Genome-wide analysis and expression profiles of the StR2R3-MYB transcription factor superfamily in Solanum Tuberosum

Yuanming Li
Gansu Agricultural University
Kui Lin-Wang
Plant and Food Research
Zhen Liu
Gansu Agricultural University
Andrew C. Allan
Plant and Food Research
Junlian Zhang
Gansu Agricultural University
Yuhui Liu (lyhui@gsau.edu.cn)
Gansu Agricultural University https://orcid.org/0000-0002-3095-5858

Research article

Keywords: potato, MYB transcription factor, Expression profiles, Anthocyanin biosynthesis, Drought stress, EAR

Posted Date: May 18th, 2019

DOI: https://doi.org/10.21203/rs.2.9669/v1

Abstract

Background: MYB transcription factors comprise one of the largest families in plant kingdom, which play a variety of functions in plant developmental processes and defence responses. However, it has not been systematically studied in Potato (Solanum tuberosum), which is the most important non-cereal crop world-wide. Results: In the present study, a total of 108 StR2R3-MYB transcription factors were identified and further phylogenetically classified into 28 subfamilies, as supported by highly conserved gene structures and motifs. Collinearity analysis showed that the segmental duplication events played a crucial role in the expansion of StR2R3-MYB gene family. Synteny analysis indicated that 37 and 13 StR2R3-MYB genes were orthologous to Arabidopsis and wheat, respectively, and these gene pairs have evolved under strong purifying selection. RNA-seq data from different tissues and abiotic stresses revealed tissue-preferential and abiotic stress-responsive StR2R3-MYB genes. We further analyzed StR2R3-MYB genes might be involved in anthocyanin biosynthesis and drought stress by using RNA-seq data of pigmented tetraploid potato cultivars and drought-sensitive and -tolerant tetraploid potato cultivars under drought stress, respectively. Moreover, EAR motifs were found in 21 StR2R3-MYB proteins and 446 pairs of proteins were predicted to interact with 21 EAR motif-containing StR2R3-MYB proteins by constructing the interaction network with medium confidence (0.4). Additionally, Gene Ontology (GO) analysis of the 21 EAR motif-containing StR2R3-MYB proteins was performed to further investigate their functions. Conclusions: In this work, we systematically identified StR2R3-MYB genes by analyzing the potato genome sequence using a set of bioinformatics approaches. Genome-wide comparative analysis of StR2R3-MYB genes and their expression analysis identified members of this superfamily may be involved in tissue-specific development, anthocyanin biosynthesis and abiotic stress responses.

Background

Transcription factors are important regulators of gene expression by involving chromatin organization, DNA methylation, dimerization and sequence-specific DNA binding to control many aspects of plant for responding to abiotic and biotic stresses and modulate developmental and metabolic processes through activating or suppressing of target genes [1, 2]. Transcription factors can be classified into many different families on the basis of DNA-binding domain in the regulatory regions of downstream target genes [3].

The MYB transcription factors (TFs) comprise one of the largest TF families in the plant kingdom, are distinguished by the highly conserved MYB domain [4, 5]. MYB family members are divided into four subfamilies depending on the number of repeat (s) in the MYB domain, which consists of 1-4 imperfect tandem repeats in N-terminus, including 1R-MYB, R2R3-MYB, 3R(R1R2R3)-MYB and 4R-MYB with one, two, three and four MYB repeats [5, 6], of which, 1R-MYBs were referred to as MYB-related proteins [7]. Generally, the R2R3-MYB members are the predominant form found in higher plants [2]. The MYB repeat is approximately 50-53 amino acid residues in length, containing three regularly spaced tryptophan (or aliphatic) residues that together form a helix-turn-helix (HTH) fold [8, 9]. The second and third α-helix of each repeat interacts with the major DNA groove at the specific recognition site C/TAACG/TG during

transcription [10, 11]. On the contrary, the C-terminal modulator region is highly divergent that is responsible for the regulatory activity of the MYB proteins [12].

The MYB TF family is present in all eukaryotes. The first gene described as encoding a MYB domain-containing protein was "Oncogene" *v-MYB* identified in avian *myeloblastosis* virus (AMV) [13], subsequently, human c-MYB proto-oncoprotein and two vertebrate MYB TFs A-MYB and B-MYB were found and proved to take part in the regulation of cell differentiation, proliferation, and apoptosis [14, 15]. In plants, *Zea mays* C1, a homolog of mammalian *c-MYB* gene, is responsible for the regulation of anthocyanin biosynthesis [16]. C1 was the first plant TF as well as the first plant *MYB* gene to be cloned and functionally characterized.

Since the first plant *MYB* gene was characterized, an increasing number of plant MYB TF family members have been identified and characterized in a wide range of plant species such as in *Arabidopsis thaliana* [5], *Brassica napus* [17], *Solanum lycopersicum* [18], *Populus trichocarpa* [19], Maize [20], Soybean [21], Apple [22] and Pineapple [23]. Furthermore, multiple MYB proteins in plants have been functionally investigated in detail. The results from these works have proved that MYB proteins, especially the R2R3-MYB proteins, play an important role in many biological processes including primary and secondary metabolism, cell fate and identity, cell development and cell cycle, response to various biotic and abiotic stresses, and hormone synthesis and signal transduction [5, 24-26]. For example, a number of MYB TFs regulating anthocyanin biosynthetic pathway have been identified from many species, such as AtMYB75 (or PAP1), AtMYB90 (or PAP2), AtMYB113 and AtMYB114in *Arabidopsis* [27, 28], MdMYBA,MdMYB1, MdMYB10 and MdMYB110ain apple [29, 30], SlMYB12 in tomato [31] and StAN1 in potato [32, 33]. An amount of MYB TFs involved in regulating responses to environmental stresses were also studied such as GbMYB5 confers drought tolerance in cotton and transgenic tobacco plants [34], LcMYB1 confers salt tolerance in transgenic *Arabidopsis* [35], overexpression of StMYB1R-1 in potato plants improved plant tolerance to drought stress [36].

Potato (*Solanum tuberosum* L.), originated from the Andean regions of Peru and Bolivia [37], is now grown worldwide as the third most important food crop plant after wheat and rice. In recent decades, pigmented potato cultivars attracted a lot of attention from consumers and researchers worldwide, as it is a rich source of anthocyanins, in particular acylated derivatives [38], which not only have key roles in protection against UV radiation and cold temperatures, and response to drought stress [39-41], but also have potential health benefits such as protection against some cancers and neuronal and cardiovascular diseases in humans [42].

Potato is very sensitive to drought with yield reduction beginning under moderate deficit of soil moisture [43, 44], genetic and molecular biological approaches were used to overcome potato yield loss under drought stress [36, 45-48], however, there is still little known about drought stress tolerance mechanisms in potato plants.

Since R2R3-MYB TFs play an important role in wider plant biological processes, and were studied in a numerous plant species, but there are only very limited reports on the functional characterization of R2R3-

MYB TFs in potato [32, 33, 36, 46, 49, 50]. Thus it is meaningful to uncover the R2R3-MYB TFs' function, evolution and expression profiles in potato based on the published sequence data [51].

In the current study, we totally identified an expanded StR2R3-MYB family with 108 members, and comprehensive analysis of the phylogenetic relationships, sequence features, gene duplications, chromosome distribution, motif recognition were further investigated. In addition, we performed a comprehensive expression analysis of *StR2R3-MYB* genes in different tissues and under abiotic stresses in doubled monoploid (DM) potato, as well as in white and pigmented potato cultivars and in drought-sensitive and -tolerant cultivars of tetraploid potato using RNA-seq data to identify *StR2R3-MYB* genes closely associated with spatial distribution, anthocyanin biosynthesis and multiple stresses. Moreover, 21 StR2R3-MYB genes containing EAR motif were obtained and an interaction network between these EAR motif-containing R2R3-MYBs with other potato genes were constructed. Additionally, Gene Ontology (GO) analysis of the 21 EAR motif-containing StR2R3-MYB proteins was performed to investigate their functions. The findings should inform the characterization of StR2R3-MYB superfamily and provide valuable information for further functional elucidation of these genes in potato.

Methods

Identification of potato *MYB* family genes

Hidden Markov Model (HMM) profile for the MYB binding domain (PF00249) was downloaded from the Pfam database (http://pfam.xfam.org/) [52] and used it as a query to search potato genome from the Potato Genome Sequencing Consortium (PGSC, http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml) by HMMER 3.1 (http://hmmer.org/download.html) with a cutoff value of 0.01. Additionally, 126 R2R3-MYB superfamily protein sequences in *Arabidopsis thaliana* were obtained from The Arabidopsis Information Resource (TAIR; http://www. arabidopsis.org/) based on previous work [2, 53]. The Pfam database, SMART website (http://smart.embl-heidelberg.de) [54] and the NCBI-CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) [55] were used to confirm the obtained MYB proteins. All potato R2R3-MYB proteins were manually inspected to ensure that the gene models contained two MYB domains.

Sequence analysis and structural characterization of *StR2R3-MYB* genes

The exon-intron organizations of the *StR2R3-MYB* genes, including intron distribution patterns, phases and intro-exon boundaries, were graphically displayed by the Gene Structure Display Server (GSDS2.0 (http://gsds.cbi.pku.edu.cn/) [56] using the CDS and genome sequence of *StR2R3-MYB* genes. The conserved motifs of StR2R3-MYB TFs was predicted by using the MEME Suite web server (http://meme-suite.org/tools/meme) analysis [57], with the following parameters: the maximum number of motifs was set to identify 20 motifs and optimum width of motifs was set from six to 100 amino acids.

Analyses of chromosome distribution, gene duplication and synteny for *StR2R3-MYB* genes

The chromosome distribution information of the *StR2R3-MYB* genes were obtained from the database of potato genome downloaded from PGSC. MapChart software [58] was used for graphical presentation of *StR2R3-MYB* gene's chromosomal location. Tandem duplication and segmental duplication between potatogenes as well as the synteny block of the orthologous *R2R3-MYB* genes between potato and *Arabidopsis*, wheat was obtained by using Multiple Collinearity Scan toolkit (MCscanX) (http://chibba.pgml.uga.edu/mcscan2/) [59] and visualized using the circos v0.69 [60]. To further estimate duplication events of StMYB genes, the synonymous (Ks) and nonsynonymous (Ka) were calculated by using KaKs_Calculator 2.0 (https://sourceforge.net/projects/kakscalculator2/) [61].

Phylogenetic analysis and classification of potato StR2R3-MYB proteins

The full-length amino acid sequences of StR2R3-MYB and AtR2R3-MYB proteins derived from PGSC and Ensembl Plants database (http://plants.ensembl.org/index. html) were used for phylogenetic analysis. Multiple sequence alignments of these R2R3-MYB proteins were performed using Clustal X with default parameters. An unrooted neighbor-joining (NJ) phylogenetic tree was constructed using MEGA 7.0 software [62], with the following parameters: Poisson model; pairwise deletion; and 1000 bootstrap replications. The potato R2R3-MYB proteins were classified into different groups according to the topology of phylogenetic tree.

Plant materials and treatments

A purple potato cultivar 'Heimeiren' (HM, purple skin and purple flesh), a white potato cultivar 'Xindaping' (XD, white skin and white flesh) and a red potato cultivar 'Lingtianhongmei' (LT, red skin and red flesh) were grown in a greenhouse at Gansu Agricultural University, China. HM and XD are local cultivars in Gansu province, LT was cultivated by Potato Research Center of Inner Mongolia Academy of Agricultural and Animal Husbandry Sciences and Inner Mongolia Lingtian Biotechnology Co., Ltd., China. Six fresh tubers (diameter 4-5cm) were harvested, and cleaned with sterilized water. Skin tissue was carefully separated from cortex tissue using a scalpel to minimize flesh contamination, flesh tissue was isolated with at least 5 mm distance from the skin to eliminate skin contamination. The skin and flesh of these potatoes were then immediately frozen in liquid nitrogen and stored in a -80 ºC freezer for later use.

For drought stress, a drought-sensitive cultivar 'Atlantic' (A) and a drought-tolerant cultivar 'Qingshu No.9' (Q) were field-grown with rainproof shed at Dingxi Academy of Agricultural Science, Gansu. A was cultivated by Dingxi Potato Research Institute and Q was cultivated by Qinghai Academy of Agriculture and Forestry Sciences of Qinghai Province, China. Two treatments were performed: drought stress and watered control with a drip irrigation system to control of watering. Each treatment had a randomized complete block design with three blocks (replications) and ten plants per block. For the first 4 weeks, all plants in both treatments were watered optimally and equally, after seedling, the plants under drought stress treatment went unwatered for whole growth period, while the plants under control treatment were still irrigated optimally during whole growth period until the foliage began to die naturally. The foliage of three pooled plants for each replicate was collected at early flowering stage, full-blooming stage and

flower-falling stage, respectively, and then immediately frozen in liquid nitrogen and stored in a -80 °C freezer for later use.

