, comprising 3,260 genes. The MEyellowgreen module contained the fewest genes (37) (Fig. S9). In accordance with the established criteria of $|r| \geq 0.7$ and $p < 0.05$, five specific modules associated with anthocyanin content were identified. The MEmagenta, MEpurple, and MEdarkgreen modules exhibited a notable positive correlation with anthocyanin content, whereas the MEyellow and MEred modules showed a significant negative correlation (Fig. 5E).

We further investigated the TFs and StARF family members within the five modules of particular importance. The MEyellow module contained 87 TFs from 38 gene families, such as MYB, bHLH, ZIP, SBP, and WRKY (Table S7). Among the genes with differential expression in tubers of different flesh colors, five *StARF* genes were distributed in the MEyellow module (*StARF1-2*, *StARF2-4*, *StARF15-1*, *StARF23-1*, and *StARF24-2*). The MEpurple module contained 36 TFs from 17 gene families, including seven StARF family members (*StARF4-3*, *StARF13-4*, *StARF14-1*, *StARF15-2*, *StARF21-1*, *StARF21-2*, and *StARF21-3*). Further analysis demonstrated that the remaining six genes exhibited differential expression in tubers with differently colored flesh, except for *StARF21-3.* A total of 81 TFs were retrieved from the modules MEdarkgreen, MEred, and MEmagenta, including 34 transcription factor families (such as MYB, bHLH, GATA, GRAS, and ERF). Nevertheless, no StARF family members were detected within the three modules.

A co-expression network was constructed using the 11 differentially expressed *StARF* genes that were identified from the MEyellow and MEpurple modules as key nodes. In the MEyellow module (Fig. 6A), the genes *StARF1-2*, *StARF2-4*, *StARF15-1*, *StARF23-1*, and *StARF24-2* were co-expressed with 69 TFs from 35 gene families, including MYB, bHLH, WRKY, ERF, and SBP. Within the MEpurple module (Fig. 6B), the genes *StARF4-3*, *StARF13-4*, *StARF14-1*, *StARF15-2*, *StARF21-1*, and *StARF21-2* were co-expressed with 18 TFs from 12 gene families, such as bZIP, SBP, and GRAS.

Protein-protein interaction network analysis

To further elucidate the potential regulatory relationship between *StARF* genes and anthocyanin biosynthesis structural genes, we used PlantCARE to identify *cis*-acting elements within the promoter regions of 27 key structural genes. The results showed that there were no typical AuxRE (TGTCTC) elements in the promoter regions of the 27 structural genes (Fig. 7A). However, the search revealed that there were nine AuxRR-core (GGTCCA T) elements in the promoter regions of seven structural genes, and one TGA-element was present in the promoter region of the *PAL* gene.

Based on the known protein-protein interaction relationships of *Arabidopsis thaliana* in the STRING database, we further predicted and constructed a protein-protein interaction network between 11 differentially expressed *StARF* genes and 27 structural genes using the homology mapping method. The results showed that the StARF23-1 protein (homologous to AtARF2) was predicted to interact with anthocyanidin synthase (ANS), a pivotal structural gene in anthocyanin biosynthesis (Fig. 7B). StARF1-2, StARF15-1, and StARF15-2 have been identified as homologs of AtARF9, whereas StARF14-1 and StARF24-2 have been classified as homologs of AtARF6. In addition, StARF2-4 is a homolog of AtARF1. These five proteins (StARF1-2, StARF14-1, StARF15-1, StARF15-2, and StARF24-2) exhibited a strong interaction relationship among themselves. However, no protein-protein interactions between them and structural genes were identified. In conclusion, these findings provide a preliminary foundation for further investigation of the potential regulatory functions of StARF TFs in anthocyanin biosynthesis.

qRT-PCR validation

To further validate the expression patterns of *StARF* genes in differently colored flesh, nine differentially expressed *StARF* genes were selected for qPCR analysis. The results demonstrated that the overall patterns were