RNA isolation and quantitative real-time RT-PCR

Total RNA extractions of skin and flesh from three potato cultivars, and of leaves from two potato cultivars, were carried out using the PureLink Plant RNA Reagent Kit (Invitrogen, USA) according to the manufacturer's instructions. The RNA was monitored by gel electrophoresis and quantified by using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, USA). Elimination of genomic DNA contamination and first-strand cDNA synthesis were carried out using oligo dT according to the manufacturer's instructions (SuperScript III, Invitrogen, USA). qRT-PCR was conducted on CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, CA, USA) using SYBR® Premix Ex Taq™ II (Takara Bio, Inc., Japan). qPCR conditions were as follows: 30s at 95 °C, followed by 40 cycles of 5 s at 95 °C, 30 s at 60 °C, followed by 65–95 °C melting curve detection. The qPCR efficiency of each gene was obtained by analysing the standard curve of a cDNA serial dilution. *StEF-1α* (AB061263) [63] was used for template normalization. Relative abundance was calculated with comparative Ct ($2^{-\Delta\Delta Ct}$) method. The primers are listed in Additional files 1: Table S1.

RNA-Seq data analysis

Total RNA of the aforementioned samples with three biological replicates was chosen for further RNA-seq library construction. Next-Generation Illumina sequencing were performed by Biomarker Technologies Corporation (Beijing, China). High quality clean reads were obtained by trimming the raw reads filtering out contaminants, adapters, phred scores less than 20 and uncertain bases N. The cleaned data were aligned to PGSC_DM_v3.4 gene models downloaded from PGSC by Bowtie2 (v2.2.9). The number of mapped clean reads were counted (by fragment) and subjected to differential expression analysis using the edgeR package (http://www.r-project.org/). Genes with the absolute value of log2FC (fold change) not less than 1 and a false discovery rate (FDR) <0.05 was considered as significant DEGs.

DEGs were annotated against nonredundant database (nr), SwissProt/UniProt Plant Proteins, KEGG, SwissProt/UniProt Plant Proteins, COG/KOG (Cluster of Orthologous Groups of proteins) and potato protein database (ftp:// ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Stuberosum) by BLASTX and the cut-off E-value was 1.0e-5. DEGs were then subjected to enrichment analysis of GO functions and KEGG pathways.

The Illumina RNA-seq data were also downloaded from the PGSC to study the expression patterns of MYB genes in various tissues and stress treatments. TBtools software [64] was used to generate the heatmap.

Identification of EAR motif-containing StR2R3-MYB proteins

The StR2R3-MYB proteins were manually searched to identify candidate genes containing motifs DLNxxP or LxLxL [65]. Specific protein interactions were constructed by using online STRING v11.0 software (Search Tool for the Retrieval of Interacting Genes/Proteins, http://string-db.org/) [66] with combined score≥400 (medium confidence). The interaction network was visualized by Cytoscape v3.7.1 [67]. The GO enrichment analysis was conducted by using the topGO package in R project.

Availability of supporting data

The raw data of the transcriptome analysis used in this study was submitted to the Sequence Read Archive (SRA) at NCBI under Project ID PRJNA528685 and PRJNA529980, and the expression data was also available Potato Genome Sequencing Consortium (PGSC, http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml). The accession number and the website listed above were publicly available. The databases used in this study were publicly accessible and no special permissions were required.

Results

Identification and classification of the potato *MYB* genes

To identify the putative MYB proteins in potato genome, a hidden Markov model (HMM) search was performed using the HMM profile of the MYB binding domain against the potato genome downloaded from PGSC database. This HMM search resulted in identification of 319 potato gene models (DMGs). Subsequently, the Pfam, SMART website and NCBI-CDD analyses were performed to verify the presence of the MYB domains. Finally, a total of 233 non-redundant potato proteins were identified, including 108 R2R3-MYB proteins (2R–MYB) (Additional files 2: Table S2), 121 MYB-related proteins (1R–MYB) proteins, 3 R1R2R3-MYB proteins (3R–MYB) and 1 4R–MYB proteins. The lengths of StR2R3-MYB ranged from 136 to 523 amino acids, the molecular weights were between 1.61 kDa and 5.73kDa and the predicted pI values ranged from 4.59 to 9.99.

The functions of several *Arabidopsis* MYB proteins, especially the R2R3 MYBs, have been well characterized, and several functional clades have been identified by phylogenetic analysis of *Arabidopsis* MYB proteins [5]. To further understand the evolutionary relationship of R2R3-MYB proteins in potato and *Arabidopsis* and speculate the putative functions of the potato R2R3-MYB proteins based on the functional clades which have been identified in *Arabidopsis*, unrooted Neighbor–Joining (NJ) phylogenetic trees were constructed based on R2R3-MYB subfamilies (Fig. 1). The resulting trees generated 31 clades for R2R3-MYB subfamilies (named C1–C31) according to the sequence similarity and topology. Remarkably, 27 clades included different numbers of R2R3-MYB proteins from the two species, while two clades (C7 and C15) only contained *Arabidopsis* R2R3-MYB members, two clades (C23 and C24) only contained potato R2R3-MYB members. Additionally, an unrooted phylogenetic tree of StR2R3-MYB TFs was also constructed individually, as shown in Fig. 2A, 28 clades for R2R3-MYB

subfamilies (named A1–A28) were generated, and 67 R2R3-MYB TFs were classified into 16 subgroups based on the phylogenetic analysis of *Arabidopsis* and potato in Figure 1.

Structure of *StR2R3-MYB* genes and conserved motifs

Since the analyses of gene structural diversity might be helpful for understanding the gene functions and evolution, the structural diversity of *StR2R3-MYB* genes was also investigated (Fig. 2B). Three members in A1 contained more exons than other subfamilies with 9-11 exons, four *StR2R3-MYB* genes (*PG0007994*, *PG0024983*, *PG0003316* and *PG0034577*, which belonged to SG22) only have one exon, furthermore, the results showed that exon/intron structures were highly conserved in the same subfamilies of *StR2R3-MYB* genes, suggesting these conserved features play crucial roles for group-specific functions (Fig. 2B).

We used online program MEME to search for conserved motifs shared by these StR2R3-MYB proteins to further study the diversification of these *StR2R3-MYB* genes in potato. In total, 20 conserved motifs were identified and designate as motif 1 to 20 (Fig. 2C). The details of the 20 motifs were referred in Additional files 3: table S3. Most of R2R3-MYB protein sequences had four highly conserved motifs and conserved motif orders. 107 out of 108 R2R3-MYB protein sequences all had motif 2, motif 15, motif 1, and motif 3 in that order, containing 23, 11, 50, and 18 amino acids, respectively, motifs 2, 15 and left part of motif 1 were composed of the R2 repeat, and right part of motif 1 and motif 3 were composed of the R3 repeat. In addition to the highly conserved motifs composed of MYB domains (R2 and R3 repeats), other conserved motifs were also found in most of the closely related members (Fig 2C and Additional files 3: Table S3), suggesting possible functional similarity among these proteins, they may play crucial roles in the regulation of target genes at transcriptional level.

Sequence conservation within the MYB domain of StR2R3-MYB proteins

The R2R3-MYB proteins share significant sequence conservation within the MYB domain regions. To investigate the homologous domain sequence features into the potato R2R3-MYB domains, multiple alignments (ClustalX 2.1) were performed using the amino acid sequences of R2 and R3 repeats and sequence logos were produced to analyze the level of conservation of R2 and R3 repeats in the R2R3-MYB proteins within each residue position (Fig. 3). The results revealed that five and four conserved amino acid residues were identical among all the members detected in the R2 and R3 MYB repeat regions, respectively (Fig. 3 and Additional files 4: Table S4).

Three regularly spaced tryptophan (Trp., W) residues were contained in typical MYB proteins, which play significant roles by interacting with specific DNA sequences [10]. For 106 out of the 108 potato R2R3-MYB proteins, the R2 repeat sequences contained three Trp residues (Fig. 3), whereas two R2R3-MYB TFs (PG0026758 and PG0002828) only contained last two Trp residues, the first Trp was replaced by phenylalanine (Phe., F) (Fig. 3 and Additional files 4: Table S4).

In R2 repeat, insertion of a glycine (Gly., G) residue was observed between Gly-24 and asparagine (Asn., N)-25 in five R2R3-MYB genes, an aspartic acid (Asp., D) residue in PG0002828 and PG0026758, an Asn residue and a threonine (Thr., T) residue insertion in PG0018750 and PG0035400 at same position were also observed (Additional files 4: Table S4). In addition, the insertion of the leucine (Leu., L) residue between the second and third helices of R2 repeat was observed in 90 StR2R3MYB proteins, which was an important step for the origin for plant-specific R2R3 MYB proteins [68], whereas a proline (Pro., P) residue was inserted at same position in PG0015087, furthermore, a 22-residue insertion (SFLFLLDLYSQSEFRARLIVWL) was observed between the second and third helices of R2 repeat in PG0001325.

In the R3 repeats, the first Trp residue was generally replaced by Phe or isoleucine (Ile., I) (Fig. 3). The second Trp residue was conserved in all the members, the third Trp residue was conserved in most of the members, except five and four R2R3-MYB members whose third Trp residue was replaced by Phe and tyrosine (Tyr., Y), respectively. Only one valine (Val., V) residue insertion was observed between Gly-4 and Lys-5 in PG0025720 (Additional files 4: Table S4).

In addition to the highly conserved Trp residues, glutamic acid (Glu., E)-10, Asp-11 in the R2 repeat, Glu-10, Gly-22, arginine (Arg., R)-35, Asn-38, lysine (Lys., K)-41, Asn-42 in the R3 repeat were also complete conserved, Leu-50 conserved in the linker region of 107 StR2R3-MYB TFs, except the Leu was replaced by Arg in PG0034577. The result showed that R3 repeat was more conserved than the R2 repeat, furthermore, the third helix of each repeat in the MYB domain was the most conserved and the first part of the HTH domain in each repeat was less conserved among the 108 StR2R3-MYB proteins.

Chromosomal location and gene duplication of potato StR2R3-MYB genes

Genome chromosomal location analyses revealed that potato *R2R3-MYB* genes were distributed throughout all 12 chromosomes, but the distribution appeared to be uneven (Fig. 4). A total of 105 *R2R3MYB* genes were mapped on 12 chromosomes, whereas three *R2R3-MYB* genes (*PG0046667*, *PG0025720*, *PG0042799*) were remained unmapped. Chr05 encompassed the largest number of 13 *R2R3-MYB* genes, followed by Chr06 with 12 *R2R3-MYB* genes. Relatively high densities of *R2R3-MYBs* were observed in some chromosomal regions, including the top and bottom of chromosomes 4, 5, 10, and 12; the bottom of chromosomes 2, 3, 6 and 7, by contrast, most of central chromosomal regions lacked *StR2R3-MYBs* (Fig. 4). The pattern is similar with the genome chromosomal location of *SlMYB* [18].

Gene duplication has long been recognized as the origin of multigene families, and has been proved to be a prominent feature of plant genome evolution. To investigate the gene duplication events in potato, tandem duplications and segmental duplications were identified by BLASTP and MCScanX method. Among the *StR2R3-MYB* genes, 6 pairs of tandemly duplicated genes were identified, of which, *PG0017525* was tandemly duplicated with a *MYB-related* gene *PG0017526* (Fig. 4 and Additional files 5: Table S5). Meanwhile, 29 segmental (26.9%) duplication pairs were found between *StR2R3-MYB* genes (Fig. 5 and Additional files 5: Table S5) with two exception (*PG0027157* and *PG0027575*) were duplicated with *MYB-related* genes *PG0004610* and *PG0008340*, respectively. Most segmentally duplicated *MYBs*

are located in collinear regions on chr2, 3, 5 and 6. (Fig. 5), and most of these segmentally duplicated genes were belonged to S1, 2, 9, 11, 18 and 20 (Additional files 5: Table S5) , suggesting the rapid expansion of *R2R3-MYB* genes in these function-specific groups in potato might be strongly linked to adaptation strategies in response to challenging environments.

To further explore the potential evolutionary processes of *StR2R3-MYB* gene family, two comparative syntenies of potato with *Arabidopsis thaliana* and *Triticum aestivum* (wheat), which were belonged to dicotyledon and monocotyledon respectively (Fig. 6A and 6B) were constructed. The results showed that 37 orthologs between potato and *Arabidopsis*, 13 orthologs between potato and wheat were identified, respectively (Additional files 6: Table S6).

The substitution rate (Ka/Ks) was an effective index to determine the positive selection pressure after duplication, Ka/Ks<1 means purifying selection, Ka/Ks = 1 stands for neutral selection, while Ka/Ks>1 signifies positive selection [69]. Thus, the Ka, Ks and Ka/Ks of each gene pair were calculated and results showed that all the segmentally and tandemly duplicated *MYB* gene pairs had Ka/Ks values of less than 1, except one pair *PG0005918* and *PG0022689* with Ka/Ks =1.01, implying that most of these genes had evolved under the effect of purifying selection (Additional files 5: Table S5 and Additional files 6: Table S6), the average Ka/Ks value of tandem duplication genes (0.31) is a little lower than that of segmental duplication genes (0.36) (Fig. 6E), and the Ka/Ks values of gene pairs between potato and *Arabidopsis*, potato and wheat orthologs were 0.3315 and 0.3722, respectively (Additional files 6: Table S6).

Expression profiles of *StR2R3-MYB* genes in different tissues

To understand the tissue-specific expression patterns of the *StR2R3-MYB* genes, the transcript abundance of 13 different tissues (leaves, roots, shoots, callus, tubers, sepals, stamens, stolons, flowers, petioles, petals, carpels and fruit) of DM potato was analyzed by using transcriptome data downloaded from the PGSC. The results showed that a total of 7.4% *StR2R3-MYB* genes (8/108) were highly expressed in all tissues (Additional files 7: Fig. S1), while 23.89% (43/108) *StR2R3-MYB* genes showed low expression with FPKM less than 2, of which, 7 genes showed no expression in all 13 tissues, some other genes displayed a tissue-specific expression pattern, such as 7 genes in leaf, 13 genes in root, 20 genes in shoot, 12 genes in tuber, 18 genes in stolon, and 25 genes in flower showed the higher transcript abundances with an FPKM > 5 (Additional files 7: Fig. S1 and Additional files 8: Table S7).

Expression profiles of *StR2R3-MYB* genes in pigmented potato cultivars

To investigate the roles of these StR2R3-MYB TFs that might be involved in anthocyanin biosynthesis, expression profiles of *StR2R3-MYB* genes in different tuber tissues (skin and flesh) of tetraploid pigmented potato cultivars were examined. StAN1, StMYBA1 and StMYB113 were found to influence anthocyanin biosynthesis in tobacco by transient assays [33]. in PGSC database, the best blast hits for StAN1, StMYBA1 and StMYB113 were PGS0013965, PG0013966 and PG0019217, respectively, PG0013965 and PG0013966 contained single MYB domain, while MYB domain was not found in PG0019217, their expression was also analyzed in this RNA-seq dataset. The results showed that in

potato tuber skin, 27 *StR2R3-MYB* genes showed no expression, expression of 28 *StR2R3-MYB* genes was very low with an FPKM value less than 1. Compared with white skin of white cultivar XD, seven *StR2R3-MYB* genes, homologous to *AtMYB102* (S11), *AtMYB69* (S21), *AtMYB13* (S2), *AtMYB74* (S11), *AtMYB93* (S24) and *AtMYB58* (S3), showed higher expression in pigmented skin of HM and LT, on the contrary, six *StR2R3-MYB* genes, homologous to *AtMYB15* (S2), *AtMYB38* (S14), *AtMYB36* (S14), *AtMYB20* and *AtMYB37* (S14), were only highly expressed in white skin. The *PG0013965* showed higher expression in pigmented skin than in white skin, while the FPKM was less than 5, whereas the expression of *PG0013966* was higher in white skin with an FPKM = 5, the *PG0019217* showed no expression (Fig. 7A and Additional files 9: Table S8).

In tuber flesh, 38 *StR2R3-MYB* genes were not expressed and the expression of 39 *StR2R3-MYBs* was lower than 1. The *PG0013965* was highly expressed in purple and red flesh, while the *PG0013966* and *PG0019217* showed lower expression in tuber flesh. In addition to *PG0013965*, six genes, which are homologous to *AtMYB48*, *AtMYB3*(S4), *AtMYB36* (S14) and *AtMYB79*, showed higher expression in pigmented flesh, while two genes, homologous to *AtMYB27* and *AtMYB15* (S2), showed higher expression in white flesh (Fig. 7B and Additional files 9: Table S8).

Furthermore, an interaction network of StR2R3MYB proteins, which showed expression with an FPKM value more than 1 in skin and flesh, was built using STRING software. The results showed that eight StR2R3-MYB TFs can directly interact with StAN1 with combined score > 400, of which, two MYBs (PG0007325 and PG0018113) were homologous to AtMYB67, two MYBs (PG0024822 and PG0018427) belonged to S21, two MYBs (PG0003316 and PG0024983) belonged to S22, one MYB (PG0027190) belonged to S23 and one MYB (PG0011243) was homologous to AtMYB72. Interestingly, the two MYBs (PG003316 and PG0024983), which belonged to S22, were also interacted with StMYBA1 (Fig. 7C), suggesting the two MYBs might play important roles in anthocyanin biosynthesis. Of these MYB TFs, *PG0007325* showed higher expression in pigmented skin, *PG0018113* showed higher expression in purple flesh and pigmented skin, *PG0024822* and *PG0018427* in S21 were higher expressed in white skin, *PG0003316* and *PG0024983* in S22 were all highly expressed in all tissues, *PG0027190* was up-regulated in pigmented flesh and *PG0011243* was higher expressed in pigmented skin.

Expression profiles of *StR2R3-MYB* genes under abiotic stress

To understand the *StR2R3-MYB* genes in response to abiotic stresses, expression profiles of *StR2R3-MYB* genes under abiotic stresses were also examined by using transcriptome data downloaded from the PGSC. The results showed that 55, 57 and 53 genes were differentially expressed with an FPKM > 1 under salt, mannitol and heat treatment, respectively (Additional files 10: Fig. S2), of which, 47, 45 and 35 genes were up-regulated, respectively. 13 genes were responded to three stress treatments with an FPKM > 1 and |log2(FC)| > 1, of which three genes (*PG0013405*, *PG0017223*, *PG0016292*) belonged to A20 (S2), two genes (*PG0026758* and *PG0002828*) belonged to A1, two genes (*PG0006930* and *PG0019535*) belong to A13 (S1), the other genes belonged to A9, A10 (S20), A14, (S14), A18 (S24), A26 (S11),

respectively. 25 genes were responded to at least two stress treatments with an FPKM > 1 and |log2(FC)| > 1 (Additional files 11: Table S9).

To further investigate the *StR2R3-MYB* genes might be in response to drought stress, one tetraploid drought-sensitive cultivar “Atlantic” (A) and one tetraploid drought-tolerant cultivar “Qingshu No.9” (Q) were subject to drought stress. The RNA-seq data showed that 42 *StR2R3-MYB* genes were up- or down-regulated in response to drought stress, of which, 15 genes were highly expressed in Q with an FPKM > 5 and |log2(FC)| > 1 at flower-falling stage (Fig. 8 and Additional files 12: Table S10). There are not much differences of these genes in A and Q at early flowering stage, then these genes were up-regulated in Q at full-blooming stage and highly up-regulated in Q at flower-falling stage under drought stress, while four genes were down-regulated in Q beginning from early flowering stage until flower-falling stage, these genes might be involved in drought stress, which is worth further investigating.

We also analyzed expression patterns of 12 selected *StR2R3-MYB* genes from different subfamilies by quantitative real-time PCR (qPCR), which showed relatively higher expression by RNA-seq related to anthocyanin biosynthesis and drought stress to further confirm the reliability of RNA-seq database. The results showed that the expression pattern determined by qPCR and FPKM is consistent, as shown in Figure 9,

PG0013965 (*StAN1*) was up-regulated in pigmented tissues, the expression in pigmented skin was lower than that in pigmented flesh, while it was not expressed in A and Q under drought stress. *PG0017223* is exclusively highly expressed in white skin, whereas *PG0015536* and *PG0030548* were only up-regulated in pigmented flesh, *PG0009033* was higher expressed in Q during three stages, *PG1026177*, *PG2026177*, *PG0024983* and *PG0013405* were highly expressed in Q at flower-falling stage, two genes *PG0015536* and *PG0030548*, showed opposite expression pattern with higher expression in A at flower-falling stage. Although the relative expression of the selected genes varied between RNA-Seq dataset and qPCR analysis, a high correlation (R^2 = 0.8394) described by a simple liner regression equation y = 0.8762x + 0.0332, suggests good consistency between the two analysis methods.

Identification of EAR motif-containing StR2R3-MYB proteins in potato

The Ethylene-responsive element binding-factor-associated amphiphilic repression (EAR) motif, defined by the consensus sequence patterns of either LxLxL or DLNxxP is the most dominant transcriptional repression motif identified in plants. In our study, 20 members of StR2R3-MYB family have been identified containing at least one LxLxL type of EAR motif, and one StR2R3-MYB protein contained DLNxxP (Table 1), these StR2R3-MYBs were belonged to S2, S4, S9, S18, S20, S22, and two StR2R3-MYBs were homologous to AtMYB41 and AtMYB48, PG0020071 (AtMYB15-like) have two LxLxL motifs. The 22 EAR motif sites were mostly found in the C-terminal region (10 out of 22) and in the N-terminal (9 out of 22), and at lower frequency in the middle (3 out of 22) regions (Table 1).

The core EAR motif sites comprising nine amino acids were analyzed by MEME website, among the 21 LxLxLx motifs, positions 1, 2, 4 and 9 were more frequently occupied by Gln, Leu, Ile and His residues, Glu

and Ser, His and Pro are more abundant in position 6 and 8, respectively (Fig. 10A). Subsequently, the protein interactions of EAR motif-containing StR2R3-MYB TFs with other potato proteins were examined by using STRING software with combined score > 400 (Fig. 10B and Additional files 13: Table S11). The results showed that the 21 EAR motif-containing StR2R3-MYB proteins were involved in at least nine interaction possibilities, the StR2R3-MYB proteins, belonged to the same subfamilies, appeared to have similar functions by regulating common target genes (Fig. 10).

Furthermore, GO assignments were used to predict the functions of the EAR motif-containing StR2R3-MYB proteins by classifying them into three independent ontologies in terms of biological process (BP), molecular function (MF) and cellular components (CC). As shown in Additional files 14: Table S12, the functions of these proteins were related to biological process (11 out of 21, 52.4%) including biological regulation (8), cellular process (8), metabolic process (8), response to stimulus (11) and developmental process (5) etc.; for molecular function (21 out of 21, 100%), nucleic acid binding (21) was the most represented GO term, followed by calmodulin binding (1); for cellular component (6 out of 21, 28.6%), major categories were cell (6) and cell part (6).

Discussion

The MYB family is one of the largest transcription factors families, which have been identified to be involved in various plant physiological and biochemical processes. R2R3-MYBs are the predominant form found in higher plants, they play important roles in the primary and secondary metabolism, developmental processes and responses to biotic and abiotic stresses [5]. In the present work, we performed a genome-wide investigation of the *R2R3-MYB* gene family in potato. A total of 108 *R2R3-MYBs* were identified, most of which remain to be functionally characterized. Subsequently, a phylogenetic tree was constructed to evaluate the evolutionary relationships of *R2R3-MYB* genes between potato and *Arabidopsis* to identify the evolution and possible functions of StR2R3-MYBs. Detailed analyses including gene structures, conserved motif composition, chromosomal location and gene duplication events were performed to further identify the characteristics of these *StR2R3-MYB* genes. In addition, the expression profiles of *StR2R3-MYB* genes in different tissues, in skin and flesh of pigmented potato cultivars as well as in response to different abiotic stresses, especially drought stress were analyzed. Finally, 21 EAR motif-containing StR2R3-MYB proteins were identified and the interaction network of these proteins with their interacting proteins were analyzed to investigate the transcriptional regulatory mechanisms of potato StR2R3-MYB gene family.

Phylogenetic analysis of the potato *R2R3-MYB* genes

Generally, paralogous and orthologous functional relationships can be confirmed by knowing the functions of certain members, it is likely that members within a subgroup may have common evolutionary origins and a conserved function. According to the phylogenetic comparative analysis, most StR2R3-MYBs are clustered with orthologs of AtR2R3-MYBs in different subgroups (Fig.1 and Fig. 2). For example, A17 consists of PG0009033, PG0004371 and PG0007304, which were clustered with AtMYB11,

AtMYB12, and AtMYB111 (S7), implicating in the control of flavonoid accumulation [70-72], A16 comprises four members clustered with AtMYB123/TT2 (S5), which controls the biosynthesis of proanthocyanidins (PAs) in the seed coat of *Arabidopsis* [73].

Duplication contributed to the *StR2R3-MYB* gene expansion

Gene duplication contributes significantly to the expansion of MYB genes in the plant kingdom, which lead to the diversification and evolution of genes. Our results showed that 6 (5.6%) and 29 (26.9%) *StR2R3-MYB* genes were identified as tandem duplication and segmental duplication, respectively, indicated that segmental duplication event was a major cause of expansion of StR2R3-MYB genes.