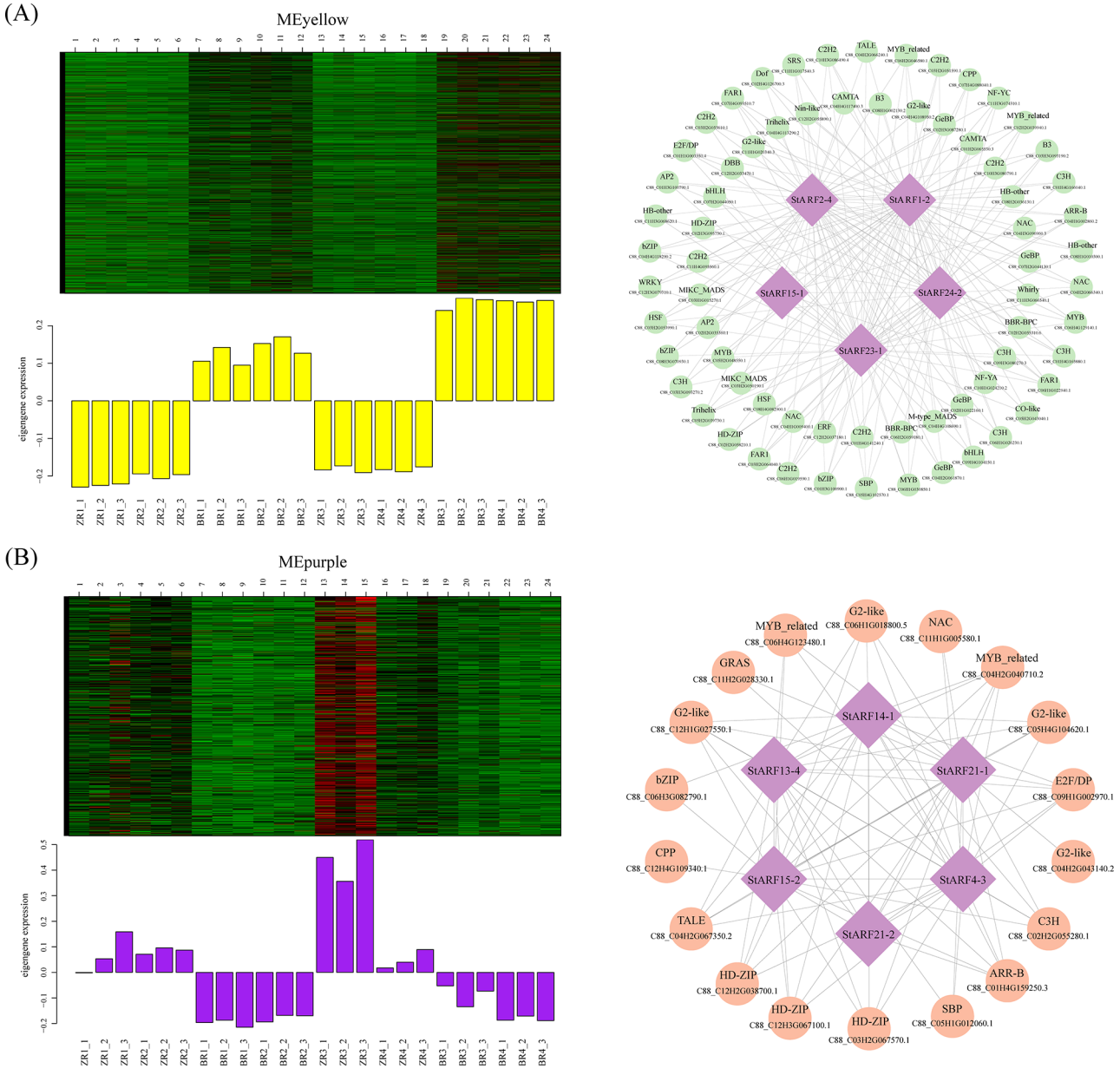

Fig. 6 WGCNA module and co-expression network of differentially expressed *StARF* genes and transcription factors in different-colored flesh. **(A)** Co-expression network of *StARF* genes and transcription factors in the MEyellow module. **(B)** Co-expression network of *StARF* genes and transcription factors in the MEpurple module

concordant, although there were discrepancies between the qPCR outcomes and the RNA-seq data (Fig. 8). Through linear regression analysis, a regression equation $y = 0.3285x - 0.1131$ was obtained, and $R^2 = 0.7876$, indicating a relatively good correlation between the two, and the RNA-seq data were reliable.

Discussion

Auxin response factors (ARFs), a group of TFs central to the auxin signaling pathway, play a critical role in regulating plant growth and development by binding to auxin response elements (AuxREs) and modulating downstream gene expression [14, 15]. To date, extensive studies have identified ARF gene families across various plant species, including 23 members in *Arabidopsis thaliana* [24], 25 in rice [25], 21 in tomato [26], 19 in grape [41], and 81 in alfalfa [16]. However, despite these advancements, a comprehensive analysis of the StARF gene family in the tetraploid potato genome remains unclear. This gap is largely attributed to the genetic complexity of cultivated potato, which is autotetraploid, highly heterozygous, and exhibits intricate genetic segregation, posing

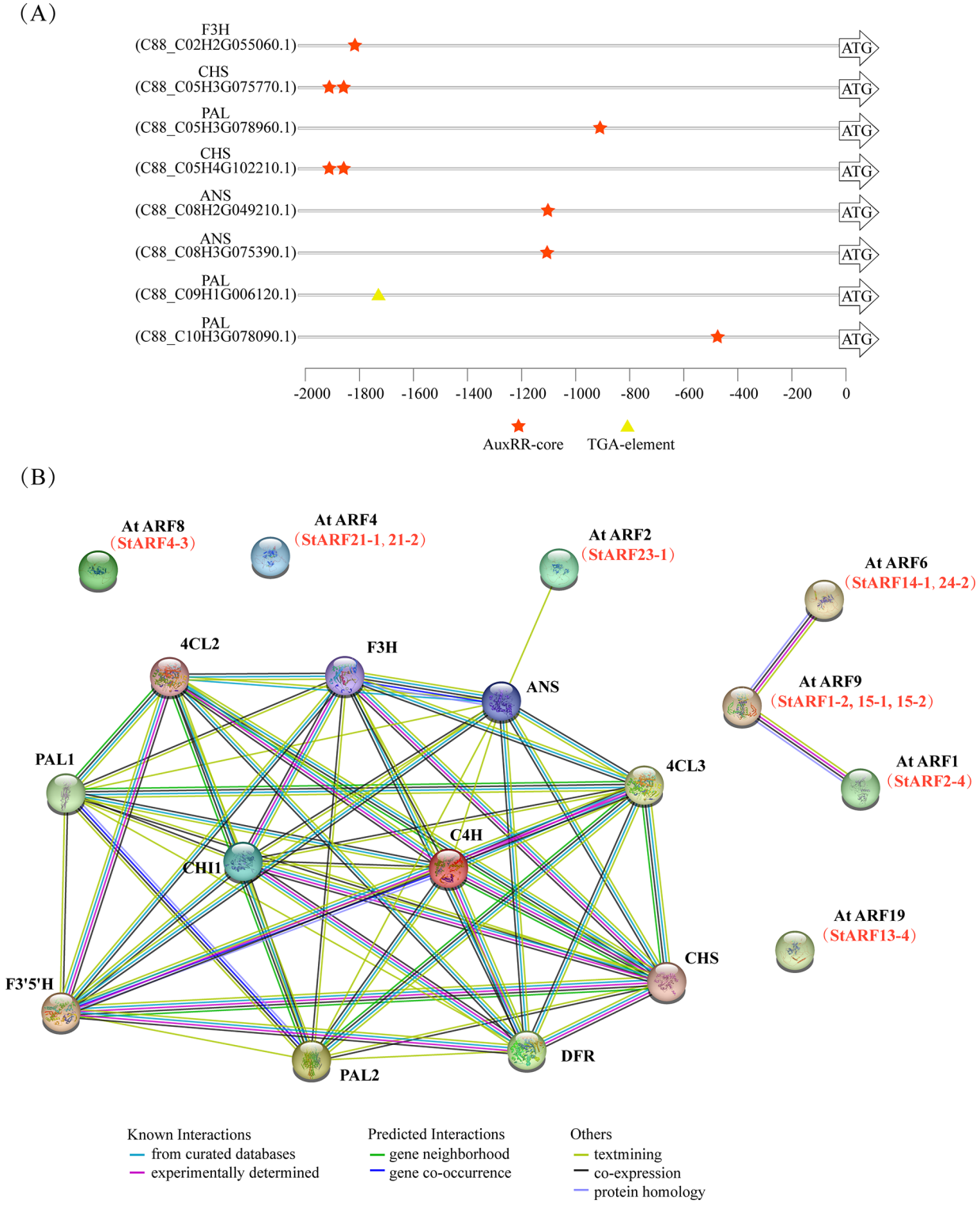

Fig. 7 Interaction analysis between differentially expressed candidate *StARF* genes and anthocyanin biosynthesis-related structural genes. **(A)** Schematic diagram of the promoter regions of anthocyanin biosynthesis structural genes with auxin-response elements AuxRR-core and TGA-element. The red pentagrams represent the AuxRR-core element and the yellow pentagrams represent the TGA-element. **(B)** Protein-protein interaction network between candidate differentially expressed *StARF* genes and structural genes related to anthocyanin biosynthesis. Black and red characters represent *Arabidopsis* and potato, respectively. The different-colored lines represent different types of interactions