Most *StR2R3-MYB* genes were tandemly and segmentally duplicated within one cluster to expand their subfamilies, the duplication event was mostly occurred in the genes belong to S1, S2, S18 and S20 (Additional files 5: Table S5), exceptionally, one *StR2R3-MYB* gene *PG0017525* was tandemly duplicated with *MYB-related* gene (*PG0017526*), two *StR2R3-MYB* genes (*PG0027157* and *PG0027575*) were segmentally duplicated with two *MYB-related* genes (*PG0004610* and *PG0008340*).

Some evolving members may have lost their original function(s) and/or acquiring new function(s) to enhance the adaptability of plants or become pseudogenes as a result of duplication event [74]. In our work, we found that some pairs of *StR2R3-MYB* genes displayed a different expression level in different tissues and under abiotic stresses, whereas some pairs of *StR2R3-MYB* genes were presented similar expression pattern. For example, in one pair, *PG1026177* and *PG2026177* were all highly expressed in drought-tolerant cultivar Q at flower-falling stage under drought stress, one gene *PG0026176*, which tandemly duplicated with *PG2026177*, was not expressed under drought stress.

Identification of the StR2R3-MYB TFs might be involved in anthocyanin biosynthesis

In *Arabidopsis*, 12 representative members of MYB family proteins, such as MYB75, MYB90, MYB113, MYB114, MYB3, MYB4, MYB7, MYB11, MYB12, MYB111, MYB32, and MYB60 were involved in controlling some major phenylpropanoid biosynthesis genes [2, 5, 72]. In our work, four members (A18) were clustered with AtMYB3, AtMYB4, AtMYB7 (S4), three members (A17) were clustered with AtMYB11, AtMYB12, and AtMYB111 (S7), one MYB PG0028125 was clustered with AtMYB60. In the RNA-seq database of pigmented skin and flesh, three members of A17 were not expressed in either white or pigmented skin, while *PG0009033* was up-regulated in pigmented flesh, in A18, *PG0030548* was up-regulated in pigmented skin and flesh, while last three members showed no difference between white and pigmented tissues. *PG0028125*, which is homologous to *AtMYB60*, showed little expression in white and pigmented tissues.

Our previous work showed that in transient assays, StMYB44 represses anthocyanin accumulation in leaves of *N. tabacum* and *N. benthamiana* by directly suppressing the activity of the dihydroflavonol reductase (DFR) promoter. StMYB44-1 (PG0003316) showed stronger repressive capacity than StMYB44-2 (PG0007994) with both predicted proteins containing the repression-associated EAR motif with some

variation (Liu et al. 2018). In our RNA-seq dataset, the two *StMYB44s* were all expressed in white and pigmented skin and flesh, as well as another *AtMYB44-like* gene *PG0024983*; one gene (*PG0030548*) belonged to S4, which was reported to repress flavonoid/anthocyanin accumulation, was highly expressed in pigmented flesh. It's also reported that the MdMYB15L, homologous to AtMYB15, might act upon MBW complexes to repress anthocyanin accumulation by interacting with MdbHLH33 [75], in our dataset, two genes *PG0017223* and *PG0016292*, homologous to *AtMYB15* (S2), were highly expressed in white skin. The results suggest that hierarchical interactions among R2R3-MYB regulators are complicated and requires further investigation, which was revealed by Albert *et al*. (2014) [76] that the network of transcriptional activators and repressors involved in anthocyanin accumulation was conserved in eudicots and by Zhou *et al*. (2018) [77] that activator-type R2R3-MYB genes induce a repressor-type R2R3-MYB gene to balance anthocyanin and proanthocyanidin accumulation.

The StR2R3-MYB TFs respond to abiotic stresses

The MYB TFs play essential roles in the regulation of gene expression to cope with environmental stresses [78, 79]. 67 *StR2R3-MYBs* were differentially expressed under salt, mannitol and heat stresses, indicating that they were major factors involving in crosstalk among different signal transduction pathways in response to abiotic stresses. However, several *StR2R3-MYB* genes appeared to take part in respond to only one stress stimulus, suggest that there are different signaling pathways related to the response to abiotic stress treatment. In addition, some genes showed opposite expression profiles under different stresses, implying the complicated signaling transduction pathways in response to abiotic stresses. We further analyzed StR2R3-MYB TFs might be involved in drought stress based on RNA-seq data and compared with the data treated with mannitol downloaded from PGSC, the results showed that 15 genes were highly expressed in drought-tolerant cultivar Q with an FPKM > 5 and |log2(FC)| > 1 at flower-falling stage (Fig. 8 and Additional files 12: Table S10), of which, 10 genes showed the same expression pattern in two RNA-seq datasets, while two genes (*PG0015536* and *PG0022689*) showed opposite expression pattern, two genes (*PG2026177* and *PG1026177*) in A22 and one gene (*PG0005848*) in A16 (S5) were highly expressed in Q, while showed lower expression in mannitol-treated RNA dataset. Among these 15 genes, only two genes (*PG0030548* and *PG0031317*), all belonged to S4, were highly expressed in A, while other genes were all highly expressed in Q. It's reported that AtMYB44/AtMYBR1 (S22) regulates ABA-mediated stomatal closure in response to abiotic stresses and three other members (AtMYB70, AtMYB73 and AtMYB77) in this subgroup are likely to be associated with stress responses [80, 81]. In our present work, one member *PG0024983* in S22 was up-regulated under salt and mannitol stresses, and highly expressed in Q at flowering-falling stage as well, the other member *PG0003316* in S22 was highly expressed under mannitol treatment, also up-regulated in Q, suggesting the two *StR2R3-MYB* genes are likely to be associated with drought response. AtMYB96 mediated the abscisic acid (ABA) signal network that confers abiotic stress tolerance [82], in our dataset, *PG0033043* and *PG0019535*, homologous to AtMYB96, were all up-regulated in DM potato under salt and mannitol stresses, and *PG0033043* was up-regulated in Q as well. The results revealed that these identified StR2R3-MYB TFs were responded to drought stress, and was worth further investigating. Our genome-wide analysis and

expression profiles of StR2R3-MYB TFs in response to various stresses, especially drought stress, provide a foundation for their functional characterization with stress tolerance.

The interaction network of EAR motif-containing StR2R3-MYB proteins

A transcription factor can act as an activator or repressor in the transcriptional regulation [83]. The EAR motif is the most dominant transcriptional repression motif identified in plants. These R2R3 MYB repressors include AtMYB3/4/6 in *Arabidopsis* [84], MdMYB16/17/111 in apple [85], FaMYB1 in strawberry [86] and PhMYB4 and PhMYB17 in petunia [87]. Our previous work also suggested that StMYB44-1 (PG0003316) exerted its strong repressive ability through the presence of a complete EAR motif (LxLxLx) in the C-terminal region. The sequence of this motif is less conserved in StMYB44-2, with a substitution of the final L with P (LxLxPx) which could account for its weaker repression [88]. In this work, 20 members of StR2R3-MYB family have been identified containing at least one LxLxL type of EAR motif, and one StR2R3-MYB contained DLNxxP, and a protein interaction network of EAR motif-containing StR2R3-MYB TFs with other potato proteins showed that the 21 EAR motif-containing StR2R3-MYB proteins were involved in at least nine interaction possibilities, The investigation of EAR motif-containing proteins will provide new insight into the link between complicated gene regulation system.

Conclusions

A comprehensive analysis of R2R3-MYB gene family in potato was carried out in the present study. A total of 108 *StR2R3-MYBs* were identified and phylogenetically divided into 28 distinct subfamilies, as supported by highly conserved gene structures and motifs. They were unevenly distributed among 12 chromosomes in potato. Collinearity analysis showed that the segmental duplication events played a crucial role in the expansion of *StR2R3-MYB* gene family. Synteny analysis indicated that 37 and 13 *StR2R3-MYB* genes were orthologous to *Arabidopsis* and wheat, respectively, and these gene pairs have evolved under strong purifying selection. Moreover, we performed a comprehensive expression analysis of *StR2R3-MYB* genes in different tissues and under abiotic stresses in doubled monoploid (DM) potato, as well as in white and pigmented potato cultivars and in drought-sensitive and -tolerant cultivars of tetraploid potato using RNA-seq data to identify *StR2R3-MYB* genes closely associated with spatial distribution, anthocyanin biosynthesis and multiple stresses. In addition, 21 StR2R3-MYB genes containing EAR motif were obtained and an interaction network between these EAR motif-containing R2R3-MYBs with other potato genes were constructed. Gene Ontology (GO) analysis of the 21 EAR motif-containing StR2R3-MYB proteins was also performed to further investigate their functions. The findings should inform the characterization of StR2R3-MYB superfamily and provide valuable information for further functional elucidation of these genes in potato.

Abbreviations

DM: doubled monoploid GO: GeneOntology

HMM: hidden Markov model Ka: non-synonymous substitution rate

Ks: synonymous substitution rate NJ: Neighbor–Joining

PG: PGSC0003DMG40 qRT-PCR: quantitative real-time PCR

TF: transcription factor

Declarations

Acknowledgements

Not applicable.

Funding

The Research Program Sponsored by the Special Funds for Discipline Construction of Gansu Agricultural University (GSAU-XKJS-2018-218), National Natural Science Foundation of China (31601356 and 31860398), Gansu Science Foundation for Distinguished Young Scholars (17JR5RA138), China Agriculture Potato Research System (CARS-09-P14) and Gansu Provincial Key Laboratory of Aridland Crop Science, Gansu Agricultural University (No. GSCS-2017-2).

Availability of data and materials

The raw data of the transcriptome analysis used in this study was submitted to the Sequence Read Archive (SRA) at NCBI under Project ID PRJNA528685 and PRJNA529980, and the expression data was also available Potato Genome Sequencing Consortium (PGSC, http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml).

Authors' contributions

KW, ACA, JZ and YL managed this project; YL, KW and YL analyzed the sequencing data; YL and ZL performed the experiments; YL, KW and YL wrote the manuscript; All authors read and approved the final manuscript for publication.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Antje F, Katja M, Braun EL, Erich G: Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J* 2011, 66(1):94-116.

2. Stracke R, Werber M, Weisshaar B: The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr Opin Plant Biol* 2001, 4(5):447-456.

3. Pabo CO, Sauer RT: Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem* 1992, 61:1053-1095.

4. Riechmann JL, Ratcliffe OJ: A genomic perspective on plant transcription factors. *Curr Opin Plant Biol* 2000, 3(5):423-434.

5. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec Lc: MYB transcription factors in *Arabidopsis*. *Trends Plant Sci* 2010, 15(10):573-581.

6. Rosinski JA, Atchley WR: Molecular Evolution of the Myb Family of Transcription Factors: Evidence for Polyphyletic Origin. *J Mol Evol* 1998, 46(1):74-83.

7. Du H, Wang YB, Xie Y, Liang Z, Jiang SJ, Zhang SS, Huang YB, Tang YX: Genome-wide identification and evolutionary and expression analyses of MYB-related genes in land plants. *DNA Res* 2013, 20(5):437-448.

8. Kanei-Ishii C, Sarai A, Sawazaki T, Nakagoshi H, He DN, Ogata K, Nishimura Y, Ishii S: The tryptophan cluster: a hypothetical structure of the DNA-binding domain of the myb protooncogene product. *J Biol Chem* 1990, 265(32):19990-19995.

9. Ogata K, Kanei-Ishii C, Sasaki M, Hatanaka H, Nagadoi A, Enari M, Nakamura H, Nishimura Y, Ishii S, Sarai A: The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nat Struct Biol* 1996, 3(2):178-187.

10. Ogata K, Nishimura Y: Specific DNA recognition by Myb protein. *Tanpakushitsu Kakusan Koso* 1995, 40(10):1592-1597.

11. Ogata K: Structure and dynamics of the transcription factor, Myb, in DNA-sequence recognition. *Seikagaku* 1998, 70(10):1233.

12. Martin C, Paz-Ares J: MYB transcription factors in plants. *Trends Genet* 1997, 13(2):67-73.

13. Klempnauer KH, Gonda TJ, Bishop JM: Nucleotide sequence of the retroviral leukemia gene *v-myb* and its cellular progenitor *c-myb*: the architecture of a transduced oncogene. *Cell* 1982, 31(2 Pt 1):453-463.

14. Weston K: Myb proteins in life, death and differentiation. *Curr Opin Genet Dev* 1998, 8(1):76-81.

15. Kranz H, Scholz K, Weisshaar B: *c-MYB* oncogene-like genes encoding three MYB repeats occur in all major plant lineages. *Plant J* 2000, 21(2):231-235.