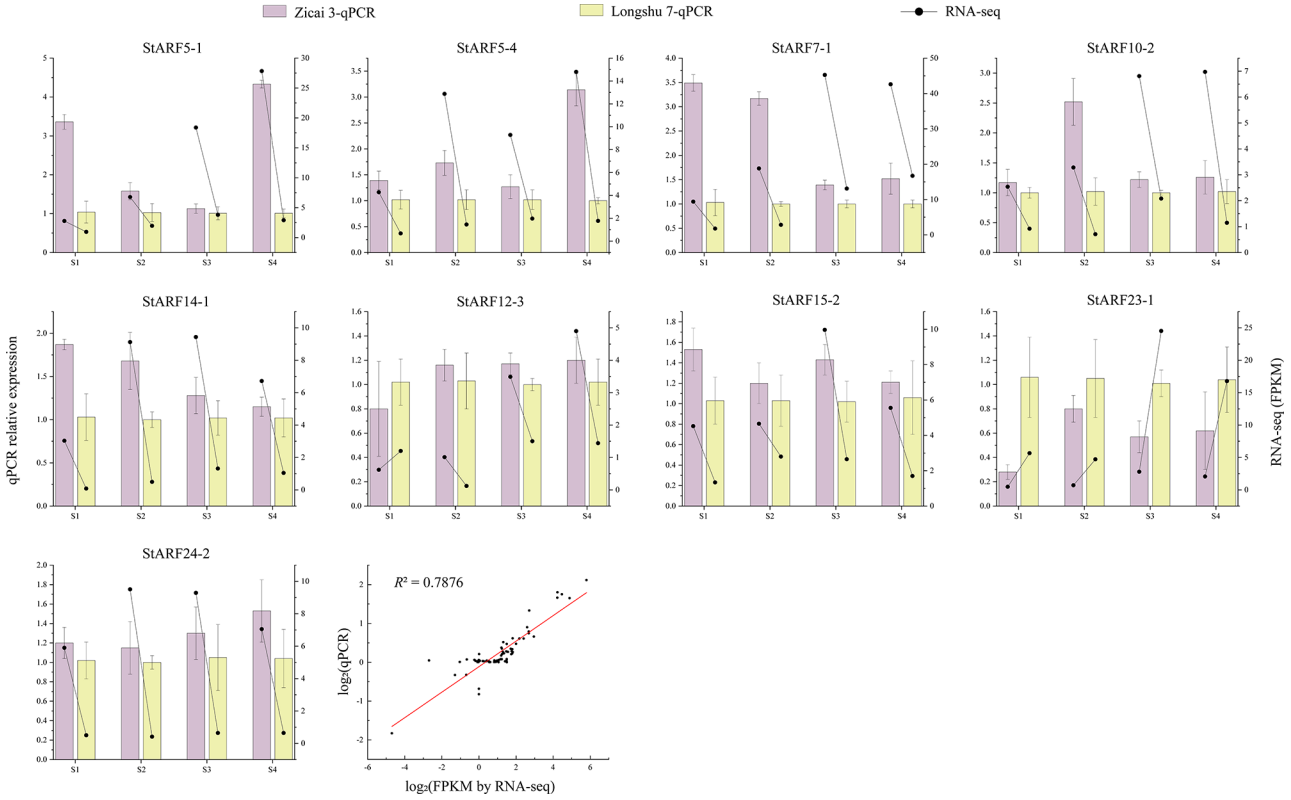

Fig. 8 qRT-PCR analysis of the differentially expressed *StARF* genes in the four tuber developmental stages of 'Zicai 3' and 'Longshu 7'. S1 represents the tuber formation stage. S2 represents the tuber bulking stage. S3 represents the starch accumulation stage. S4 represents the tuber maturation stage

significant challenges for genome assembly and functional studies [38, 40]. Recent breakthroughs in genome sequencing and assembly technologies have begun to address these challenges, enabling the reconstruction of complex polyploid genomes [42]. Notably, Bao et al. [40] published data on the complete genome (C88.v1) of autotetraploid potato comprising 48 high-quality chromosomes in 2022. This advancement provides dependable genomic resources for a comprehensive investigation of the StARF gene family in tetraploid potato as well as crucial insights for biological research and molecular breeding of cultivated potato.

In this study, using bioinformatics methods, we identified the StARF gene family in the tetraploid potato genome for the first time. This gene family comprises 92 *StARF* genes, including alleles, which are distributed across 48 chromosomes. Phylogenetic tree analysis based on *Arabidopsis thaliana* and potato revealed that the 92 StARF proteins could be classified into four subfamilies. Notably, the V subfamily contains no StARF proteins but included eight AtARF proteins. Previous studies have shown that *AtARF* genes in this subfamily originate from a single *AtARF* gene [28, 29]. Similar observations have been made for other plant species, including maize [27], grape [41], and apple [28]. It is plausible that these plants and tetraploid potato have lost only a few *ARF* genes during evolution. Moreover, the count of *StARF* genes identified in this study was considerably higher than that observed in diploid plants, including *Arabidopsis thaliana* (23 *AtARFs*) [24], rice (25 *OsARFs*) [25], tomato (21 *SlARFs*) [26], maize (31 *ZmARFs*) [27], soybean (51 *GmARFs*) [14], and diploid potato (20 *StARFs*) [43]. Compared to cultivated alfalfa (81 *MsARFs*) [16], which is also an autotetraploid, the discrepancy in the number of *ARF* genes is relatively minor. While the genome size of cultivated alfalfa is 2.738 Gb [16], tetraploid potato can reach 3 Gb [40], significantly larger than the genomes of *Arabidopsis*, rice, tomato, and diploid potato. However, previous studies have shown that there is no proportionality between the number of *ARF* genes and genome size [44]. Therefore, it is postulated that the amplification of *StARF* gene numbers may be related to the polyploidization of the cultivated potato genome, resulting in the emergence of additional homologous genes among the different chromosome sets of family members. This hypothesis is analogous to a genome-wide analysis of the MADS-box gene family in wheat [45]. In addition, gene replication events serve as the main driving factor for the formation and expansion of gene families [16]. Such events result in the production of additional copies of several genes, which give rise to functional differences in the encoded proteins [46]. Analysis of gene duplication events in 92

StARF family members revealed that all members had experienced gene duplication events, with these events being WGD or segmental duplication events. It suggests that WGD and segmental duplication events are the primary forces driving the expansion of the StARF gene family in tetraploid potato.

The analysis of the expression patterns of the *StARF* genes in 12 tissues revealed significant differences in the expression levels of *StARF* genes among the different tissues. Over two-thirds of the *StARF* genes display tissue-specific expression patterns. The expression of a gene in a specific tissue indicates the relationship between the gene and the function of the tissue [47]. In *Arabidopsis*, *AtARF6* and *AtARF8* have been demonstrated to mediate floral organs [48]. In this study, *StARF24-2* and *StARF24-3* were in the same cluster as *AtARF6*, and these two *StARF* genes exhibited relatively high expression levels in flower buds and petals. Thus, it can be postulated that *StARF24-2* and *StARF24-3* perform analogous functions in modulating floral organogenesis. *StARF23-3* was highly expressed in fruits and clusters and was closely correlated with *AtARF2*. Previous research has demonstrated that *AtARF2* is involved in the regulation of fruit ripening [49]. Thus, *StARF23-3* may play an important role in fruit development. Furthermore, the analysis revealed that some alleles exhibited differential expression. For instance, no expression data for *StARF7-4* was identified in any tissue. In contrast, the duplicated gene pairs *StARF7-1* and *StARF7-2* exhibited high expression levels in the apical buds, whereas *StARF7-3* demonstrated a low expression level in the same tissue. This phenomenon may be attributed to the sequence divergence experienced by the alleles, which gives rise to functional divergence, resulting in the evolution of a certain gene in a novel functional direction influencing the expression pattern of the genes [17, 27].

Anthocyanins, water-soluble natural pigments in plant cell vacuoles, belong to flavonoid compounds. They endow plant organs with bright colors, confer resilience to low temperatures, ultraviolet radiation, and diseases in plants, and bring health benefits to humans [1–3]. Colored potato, rich in anthocyanins, is valuable for food and industrial processing. Anthocyanin biosynthesis is regulated by multiple TFs. The MYB, bHLH, and WD40 families play key roles, either acting independently or forming an MBW complex to regulate structural gene transcription [5]. Other TFs like bZIP [6], ERF [7], SPL [8], and WRKY [9] are also involved. Although recent studies in *Arabidopsis* and apple have shown that ARF TFs can regulate anthocyanin synthesis [10–13], research on whether StARF family members are involved in colored potato anthocyanin synthesis is scarce. Our study aimed to explore the potential role of *StARF* genes in colored potato anthocyanin biosynthesis. We analyzed the expression of *StARF* genes in potato tubers with different-colored flesh and examined their correlation with anthocyanin content and key structural genes. The results showed 41 *StARF* genes had differential expression, and 20 of them correlated with anthocyanin content and key structural genes. For example, *StARF2-4* had a significant positive correlation with anthocyanin content and 23 structural genes, while *StARF23-1* had a significant negative correlation with anthocyanin content and 17 structural genes. To find key *StARF* genes, we used WGCNA and identified 12 *StARF* genes in MEyellow and MEpurple modules. Except *StARF21-3*, 11 had differential expression and correlated with structural genes. We identified 87 TFs regulating the expression of *StARF* genes and constructed a co-expression network. Many co-expressed TFs belonged to families involved in anthocyanin biosynthesis, such as MYB, bHLH, bZIP, and SBP [5, 6, 8]. For example, AtMYB111 has been demonstrated to bind to the promoters of structural genes, including *CHS*, CHI, and *F3H*, activating their expression and promoting anthocyanin synthesis [50, 51]. GAI is one of the earliest members of the GRAS transcription factor family and has been the subject of numerous studies [52]. It has been demonstrated that AtGAI interacts with AtMYB111 to regulate the expression of structural genes [53]. In this study, StGAI (*C88_C11H2G028330.1*, which is homologous to AtGAI) demonstrated a high connectivity with StARF21-1 in the co-expression network, indicating that GAI may have a regulatory relationship with StARF21-1. AtPHL3 is a crucial transcription factor in *Arabidopsis*, belonging to the MYB-CC subfamily of the MYB transcription factor family [54]. A previous study demonstrated that AtPHL3 is closely associated with anthocyanin synthesis and exerts a positive regulatory effect on JA-induced anthocyanin accumulation [55]. In this study, StPHL3 (*C88_C04H4G108950.2*, homologous to AtPHL3) was connected to StARF1-2 and StARF2-4 in the co-expression network, indicating a potential interaction. Moreover, the promoter regions of *StARF* genes had many MYB and bHLH binding sites, indicating possible interaction with other TFs in colored potato to affect anthocyanin synthesis.

We also predicted *cis*-acting elements in the promoter regions of 27 structural genes. Although the typical AuxREs were absent, they contained AuxRR-core and TGA-element, which ARF TFs can bind to [56, 57]. This implies that StARF TFs may directly regulate potato anthocyanin biosynthesis genes and affect anthocyanin accumulation. The protein-protein interaction network predicted that StARF23-1 interacted with ANS, and there was a significant correlation between them. Thus, StARF23-1 may be a key candidate gene for anthocyanin synthesis. Overall, the different expression patterns of *StARF* genes in differently colored flesh suggest

their potential involvement in colored potato anthocyanin biosynthesis, which broadens our understanding of the regulatory mechanisms of anthocyanin synthesis and provides new insights for improving potato quality through genetic engineering.

Conclusions

In this study, the ARF gene family was first identified at the genome-wide level in tetraploid potato, and 92 StARF family members were identified. A comprehensive analysis was conducted on the physicochemical properties, phylogeny, gene structure, *cis*-acting elements, gene duplication events, and tissue expression patterns of StARF family members. Moreover, RNA-seq data were used to compare the *StARF* genes that exhibited differential expression in potato of different colors. Correlation analysis demonstrated that the *StARF* genes with differential expression were significantly correlated with the key structural genes of anthocyanins. Further through WGCNA analysis, 11 differentially expressed *StARF* genes were identified as candidate genes involved in anthocyanin biosynthesis in colored potato, and a corresponding co-expression network was constructed. In addition, protein-protein interaction prediction analysis indicated that StARF23-1 may serve as a key regulatory factor for anthocyanin synthesis, a hypothesis that warrants further investigation. In conclusion, these results provide potential candidate genes for further investigation into the role of StARF gene family in anthocyanin biosynthesis in colored potato, which establish a foundation for further elucidation of the intricate regulatory mechanisms underlying anthocyanin synthesis in colored potato.