16. Paz-Ares J, Ghosal D, Wienand U, Peterson P, Saedler H: The regulatory *c1* locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J* 1987, 6(12):3553.

17. Hajiebrahimi A, Owji H, Hemmati S: Genome-wide identification, functional prediction, and evolutionary analysis of the R2R3-MYB superfamily in *Brassica napus*. *Genome* 2017, 60(10):797-814.

18. Li ZJ, Peng RH, Tian YS, Han HJ, Xu J, Yao QH: Genome-wide identification and analysis of the MYB transcription factor superfamily in *Solanum lycopersicum*. *Plant Cell Physiol* 2016, 57(8):1657-1677.

19. Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM: Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Physiol* 2009, 149(2):981-993.

20. Du H, Feng BR, Yang SS, Huang YB, Tang YX: The R2R3-MYB transcription factor gene family in Maize. *PLoS ONE* 2012, 7(6):12.

21. Du H, Yang SS, Liang Z, Feng BR, Liu L, Huang YB, Tang YX: Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biol* 2012, 12:22.

22. Cao ZH, Zhang SZ, Wang RK, Zhang RF, Hao YJ: Genome wide analysis of the apple MYB transcription factor family allows the identification of *MdoMYB121* gene confering abiotic stress tolerance in plants. *PLoS ONE* 2013, 8(7):13.

23. Liu CY, Xie T, Chen CJ, Luan AP, Long JM, Li CH, Ding YQ, He YH: Genome-wide organization and expression profiling of the R2R3-MYB transcription factor family in pineapple (*Ananas comosus*). *BMC Genomics* 2017, 18:16.

24. Ambawat S, Sharma P, Yadav NR, Yadav RC: MYB transcription factor genes as regulators for plant responses: an overview. *Physiol Mol Biol Plants* 2013, 19(3):307-321.

25. Feller A, Machemer K, Braun EL, Grotewold E: Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J* 2011, 66(1):94-116.

26. Jin H, Martin C: Multifunctionality and diversity within the plant MYB-gene family. *Plant Mol Biol* 1999, 41(5):577-585.

27. Gonzalez A, Zhao M, Leavitt JM, Lloyd AM: Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J* 2008, 53(5):814-827.

28. 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):660-677.

29. Espley RV, Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC: Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *Plant J* 2007, 49(3):414-427.

30. Chagné D, Lin-Wang K, Espley RV, Volz RK, How NM, Rouse S, Brendolise C, Carlisle CM, Kumar S, De Silva N: An ancient duplication of apple MYB transcription factors is responsible for novel red fruit-flesh phenotypes. *Plant Physiol* 2013, 161(1):225-239.

31. Ballester AR, Molthoff J, De RV, Hekkert B, Orzaez D, Fernándezmoreno JP, Tripodi P, Grandillo S, Martin C, Heldens J: Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SlMYB12 leads to pink tomato fruit color. *Plant Physiol* 2010, 152(1):71-84.

32. Jung CS, Griffiths HM, De Jong DM, Cheng S, Bodis M, Kim TS, De Jong WS: The potato *developer* (*D*) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. *Theor Appl Genet* 2009, 120(1):45-57.

33. Liu Y, Kui LW, Espley RV, Wang L, Yang H, Yu B, Andrew D, Erika VG, Wang J, Zhang J: Functional diversification of the potato R2R3 MYB anthocyanin activators AN1, MYBA1, and MYB113 and their interaction with basic helix-loop-helix cofactors. *J Exp Bot* 2016, 67(8):2159-2176.

34. Chen TZ, Li WJ, Hu XH, Guo JR, Liu AM, Zhang BL: A cotton MYB transcription factor, GbMYB5, is positively involved in plant adaptive response to drought stress. *Plant Cell Physiol* 2015, 56(5):917-929.

35. Cheng L, Li X, Huang X, Ma T, Liang Y, Ma X, Peng X, Jia J, Chen S, Chen Y: Overexpression of sheepgrass R1-MYB transcription factor LcMYB1 confers salt tolerance in transgenic *Arabidopsis*. *Plant Physiol Biochem* 2013, 70(1):252-260.

36. Shin D, Moon SJ, Han S, Kim BG, Park SR, Lee SK, Yoon HJ, Lee HE, Kwon HB, Baek D *et al*: Expression of *StMYB1R-1*, a novel potato single MYB-like domain transcription factor, increases drought tolerance. *Plant Physiol* 2011, 155(1):421-432.

37. Spooner DM, Karen ML, Gavin R, Robbie W, Bryan GJ: A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proc Natl Acad Sci U S A* 2005, 102(41):14694-14699.

38. Fossen T, Andersen Ø: Anthocyanins from tubers and shoots of the purple potato, *Solanum tuberosum*. *J Hortic Sci Biotec* 2000, 75(3):360-363.

39. Christie PJ, Alfenito MR, Walbot V: Impact of low-temperature stress on general phenylpropanoid and anthocyanin pathways: enhancement of transcript abundance and anthocyanin pigmentation in maize seedlings. *Planta* 1994, 194(4):541-549.

40. André CM, Schafleitner R, Legay S, Lefèvre I, Aliaga CAA, Nomberto G, Hoffmann L, Hausman J-F, Larondelle Y, Evers D: Gene expression changes related to the production of phenolic compounds in potato tubers grown under drought stress. *Phytochemistry* 2009, 70(9):1107-1116.

41. Castellarin SD, Pfeiffer A, Sivilotti P, Degan M, Peterlunger E, Di Gaspero G: Transcriptional regulation of anthocyanin biosynthesis in ripening fruits of grapevine under seasonal water deficit. *Plant Cell Environ* 2007, 30(11):1381-1399.

42. He J, Giusti MM: Anthocyanins: natural colorants with health-promoting properties. *Annu Rev Food Sci Technol* 2010, 1:163-187.

43. Gregory PJ, Simmonds LP: Water relations and growth of potatoes. *World Crops* 1992:214-246.

44. Jefferies RA: Responses of potato genotypes to drought. I. Expansion of individual leaves and osmotic adjustment. *Ann Appl Biol* 2010, 122(1):93-104.

45. Schafleitner R, Gutierrez Rosales RO, Gaudin A, Alvarado Aliaga CA, Martinez GN, Tincopa Marca LR, Bolivar LA, Delgado FM, Simon R, Bonierbale M: Capturing candidate drought tolerance traits in two native Andean potato clones by transcription profiling of field grown plants under water stress. *Plant Physiol Biochem* 2007, 45(9):673-690.

46. Evers D, Lefevre I, Legay S, Lamoureux D, Hausman JF, Rosales ROG, Marca LRT, Hoffmann L, Bonierbale M, Schafleitner R: Identification of drought-responsive compounds in potato through a combined transcriptomic and targeted metabolite approach. *J Exp Bot* 2010, 61(9):2327-2343.

47. Moon SJ, Han SY, Kim DY, Yoon IS, Shin D, Byun MO, Kwon HB, Kim BG: Ectopic expression of a hot pepper bZIP-like transcription factor in potato enhances drought tolerance without decreasing tuber yield. *Plant Mol Biol* 2015, 89(4-5):421-431.

48. Wang L, Liu Y, Feng S, Yang J, Li D, Zhang J: Roles of Plasmalemma Aquaporin Gene *StPIP1* in Enhancing Drought Tolerance in Potato. *Frontiers in Plant Science* 2017, 8:616.

49. D'Amelia V, Aversano R, Batelli G, Caruso I, Castellano Moreno M, Castro-Sanz AB, Chiaiese P, Fasano C, Palomba F, Carputo D: High AN1 variability and interaction with basic helix-loop-helix co-factors related to anthocyanin biosynthesis in potato leaves. *Plant J* 2014, 80(3):527-540.

50. Tai HH, Goyer C, Murphy AM: Potato MYB and bHLH transcription factors associated with anthocyanin intensity and common scab resistance. *Botany* 2013, 91(10):722-730.

51. Consortium PGS: Genome sequence and analysis of the tuber crop potato. *Nature* 2011, 475(7355):189-195.

52. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A *et al*: The Pfam protein families database: towards a more sustainable future. *Nucleic*

Acids Res 2016, 44(D1):D279-D285.

53. Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q *et al*: The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol Biol* 2006, 60(1):107-124.

54. Letunic I, Doerks T, Bork P: SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res* 2015, 43(D1):D257-D260.

55. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu SN, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI *et al*: CDD: NCBI's conserved domain database. *Nucleic Acids Res* 2015, 43(D1):D222-D226.

56. Hu B, Jin JP, Guo AY, Zhang H, Luo JC, Gao G: GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 2015, 31(8):1296-1297.

57. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS: MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 2009, 37(Web Server issue):W202-208.

58. Voorrips RE: MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 2002, 93(1):77-78.

59. Wang Y, Tang H, Debarry Jeremy D, Tan X, Li J, Wang X, Lee T-H, Jin H, Marler B, Guo H: MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* 2012, 40(7):e49-e49.

60. Krzywinski M, Schein JI: Circos: an information aesthetic for comparative genomics. *Genome Res* 2009, 19(9):1639-1645.

61. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J: KaKs_Calculator 2.0:A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. *Genom Proteom bioinf* 2010, 8(1):77-80.

62. Kumar S, Stecher G, Tamura K: MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016, 33(7):1870.

63. Tang X, Zhang N, Si H, CalderÃ³n-Urrea A: Selection and validation of reference genes for RT-qPCR analysis in potato under abiotic stress. *Plant Methods* 2017, 13(1):85.

64. Chen C, Xia R, Chen H, He Y: TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. *bioRxiv* 2018:289660.

65. Kagale S, Rozwadowski K: EAR motif-mediated transcriptional repression in plants: An underlying mechanism for epigenetic regulation of gene expression. *Epigenetics* 2011, 6(2):141-146.

66. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P *et al*: The STRING database in 2017: quality-controlled protein-protein association networks, made

broadly accessible. *Nucleic Acids Res* 2017, 45(D1):D362-D368.

67. 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.

68. Williams CE, Grotewold E: Differences between plant and animal Myb domains are fundamental for DNA binding activity, and chimeric Myb domains have novel DNA binding specificities. *J Biol Chem* 1997, 272(1):563-571.

69. Lynch M, ., Conery JS: The evolutionary fate and consequences of duplicate genes. *Science* 2000, 290(5494):1151-1155.

70. 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(5):5018.

71. Pandey A, Misra P, Choudhary D, Yadav R, Goel R, Bhambhani S, Sanyal I, Trivedi R, Trivedi PK: AtMYB12 expression in tomato leads to large scale differential modulation in transcriptome and flavonoid content in leaf and fruit tissues. *Sci Rep* 2015, 5(12412):12412.

72. Stracke R, Ishihara HG, 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* 2010, 50(4):660-677.

73. Lepiniec L, Debeaujon I, J M, Baudry A, Pourcel L, Nesi N, Caboche M: Genetics and biochemistry of seed flavonoids. *Annu Rev Plant Biol* 2006, 57(1):405-430.

74. Dias AP, Braun EL, Mcmullen MD, Grotewold E: Recently duplicated maize R2R3 myb genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. *Plant Physiol* 2003, 131(2):610-620.

75. Xu HF, Yang GX, Zhang J, Wang YC, Zhang TL, Wang N, Jiang SH, Zhang ZY, Chen XS: Overexpression of a repressor MdMYB15L negatively regulates anthocyanin and cold tolerance in red-fleshed callus. *Biochem Biophys Res Commun* 2018, 500(2):405-410.

76. Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, Brendolise C, Boase MR, Ngo H, Jameson PE, Schwinn KE: A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *Plant Cell* 2014, 26(3):962-980.

77. Zhou H, Lin-Wang K, Wang F, Espley RV, Ren F, Zhao J, Ogutu C, He H, Jiang Q, Allan AC *et al*: Activator-type R2R3-MYB genes induce a repressor-type R2R3-MYB gene to balance anthocyanin and proanthocyanidin accumulation. *New Phytol* 2019, 221(4):1919-1934.

78. Takashi H, Kazuo S: Research on plant abiotic stress responses in the post-genome era: past, present and future. *Plant J* 2010, 61(6):1041-1052.

79. Ahuja I, Vos RCHD, Bones AM, Hall RD: Plant molecular stress responses face climate change. *Trends Plant Sci* 2010, 15(12):664-674.

80. Jung C, Seo JS, Han SW, Koo YJ, Kim CH, Song SI, Nahm BH, Choi YD, Cheong J-J: Overexpression of AtMYB44 enhances stomatal closure to confer abiotic stress tolerance in transgenic Arabidopsis. *Plant Physiol* 2008, 146(2):623-635.

81. Seo JS, Sohn HB, Noh K, Jung C, An JH, Donovan CM, Somers DA, Kim DI, Jeong SC, Kim CG *et al*: Expression of the Arabidopsis *AtMYB44* gene confers drought/salt-stress tolerance in transgenic soybean. *Mol Breed* 2012, 29(3):601-608.