Materials and methods

Identification and physicochemical characterization of StARF proteins in tetraploid potato

The 23 ARF amino acid sequences of *Arabidopsis thaliana* were obtained from the TAIR database (https://www.arabidopsis.org), and related data for the tetraploid potato genome were sourced from SpudDB (https://spuddb.uga.edu/c88_potato_download.shtml). In this study, two methods were used to identify members of the StARF gene family in tetraploid potato. The first method used the AtARF protein sequence as the seed sequence and performed a BLASTP (E-value ≤ 1e-10) search with the tetraploid potato genome sequence to obtain homologous sequences. Secondly, the Hidden Markov Model file of the ARF protein domain (PF06507) was obtained from the Pfam database (http://pfam.xfam.org), and a sequence containing the ARF conserved domain was searched in the tetraploid potato protein sequences using HMMER 3.3.2 (http://hmmer.org/download.html) [58]. The StARF protein sequences identified using the two aforementioned methods were integrated, and any repeat was subsequently removed. The conserved domains of the candidate members were additionally validated via the websites NCBI-CDD (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) [59] and SMART (https://smart.embl.de) [60]. Sequences lacking ARF protein characteristic domains were removed, and the remaining sequences were deemed members of the tetraploid potato ARF gene family.

The physicochemical properties and subcellular locations of StARF protein sequences were predicted using ExPASy (https://web.expasy.org/protparam/) [61] and WoLF PSORT (https://wolfpsort.hgc.jp) [62], respectively. The position of the *StARF* gene on the chromosome was visualized using TBtools v1.108 tool [63], and the gene was renamed following the order of the *StARF* gene within the chromosome.

Multiple sequence alignment and phylogenetic analysis of StARF gene family in tetraploid potato

The ClustalW program [64] was used to execute multiple sequence alignments between the protein sequences of the 23 reported AtARF family members and the identified StARF family members. A phylogenetic tree was constructed using MEGA 11 software [65] using the neighbor-joining (NJ) method. Bootstrap was set to 1,000, and the remaining parameters were set to their default values. The identified StARF family members were classified into subfamilies according to established criteria for the ARF gene family subfamily of the model plant *Arabidopsis* [24]. ARF protein sequences of rice (25 OsARF proteins) [25], tomato (21 SlARF proteins) [26], and maize (31 ZmARF proteins) [27] were acquired from the PlantTDB 5.0 database (https://plantregmap.gao-lab.org) [66]. A phylogenetic tree of the ARF gene family between potato and different plants was constructed to further elucidate the genetic relationships of the ARF genes among different plants.

Analysis of conserved motifs and gene structure of StARF proteins

The exon-intron position information of the *StARF* gene was drawn out depending on the gff3 file, while the gene structure of the StARF family members was analyzed visually using TBtools. The StARF protein sequences were submitted to the MEME website (http://meme-suite.org/meme/tools/meme) [67] for the analysis of conserved motifs. The motif search value was set to 10 and the remaining parameters were set to their default values. Furthermore, StARF protein sequences were used to conduct a multiple sequence alignment using DNAMAN v9.0 software, and a protein domain diagram was generated using TBtools based on the analysis results of the

protein domain of the StARF gene family on the NCBI-CDD website.

Cis-acting elements located in the promoter region of *StARF* genes

A 2 kb sequence upstream of *StARF* genes was extracted from the tetraploid potato genome database. The online tool PlantCARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html) [68] was used to predict potential *cis*-acting elements within the sequence, and TBtools were used to visualize the resulting data.

Analysis of collinearity and evolutionary selection of StARF gene family

The duplication events of the *StARF* genes were analyzed using McScanX software, a built-in function of TBtools software. The non-synonymous (Ka) and synonymous (Ks) substitution rates and Ka/Ks values of the duplicated gene pairs were then calculated. Furthermore, the collinearity of *ARF* genes among plants was analyzed based on the whole-genome sequences and annotation files, among which the ones of *Arabidopsis* and rice were downloaded from the Ensembl Plants database (https://plants.ensembl.org/index.html), while those of tomato and maize were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov).

Expression patterns analysis of *StARF* genes in tetraploid potato

Raw data from the RNA-seq (project number: PRJCA007997) were obtained from the National Genomics Data Center (NGDC, https://bigd.big.ac.cn), encom passing 12 tissues of the potato cultivar C88, including apical buds, flower buds, stamens, petals, young leaves, stems, underground stolons, small tubers (1–5 cm), medium tubers (5–10 cm), large tubers (> 10 cm), and in vitro shoots and fruits. The FPKM values of the *StARF* genes in various tissues were extracted, and a heat map of tissue expression was generated using TBtools software. Furthermore, we used RNA-seq data (project number: PRJNA1033052) obtained from the four tuber development stages of 'Zicai 3' and 'Longshu 7' by our lab in the previous stage, and then aligned them with the tetraploid potato C88.v1 reference genome (https://spuddb.uga.edu/c88_potato_download.shtml) [40]. The FPKM value was used as an indicator to evaluate gene expression levels. The data were used to identify differentially expressed *StARF* genes in potato flesh of different colors. The criteria for identifying differentially expressed genes were $|\log_2(\text{Fold Change})| > 1$ and FDR < 0.05.

To further identify candidate *StARF* genes linked to anthocyanin synthesis, Pearson's correlation analysis was performed to assess the correlations between the differentially expressed *StARF* genes and anthocyanin content, and between these genes and key structural genes involved in anthocyanin biosynthesis. The R package was used for visualization, and a P-value of less than 0.05 was deemed statistically significant.

WGCNA identified candidate modules related to anthocyanin content in the tuber flesh

A co-expression network was constructed using the WGCNA package [69] in R software based on transcriptome data from different tuber development stages of 'Zicai 3' and 'Longshu 7'. The soft threshold power was set to 9 ($\beta = 9$) to ensure that the network was suitable for scale-free topology. Dissimilarity between genes was used to hierarchically cluster genes in the network, establishing a hierarchical clustering tree. Subsequently, the tree was divided into discrete modules (with at least 30 genes per module) by using the dynamic cutting method. Modules with correlation coefficients exceeding 0.75 were then merged. To identify specific anthocyanin-related modules, the correlation coefficient (r) and P-values of the module eigengene (ME) and phenotype (anthocyanin content in the potato tuber flesh) were calculated. Cytoscape v.3.7.1 software [70] was used to visualize the gene interactions within the module and to construct a gene co-expression network diagram.

The STRING website (https://cn.string-db.org) [71] was used to examine the protein interaction network of the differentially expressed *StARF* genes.