82. Seo PJ, Xiang F, Qiao M, Park J-Y, Lee YN, Kim S-G, Lee Y-H, Park WJ, Park C-M: The MYB96 transcription factor mediates abscisic acid signaling during drought stress response in *Arabidopsis*. *Plant Physiol* 2009, 151(1):275-289.

83. Latchman DS: Transcription factors: an overview. *Int J Biochem Cell B* 1997, 29(12):1305-1312.

84. Jin H, Cominelli E, Bailey P, Parr A, Mehrtens F, Jones J, Tonelli C, Weisshaar B, Martin C: Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. *EMBO J* 2000, 19(22):6150-6161.

85. Lin-Wang K, Micheletti D, Palmer J, Volz R, Lozano L, Espley R, Hellens RP, Chagne D, Rowan DD, Troggio M *et al*: High temperature reduces apple fruit colour via modulation of the anthocyanin regulatory complex. *Plant Cell Environ* 2011, 34(7):1176-1190.

86. Aharoni A, De Vos C, Wein M, Sun Z, Greco R, Kroon A, Mol JN, O'Connell AP: The strawberry FaMYB1 transcription factor suppresses anthocyanin and flavonol accumulation in transgenic tobacco. *Plant J* 2001, 28(3):319-332.

87. Albert NW, Lewis DH, Zhang H, Schwinn KE, Jameson PE, Davies KM: Members of an R2R3-MYB transcription factor family in *Petunia* are developmentally and environmentally regulated to control complex floral and vegetative pigmentation patterning. *Plant J* 2011, 65(5):771-784.

88. Liu Y, Lin-Wang K, Espley RV, Wang L, Li Y, Liu Z, Zhou P, Zeng L, Zhang X, Zhang J *et al*: StMYB44 negatively regulates anthocyanin biosynthesis at high temperatures in tuber flesh of potato. *J Exp Bot* 2019.

Tables

Table 1 potato StR2R3-MYB proteins with conserved EAR motifs

	Homologous to At	Gene ID	Size	EAR motif sequence	Location
1	MYB15-like S2	PG0020071	259	NYLDLSLME	220
		PG0020071	259	NELMLELPE	251
2	MYB32-like S4	PG0006176	255	PDLNLELRI	188
3	MYB4-like S4	PG0013215	269	PDLNLELRI	189
4	MYB3-like S4	PG0030548	179	LLDLNSLP*	173
5	MYB17-like S9	PG0000027	310	SALQLLLDF	278
6	MYB17-like S9	PG0021654	320	SALQLLLDF	274
7	MYB101-like S18	PG0028949	485	KGLPLTLPS	200
8	MYB101-like S18	PG0013897	477	VSLSLTLAS	237
9	MYB108-like S20	PG0008761	298	QLLILELHS	89
10	MYB62-like S20	PG0014550	274	QLLILELHS	84
11	MYB108-like S20	PG0027157	325	QLLILELHS	104
12	MYB108-like S20	PG2004611	235	QLLILQLHS	68
13	MYB108-like S20	PG1004611	296	QLLILQLHS	100
14	MYB108-like S20	PG0004612	231	QLLILQLHF	68
15	MYB62-like S20	PG0005641	279	QLLILELHS	88
16	MYB70-like S22	PG0044858	305	THLSLSLPG	227
17	MYB44-like S22	PG0003316	321	TSLCLSLPG	206
18	AtMYB73-like S22	PG0024983	372	TSLSLSLPG	205-209
19	MYB41-like	PG0020012	244	FWLELYLAA	232
20	MYB15-like	PG0015087	244	DELILNLHA	77
21	MYB48-like	PG0015536	219	ERLVLELHS	72

PG represents PGSC0003DMG40.

Figures

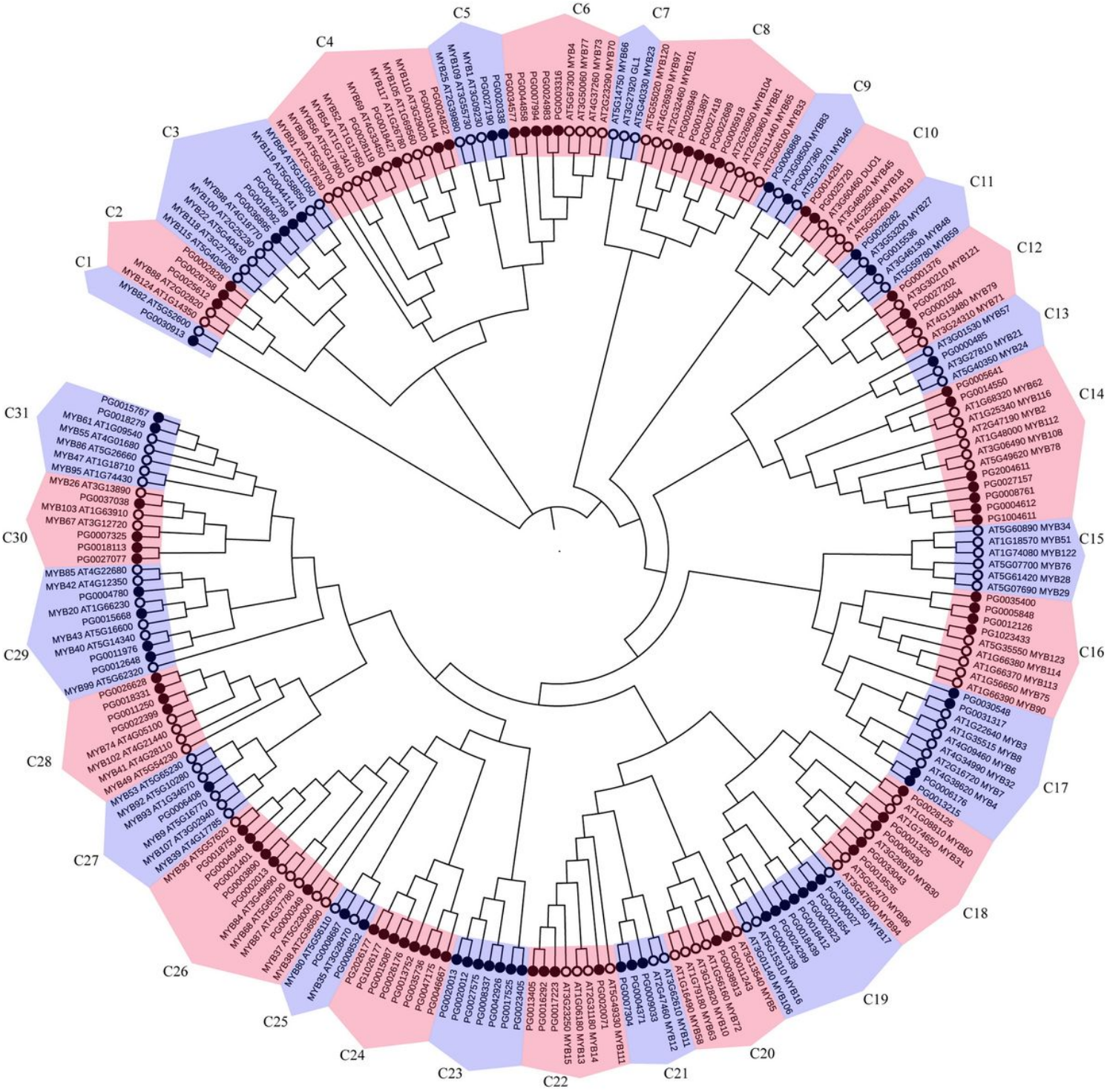

Figure 1

Phylogenetic relationships of potato and Arabidopsis R2R3-MYB proteins. The complete amino acid sequences of the 126 Arabidopsis, and 108 potato R2R3-MYB were aligned by ClustalW, and the Neighbor-Joining tree was constructed using MEGA 7 with 1000 bootstrap replicates. The hollow circles represent the Arabidopsis MYB proteins; the solid black circles represent the potato R2R3-MYB subfamily proteins. The R2R3-MYB of potato and Arabidopsis were classified into 31 clades (C1-C31), PG represents PGSC0003DMG40.

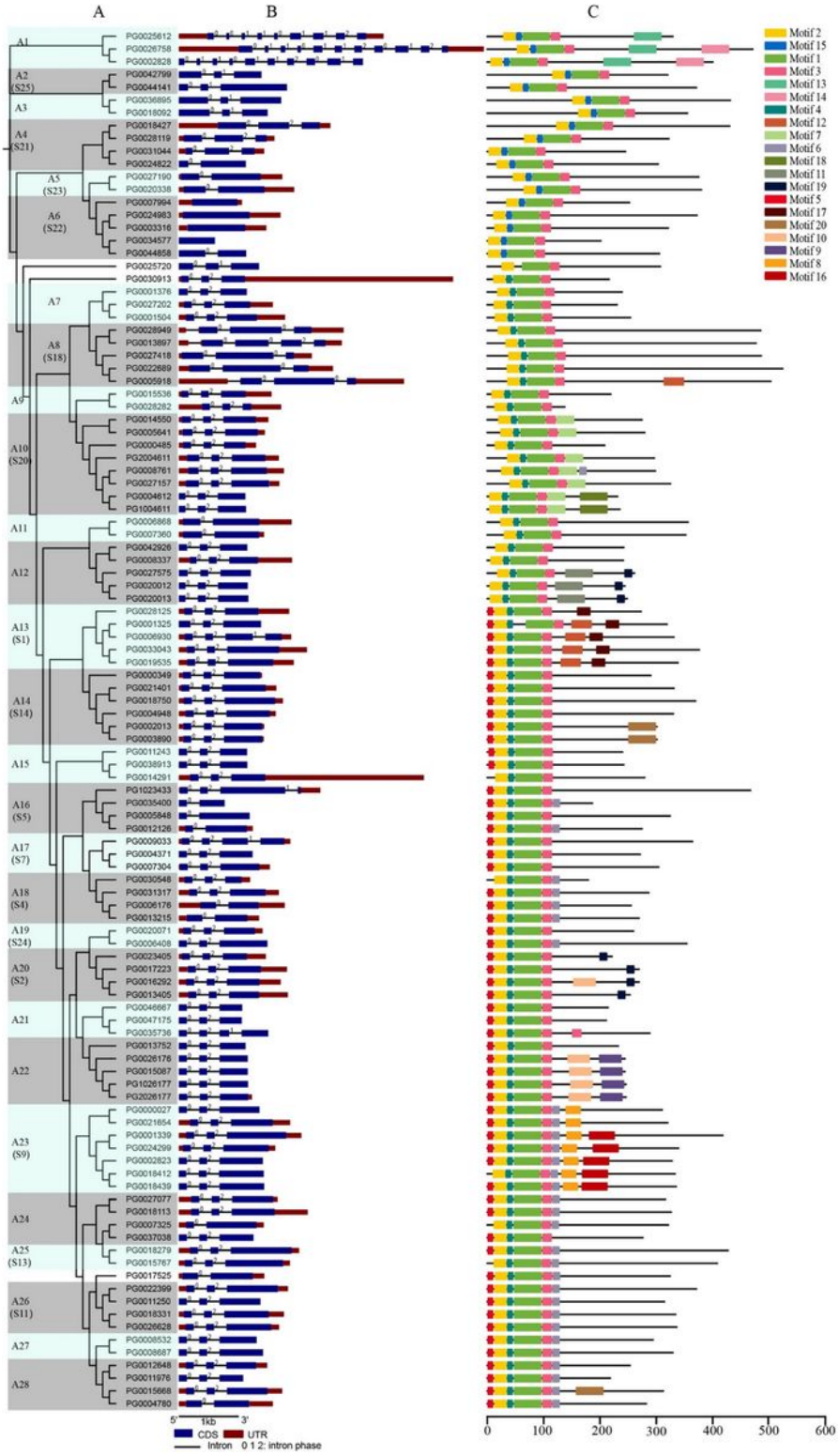

Figure 2

Phylogenetic relationships (A), gene structure (B) and motif composition (C) of StR2R3-MYB genes. A: The neighbor-joining (NJ) tree on the left includes 108 StR2R3-MYB proteins (using ClustalW for alignment, MEGA7 for constructing the neighbor-joining tree). The MYB proteins were clustered into 28 subfamilies, sequentially designated as A1 to A28. The subfamily name which was designated as previously reports of AtMYB proteins in Arabidopsis are also marked (Dubos et al., 2010). Three proteins

did not fit well into clusters. B: Exon/intron structures of R2R3-MYB genes from potato. Exon(s) and intron(s) are represented by blue boxes and black lines, respectively, red boxes represent 3’ UTR and 5’ UTR. The numbers 0, 1 and 2 represent the phases of the introns. The exon–intron structures of these genes were graphically displayed by the Gene Structure Display Server using the CDS and genome sequence of StR2R3-MYB genes. C: motif compositions of StR2R3-MYB TFs. The protein sequences of StR2R3-MYB TFs were used to predict the conserved motifs by using the MEME Suite web server.