Plant materials, RNA extraction and qRT-PCR analysis

The tetraploid variety 'Zicai 3,' which is purple-skinned and has deep-purple flesh, was independently developed by the Agriculture College of Inner Mongolia Agricultural University. The yellow-skinned and yellow-fleshed variety designated 'Longshu 7' was developed by the Potato Research Institute of the Gansu Academy of Agricultural Sciences. The two varieties were used as experimental materials (Fig. 5A). The experimental materials were planted at the experimental base of the Agricultural College of Inner Mongolia Agricultural University (40°46'N, 110°45'E). During the growth period, fertilization was applied in a manner appropriate to the growth of the plant, weeds were controlled, and the soil was cultivated to ensure normal growth of the plant.

From July 2022 onward, tuber flesh samples of 'Zicai 3' and 'Longshu 7' were collected at four stages of tuber development: tuber formation, tuber bulking, starch accumulation, and tuber maturation. Following rapid freezing in liquid nitrogen, the samples were stored at -80 °C for subsequent analysis. Three biological replicates were used for each tuber development stage. Total RNA was extracted using the Trizol method and reverse-transcribed using a FastKing RT Kit (with gDNAse) (Tiangen KR116, Beijing, China). The differentially expressed

StARF genes were selected for qRT-PCR verification. The qPCR analysis was conducted using MonAmp SYBR Green qPCR Mix (Monad, Suzhou, China) on a QuantStudio 3 & 5 PCR instrument (Thermo Fisher Scientific, USA). *EF-1a* was used as a housekeeping gene, and the relative expression of the gene was calculated using the $2^{-\Delta\Delta Ct}$ method [72]. The Origin 2022 software was used for plotting. The primers used for qRT-PCR are provided in Table S8.

Determination of the anthocyanin content

The anthocyanin content in the potato flesh at different tuber development stages was determined using a total anthocyanin content kit (Geruisi, Suzhou, China). Previous studies have shown that the tuber flesh of 'Longshu 7' is yellow, mainly because of the color development effect caused by carotenoid substances, and there is no anthocyanin accumulation [73]. In this study, the anthocyanin content in the potato flesh of 'Longshu 7' was not determined, and its value was set to 0.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06366-4.

Supplementary Material 1: Fig. S1. Phylogenetic tree of the ARF gene family in Arabidopsis, rice, tomato, maize, and tetraploid potato. Different subfamilies are represented by arcs of different colors. Blue triangle, green circle, red square, yellow triangle and purple pentagram represent Arabidopsis, rice, tomato, maize, and potato, respectively.

Supplementary Material 2: Fig. S2. Gene structure analysis of the StARF family members. Yellow rectangular boxes, green rectangular boxes, and gray line distributions represent exons, untranslated regions (UTRs), and introns.

Supplementary Material 3: Fig. S3. Phylogenetic tree, conserved motifs, and protein domains analyses of StARF family members. (A) Phylogenetic tree of 92 StARF family members. (B) The conserved motifs of StGATA proteins are shown in different colors, and their related information is presented in the top-right legend. (C) StARF conserved protein domains.

Supplementary Material 4: Fig. S4. Multiple sequence alignment of StARF domains.

Supplementary Material 5: Fig. S5. Analysis of cis-acting elements 2 kb upstream of the StARF genes. The boxes of different colors represent cis-acting elements, as shown in the legend at the top-right.

Supplementary Material 6: Fig. S6. StARF genes expression patterns in 12 potato tissues. StARF gene expression values were normalized by log2-transformed (FPKM+1).

Supplementary Material 7: Fig. S7. Anthocyanin content in the tuber flesh of 'Zicai 3' during four tuber development stages.

Supplementary Material 8: Fig. S8. Identification and expression analysis of key structural genes in the anthocyanin biosynthesis pathway of tetraploid potato.

Supplementary Material 9: Fig. S9. Number of genes within the modules.

Supplementary Material 10: Table S1. Physicochemical characteristics and subcellular localization of StARF proteins in tetraploid potato.

Supplementary Material 11: Table S2. Motifs sequence of the StARF gene family in tetraploid potato.

Supplementary Material 12: Table S3. Gene duplication of ***StARF*** genes.

Supplementary Material 13: Table S4. Ka/Ks ratios of ***StARF*** genes in tetraploid potato.

Supplementary Material 14: Table S5. Collinearity analysis of ***ARF*** genes between tetraploid potato and ***Arabidopsis***, rice, tomato, and maize.

Supplementary Material 15: Table S6. The FPKM values of 41 differentially expressed ***StARF*** genes.

Supplementary Material 16: Table S7. The information of TFs in five modules.

Supplementary Material 17: Table S8. Primers for qRT-PCR study.

Supplementary Material 18

Acknowledgements
Not applicable.

Author contributions
Author contributions X.Z. and X.Y. conceived and designed the experiments. R.F., X.H., H.W. and W.X. performed the experiments. X.Z. analyzed the data and wrote the manuscript. X.Y. and Z.Y. revised and edited the manuscript. All authors reviewed the manuscript.

Funding
This research was funded by the Outstanding Youth Fund Cultivation Project of the Inner Mongolia Agricultural University (BR220403), the Inner Mongolia Natural Science Foundation (2024MS03018), the Inner Mongolia Agricultural University Research Innovation Project for Graduate Students (DC2400000900), and the Inner Mongolia Major Science and Technology Project (ZDZX2018019).

Data availability
Data availabilityAll data generated or analyzed during this study are included in this published article and its supplementary files. The datasets analysed during the current study are available in the NCBI Sequence Read Archive database (project number: PRJNA1033052). Further inquiries can be directed to the corresponding author.

Declarations

Ethics approval and consent to participate
Experimental research on plants in this study complied with institutional, national, or international guidelines and legislation.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 2 January 2025 / Accepted: 7 March 2025
Published online: 17 March 2025

References

1. Landi M, Tattini M, Gould K. Multiple functional roles of anthocyanins in plant-environment interactions. Environ Exp Bot. 2015;119:4–17.
2. Ghosh D, Konishi T. Anthocyanins and anthocyanin-rich extracts: role in diabetes and eye function. Asia Pac J Clin Nutr. 2007;16(2):200–8.
3. Wu Y, Han T, Lyu L, Li W, Wu W. Research progress in Understanding the biosynthesis and regulation of plant anthocyanins. Sci Hortic. 2023;321:112374.
4. Mathura SR, Sutton F, Bowrin V. Genome-wide identification, characterization, and expression analysis of the sweet potato (*Ipomoea batatas* [L.] Lam.) *ARF*, Aux/*IAA*, *GH3*, and *SAUR* gene families. BMC Plant Biol. 2023;23(1):622.
5. Chen L, Cui Y, Yao Y, An L, Bai Y, Li X, Yao X, Wu K. Genome-wide identification of WD40 transcription factors and their regulation of the MYB-bHLH-WD40