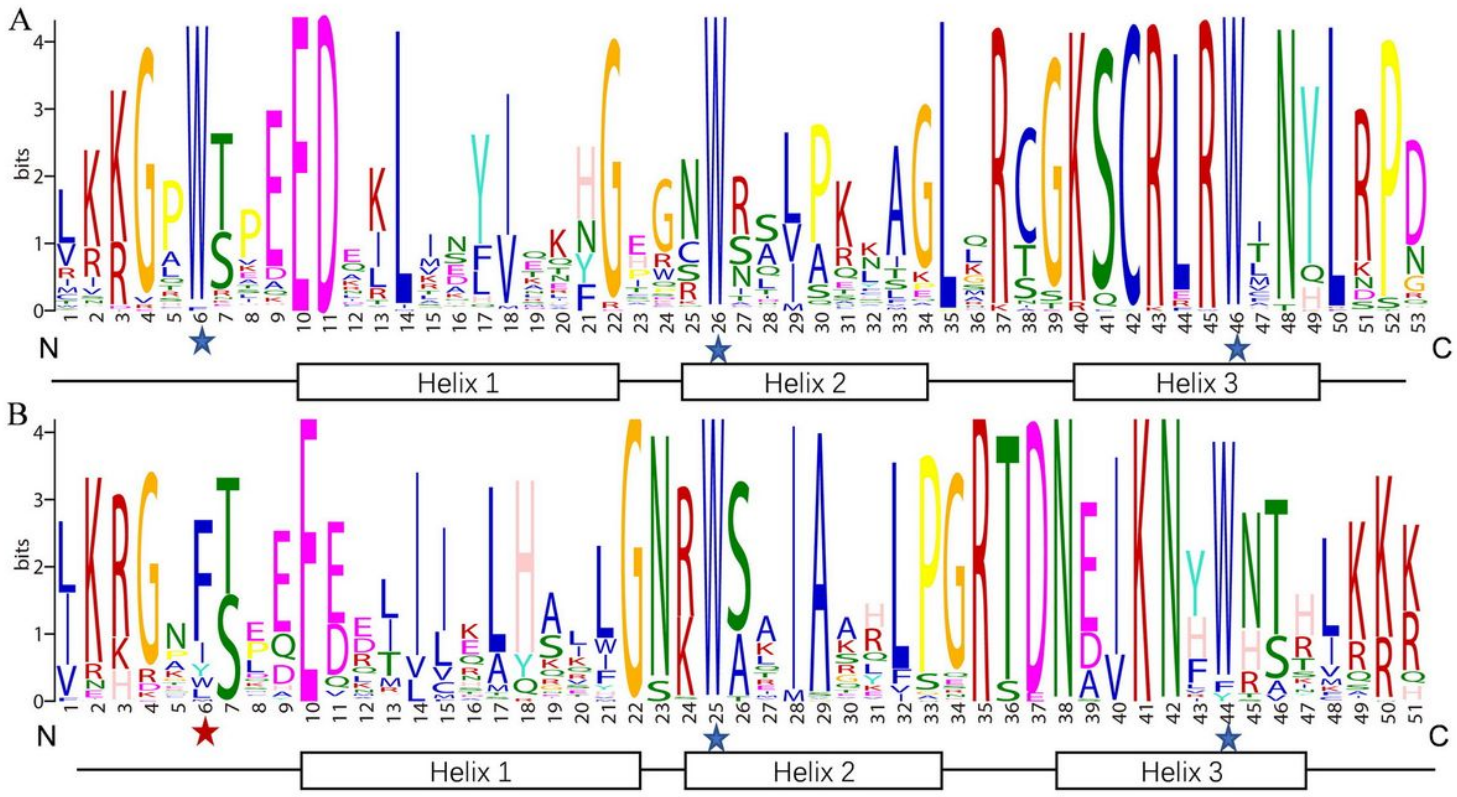

Figure 3

Sequence logos of the R2 (A) and R3 (B) MYB repeats are based on the full-length alignments of all 108 StR2R3-MYB domains. Multiple alignment analysis of the R2R3-MYB domains was performed with Clustal X. The overall height of each stack indicates the conservation of the sequence at that position and the bit score exhibits the relative frequency of the corresponding acid. The position of the three α-helices are marked (Helix 1 to 3). The conserved tryptophan residues (Trp, W) in the MYB domain are marked with blue asterisks. The replaced residues in the R3 repeat are shown by red asterisks.

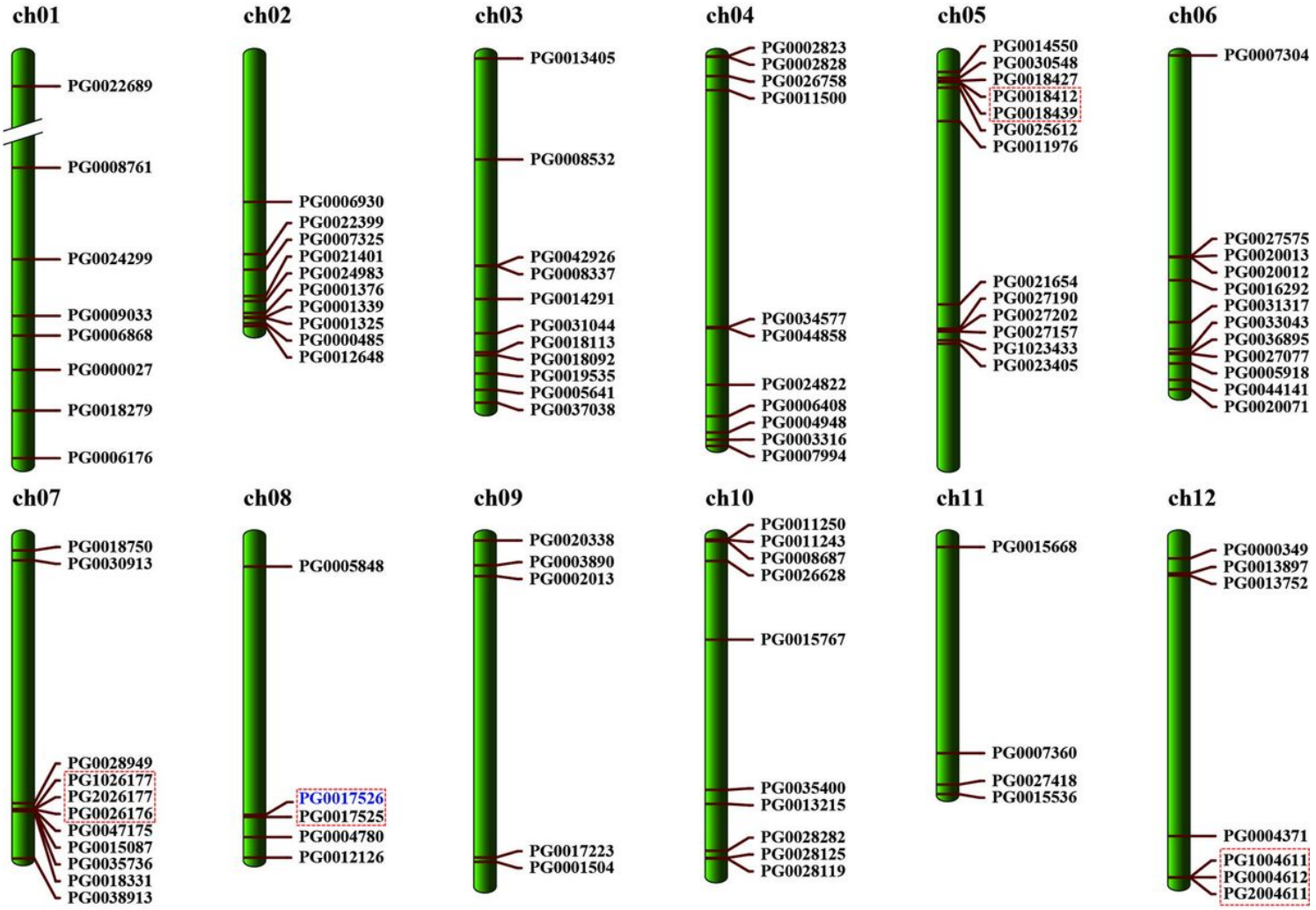

Figure 4

Chromosomal locations of potato R2R3-MYB genes. 105 out of 108 StR2R3-MYB genes are mapped to the 12 chromosomes. The chromosomal position of each StR2R3-MYB gene was mapped according to the potato genome database (PGSC). The chromosome number is indicated at the top of each chromosome. Red boxes indicate tandemly duplicated gene clusters on the chromosomes. The gene (PG0017526) marked in blue indicates a MYB-related gene.

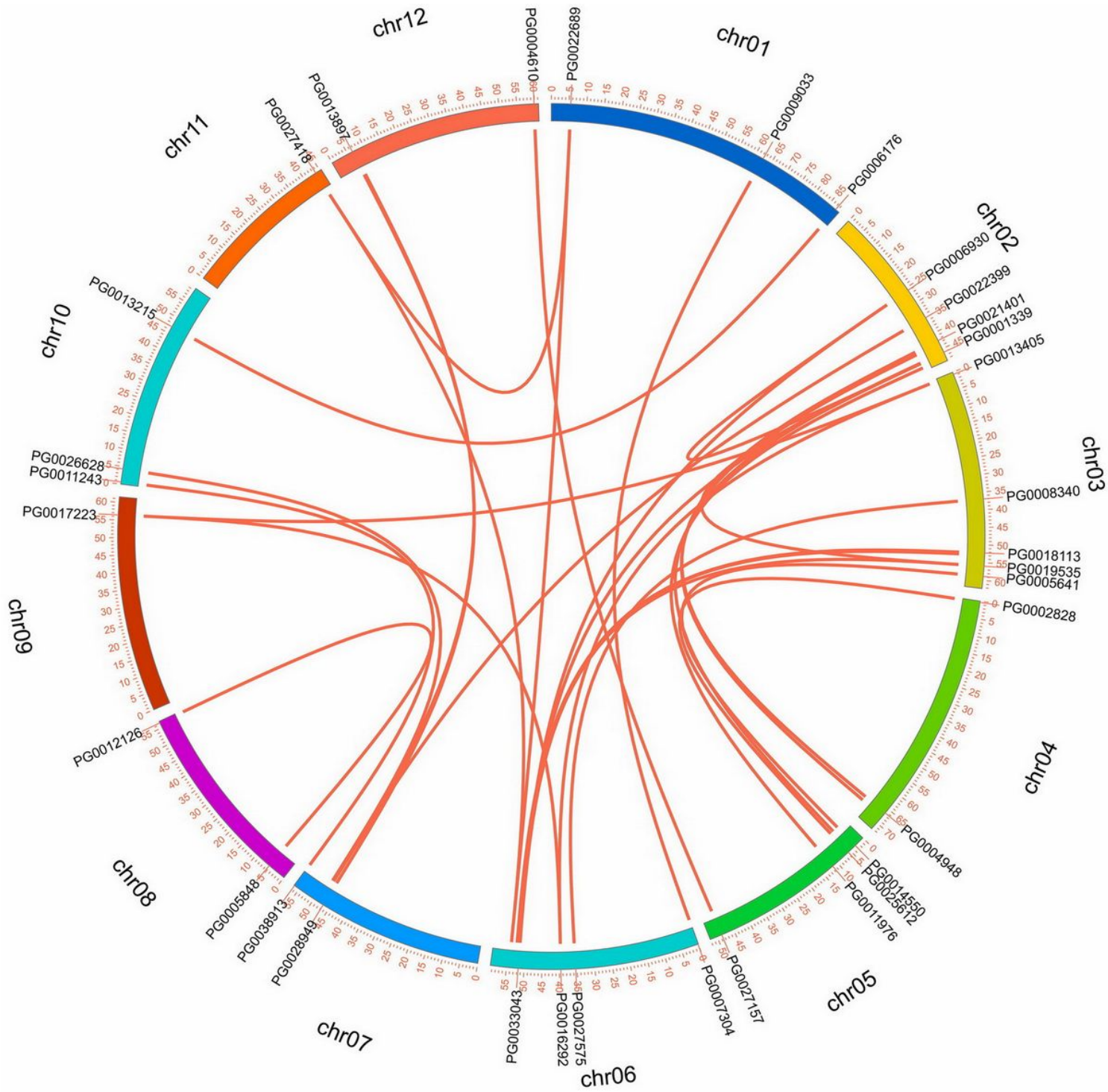

Figure 5

Genomic locations of StR2R3-MYB genes and segmentally duplicated gene pairs in the potato genome. The red lines indicate 29 segmentally duplicated R2R3-MYB gene pairs. The chromosome number is indicated at the top of each chromosome.