(MBW) complex related to anthocyanin synthesis in Qingke (*Hordeum vulgare* L. Var. Nudum Hook. f). BMC Genomics. 2023;24(1):166.
6. Chen M, Cao X, Huang Y, Zou W, Liang X, Yang Y, Wang Y, Wei J, Li H. The bZIP transcription factor MpbZIP9 regulates anthocyanin biosynthesis in *Malus* 'pinkspire' fruit. Plant Sci. 2024;342:112038.
7. Zhang S, Liu S, Ren Y, Zhang J, Han N, Wang C, Wang D, Li H. The ERF transcription factor ZbERF3 promotes ethylene-induced anthocyanin biosynthesis in *Zanthoxylum bungeanum*. Plant Sci. 2024;349:112264.
8. Wang Y, Liu W, Wang X, Yang R, Wu Z, Wang H, Wang L, Hu Z, Guo S, Zhang H, et al. MiR156 regulates anthocyanin biosynthesis through *SPL* targets and other MicroRNAs in Poplar. Hortic Res. 2020;7:118.
9. Wang N, Song G, Zhang F, Shu X, Cheng G, Zhuang W, Wang T, Li Y, Wang Z. Characterization of the WRKY gene family related to anthocyanin biosynthesis and the regulation mechanism under drought stress and Methyl jasmonate treatment in *Lycoris radiata*. Int J Mol Sci. 2023;24(4):2423.
10. Jiang W, Xia Y, Su X, Pang Y. ARF2 positively regulates flavonols and proanthocyanidins biosynthesis in *Arabidopsis Thaliana*. Planta. 2022;256(2):44.
11. Wang YC, Wang N, Xu HF, Jiang SH, Fang HC, Su MY, Zhang ZY, Zhang TL, Chen XS. Auxin regulates anthocyanin biosynthesis through the Aux/IAA–ARF signaling pathway in Apple. Hortic Res. 2018;5:59.
12. Wang CK, Han PL, Zhao YW, Yu JQ, You CX, Hu DG, Hao YJ. Genome-wide analysis of auxin response factor (ARF) genes and functional identification of *MdARF2* reveals the involvement in the regulation of anthocyanin accumulation in Apple. N Z J Crop Hortic Sci. 2021;49(2–3):78–91.
13. Li HL, Liu ZY, Wang XN, Han Y, You CX, An J-P. E3 ubiquitin ligases SINA4 and SINA11 regulate anthocyanin biosynthesis by targeting the IAA29–ARF5-1–ERF3 module in Apple. Plant Cell Environ. 2023;46(12):3902–18.
14. Ali S, Wang W, Zhang Z, Xie L, Boer DR, Khan N. Genome-wide identification, expression and interaction analysis of ARF and AUX/IAA gene family in soybean. Front Biosci. 2022;27(8):251.
15. Boer DR, Freire-Rios A, van den Berg WAM, Saaki T, Manfield IW, Kepinski S, López-Vidrieo I, Franco-Zorrilla J, de Vries SC, Solano R, et al. Structural basis for DNA binding specificity by the auxin-dependent ARF transcription factors. Cell. 2014;156(3):577–89.
16. Chen F, Zhang J, Ha X, Ma H. Genome-wide identification and expression analysis of the auxin-response factor (ARF) gene family in *Medicago sativa* under abiotic stress. BMC Genomics. 2023;24(1):498.
17. Wen J, Guo P, Ke Y, Liu M, Li P, Wu Y, Ran F, Wang M, Li J, Du H. The auxin-response factor gene family in allopolyploid *Brassica napus*. PLoS ONE. 2019;14(4):e0214885.
18. Tiwari SB, Hagen G, Guilfoyle T. The roles of auxin response factor domains in auxin-responsive transcription. Plant Cell. 2003;15(2):533–43.
19. Guilfoyle TJ, Hagen G. Auxin response factors. Curr Opin Plant Biol. 2007;10(5):453–60.
20. Ulmasov T, Hagen G, Guilfoyle TJ. ARF1, a transcription factor that binds to auxin response elements. Science. 1997;276(5320):1865–8.
21. Hagen G, Guilfoyle T. Auxin-responsive gene expression: genes, promoters and regulatory factors. Plant Mol Biol. 2002;49:373–85.
22. Liscum E, Reed JW. Genetics of Aux/IAA and ARF action in plant growth and development. Plant Mol Biol. 2002;49:387–400.
23. 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):1738–52.
24. 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):444–63.
25. 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.
26. 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):2059–73.
27. 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.
28. Luo XC, Sun MH, Xu RR, Shu HR, Wang JW, Zhang SZ. Genome-wide identification and expression analysis of the ARF gene family in Apple. J Genet. 2014;93(4):785–97.
29. Chen J, Wang S, Wu F, Wei M, Li J, Yang F. Genome–wide identification and functional characterization of auxin response factor (ARF) genes in eggplant. Int J Mol Sci. 2022;23(11):6219.
30. Feng L, Li G, He Z, Han W, Sun J, Huang F, Di J, Chen Y, The. *ARF*, *GH3*, and *Aux/IAA* gene families in castor bean (*Ricinus communis* L.): Genome-wide identification and expression profiles in high-stalk and dwarf strains. Ind. Crops Prod. 2019;141:111804.
31. Zhang T, Ge Y, Cai G, Pan X, Xu L. WOX-ARF modules initiate different types of roots. BMC Plant Bio. 2023;42:112966.
32. Yamauchi T, Tanaka A, Inahashi H, Nishizawa NK, Tsutsumi N, Inukai Y, Nakazono M. Fine control of aerenchyma and lateral root development through AUX/IAA- and ARF-dependent auxin signaling. PNAS. 2019;23(1):622.
33. Yan M, Yan Y, Wang P, Wang Y, Piao X, Di P, Yang DC. Genome-wide identification and expression analysis of auxin response factor (ARF) gene family in *Panax ginseng* indicates its possible roles in root development. Plants. 2023;12(23):3943.
34. Hu J, Li X, Sun TP. Four class A AUXIN RESPONSE FACTORS promote tomato fruit growth despite suppressing fruit set. Nat Plants. 2023;9(6):706–19.
35. Shen X, He J, Ping Y, Guo J, Hou N, Cao F, Li X, Geng D, Wang S, Chen P, et al. The positive feedback regulatory loop of miR160-Auxin response factor 17-HYPONASTIC LEAVES 1 mediates drought tolerance in Apple trees. Plant Physiol. 2022;188(3):1686–708.
36. Wang C, Li X, Zhuang Y, Sun W, Cao H, Xu R, Kong F, Zhang D. A novel miR160a–GmARF16–GmMYC2 module determines soybean salt tolerance and adaptation. New Phytol. 2024;241(5):2176–92.
37. Wang K, Li J, Fan Y, Yang J. Temperature effect on rhizome development in perennial rice. Rice. 2024;17(1):32.
38. Li J, Yu X, Zhang S, Yu Z, Li J, Jin X, Zhang X, Yang D. Identification of starch candidate genes using SLAF-seq and BSA strategies and development of related SNP-CAPS markers in tetraploid potato. PLoS ONE. 2021;16(12):e0261403.
39. He M, Ma X, Zhou Y, Wang F, Fang G, Wang J. Combined metabolome and transcriptome analyses reveals anthocyanin biosynthesis profiles between purple and white potatoes. Int J Mol Sci. 2024;25(23):12884.
40. Bao Z, Li C, Li G, Wang P, Peng Z, Cheng L, Li H, Zhang Z, Li Y, Huang W, et al. Genome architecture and tetrasomic inheritance of autotetraploid potato. Mol Plant. 2022;15(7):1211–26.
41. 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(10):1365–75.
42. Zhang J, Zhang X, Tang H, Zhang Q, Hua X, Ma X, Zhu F, Jones T, Zhu X, Bowers J, et al. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. Nat. Genet. 2018;50(11):1565–73.
43. Song S, Hao L, Zhao P, Xu Y, Zhong N, Zhang H, Liu N. Genome-wide identification, expression profiling and evolutionary analysis of auxin response factor gene family in potato (*Solanum tuberosum* group Phureja). Sci Rep. 2019;9(1):1755.
44. Zhai Y, Shen X, Sun Y, Liu Q, Ma N, Zhang X, Jia Q, Liang Z, Wang D. Genome-wide investigation of ARF transcription factor gene family and its responses to abiotic stress in Coix (*Coix lacryma-jobi* L). Protoplasma. 2023;260:1389–405.
45. Schilling S, Kennedy A, Pan S, Jermiin LS, Melzer R. Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. New Phytol. 2020;225(1):511–29.
46. Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. Gene duplication as a major force in evolution. J Genet. 2013;92:155–61.
47. Zhang L, Guan R, Li G, Su K, Duan L, Sun W, Meng X, Wan H, Wang S, Chen S, et al. Genomic identification of ARF transcription factors and expression analysis in *Cannabis sativa* L. Ind. Crops Prod. 2022;186:115118.
48. Wu MF, Tian Q, Reed JW. Arabidopsis microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. Development. 2006;133(21):4211–8.
49. Li Y, Han S, Qi Y. Advances in structure and function of auxin response factor in plants. J Integr Plant Biol. 2023;65(3):617–32.
50. Pandey A, Misra P, Bhambhani S, Bhatia C, Trivedi PK. Expression of *Arabidopsis* MYB transcription factor, *AtMYB111*, in tobacco requires light to modulate flavonol content. Sci Rep. 2014;4(1):5018.
51. Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis Thaliana* seedling. Plant J. 2007;50(4):660–77.
52. Xu W, Chen Z, Ahmed N, Han B, Cui Q, Liu A. Genome-wide identification, evolutionary analysis, and stress responses of the GRAS gene family in castor beans. Int J Mol Sci. 2016;17(7):1004.

53. Tan H, Man C, Xie Y, Yan J, Chu J, Huang J. A crucial role of GA-regulated flavonol biosynthesis in root growth of *Arabidopsis*. Mol Plant. 2019;12:521–37.
54. Li M, Han X, Du H, Mu Z. Genome-wide identification and characterization of *MYB-CC* gene factor family in sorghum and their response to phosphate starvation. Russ J Plant Physiol. 2022;69:120.
55. He K, Du J, Han X, Li H, Kui M, Zhang J, Huang Z, Fu Q, Jiang Y, Hu Y. PHOSPHATE STARVATION RESPONSE1 (PHR1) interacts with JASMONATE ZIM-DOMAIN (JAZ) and MYC2 to modulate phosphate deficiency-induced jasmonate signaling in *Arabidopsis*. Plant Cell. 2023;35(6):2132–56.
56. Huang KL, Ma GJ, Zhang ML, Xiong H, Wu H, Zhao CZ, Liu CS, Jia HX, Chen L, Kjorven J, et al. The ARF7 and ARF19 transcription factors positively regulate *PHOSPHATE STARVATION RESPONSE1* in *Arabidopsis* roots. Plant Physiol. 2018;178(1):413–27.
57. Chen J, Li Y, Li Y, Li Y, Wang Y, Jiang C, Choisy P, Xu T, Cai Y. Pei dong. AUXIN RESPONSE FACTOR 18–HISTONE DEACETYLASE 6 module regulates floral organ identity in Rose (*Rosa hybrida*). Plant Physiol. 2021;186(2):1074–87.
58. Finn RD, Clements J, Eddy S. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39(2):29–37.
59. Yang M, Derbyshire MK, Yamashita RA, Marchler-Bauer A. NCBI's conserved domain database and tools for protein domain analysis. Curr Protoc Bioinf. 2020;69(1):e90.
60. Letunic I, Bork P. 20 Years of the SMART protein domain annotation resource. Nucleic Acids Res. 2018;46(1):493–6.
61. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, et al. Expasy: SIB bioinformatics resource portal. Nucleic Acids Res. 2012;40(1):597–603.
62. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007;35(2):585–7.
63. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13(8):1194–202.
64. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. Curr. Protoc. Bioinformatics. 2003;(1):2.3.1–2.3.22.
65. Hall BG. Building phylogenetic trees from molecular data with MEGA. Mol Biol Evol. 2013;30(5):1229–35.
66. Tian F, Yang DC, Meng YQ, Jin J, Gao G. PlantRegMap: charting functional regulatory maps in plants. Nucleic Acids Res. 2020;48(1):1104–13.
67. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li W, Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37(2):202–8.
68. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé 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):325–7.
69. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9(1):559.
70. 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):2498–504.
71. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023;51(1):638–46.
72. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods. 2001;25(4):402–8.
73. Zhang Z, Zhou D, Li S, Pan J, Liang J, Wu X, Wu XN, Krall L, Zhu G. Multiomics analysis reveals the chemical and genetic bases of pigmented potato tuber. J Agric Food Chem. 2023;71(43):16402–16.

Publisher's note