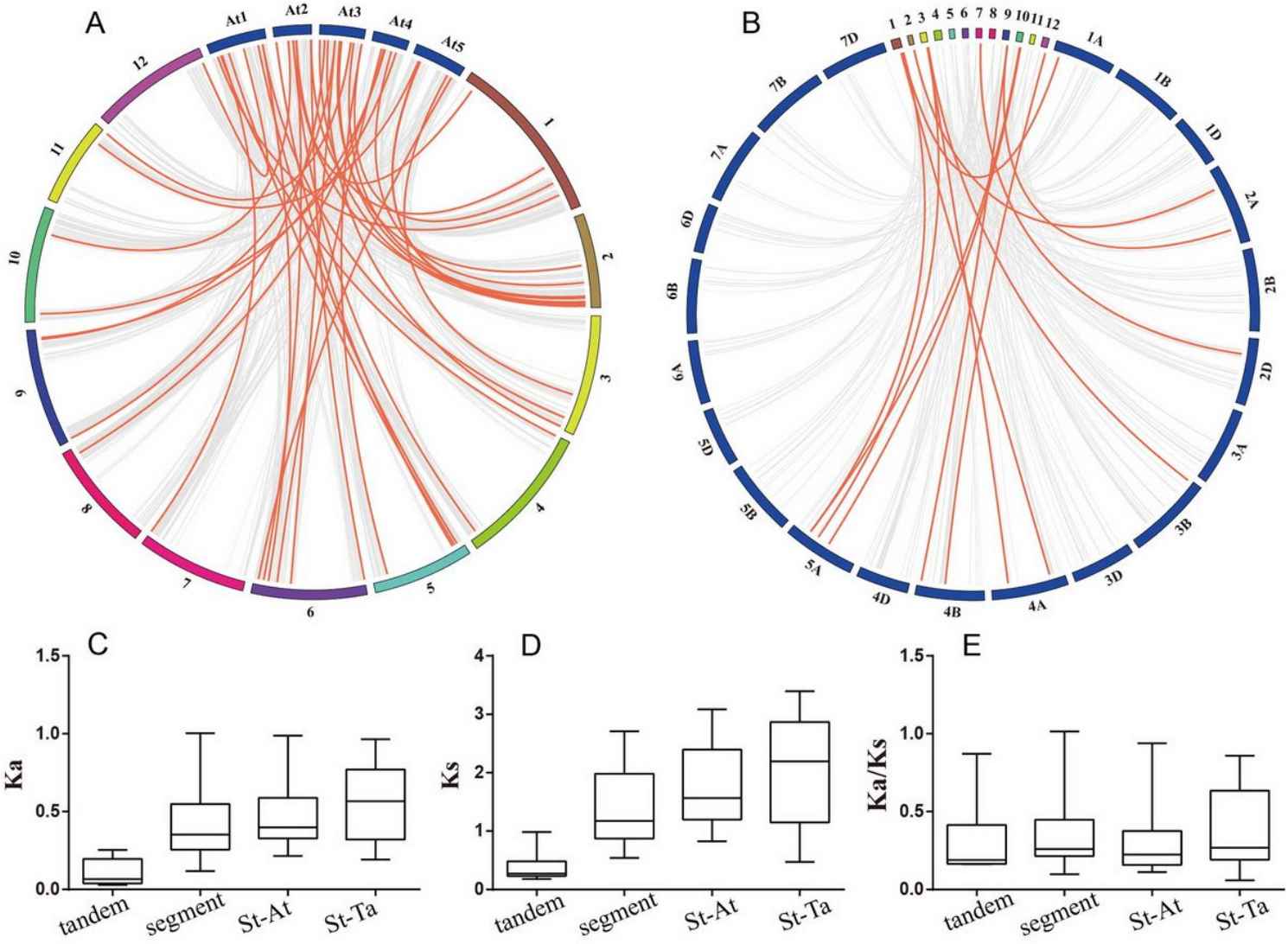

Figure 6

Comparative mapping shows the orthologous relationships of StMYB TFs with Arabidopsis (A) wheat (B). Average values of Ka, Ks and Ka/Ks, respectively of duplicated genes were shown in C-E. The horizontal axes in (C–E) represent tandem duplication (tandem), segmental duplication (segment) and the duplication between potato and Arabidopsis (St-At) and wheat (St-Ta), respectively.

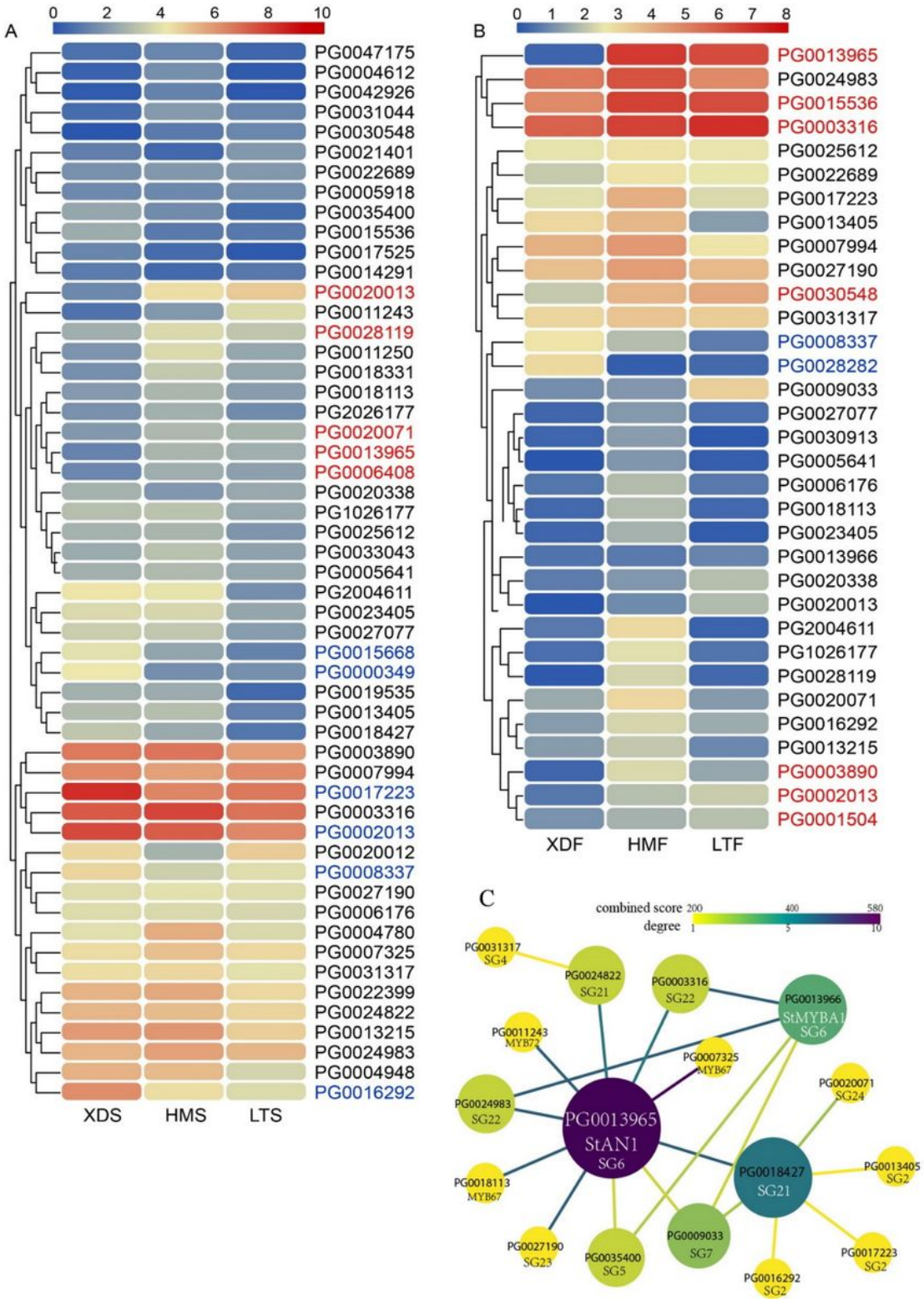

Figure 7

The expression profiles of StR2R3-MYB genes with an FPKM > 1 in white and pigmented potato tuber skin (A) and white and pigmented potato tuber flesh (B) and the interaction network of StR2R3-MYB proteins in skin and flesh (C). XDS, HMS and LTS represent white skin of white potato cultivar (XD), purple skin of purple potato cultivar (HM) and red skin of red potato cultivar (LT); XDF, HMF and LTF represent white flesh of white potato cultivar (XD), purple flesh of purple potato cultivar (HM) and red flesh of red

potato cultivar (LT). The gene IDs marked in red represent up-regulated R2R3-MYB genes in pigmented tissues, and the gene IDs marked in blue represent up-regulated R2R3-MYB genes in white tissue. The edge color was shown from yellow to purple in accordance with the combined score (confidence) > 200, the node size was in accordance with the node degree from yellow to purple to highlight the most-connected genes.

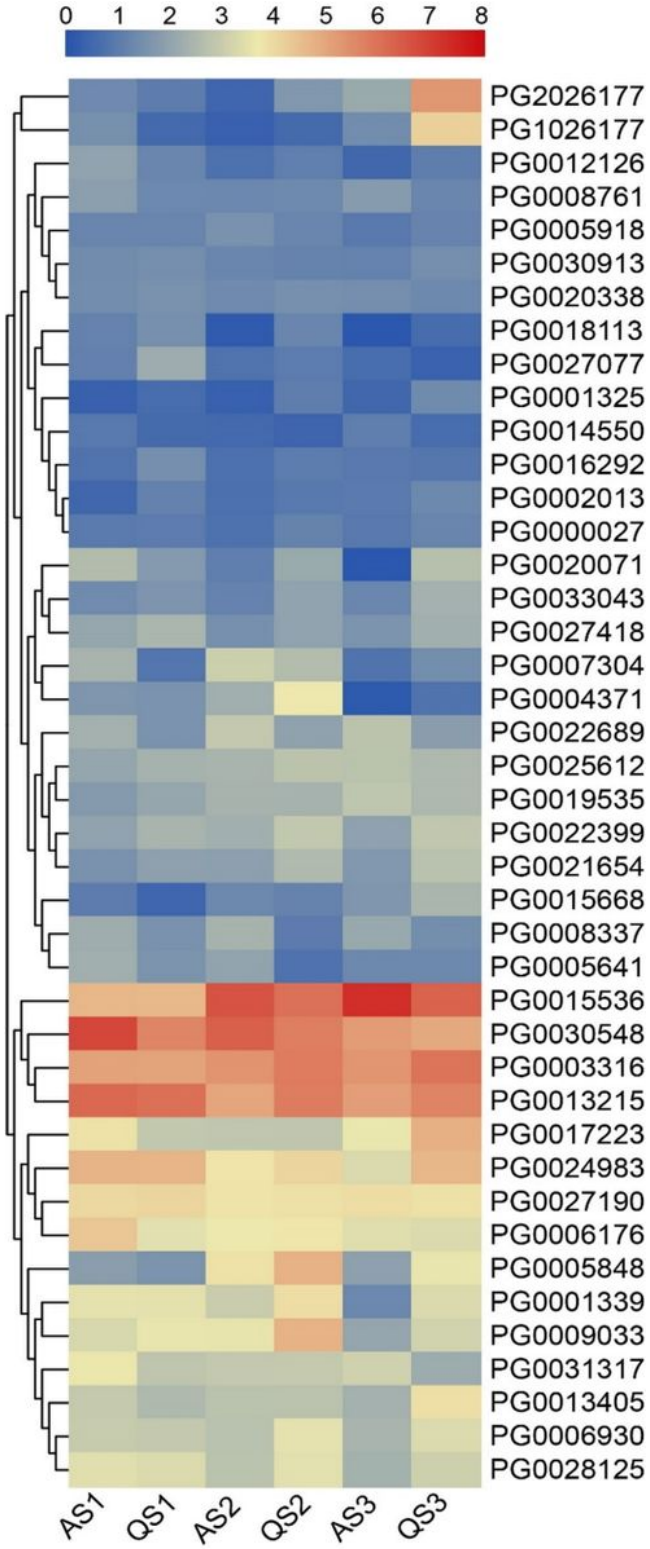

Figure 8

The expression profiles of StR2R3-MYB genes in drought-sensitive cultivar (Atlantic, A) and drought-tolerant cultivar (Qingshu No.9, Q) with an FPKM > 1 under drought stress at early flowering stage (S1), full-blooming stage (S2) and flower-falling stage (S3).

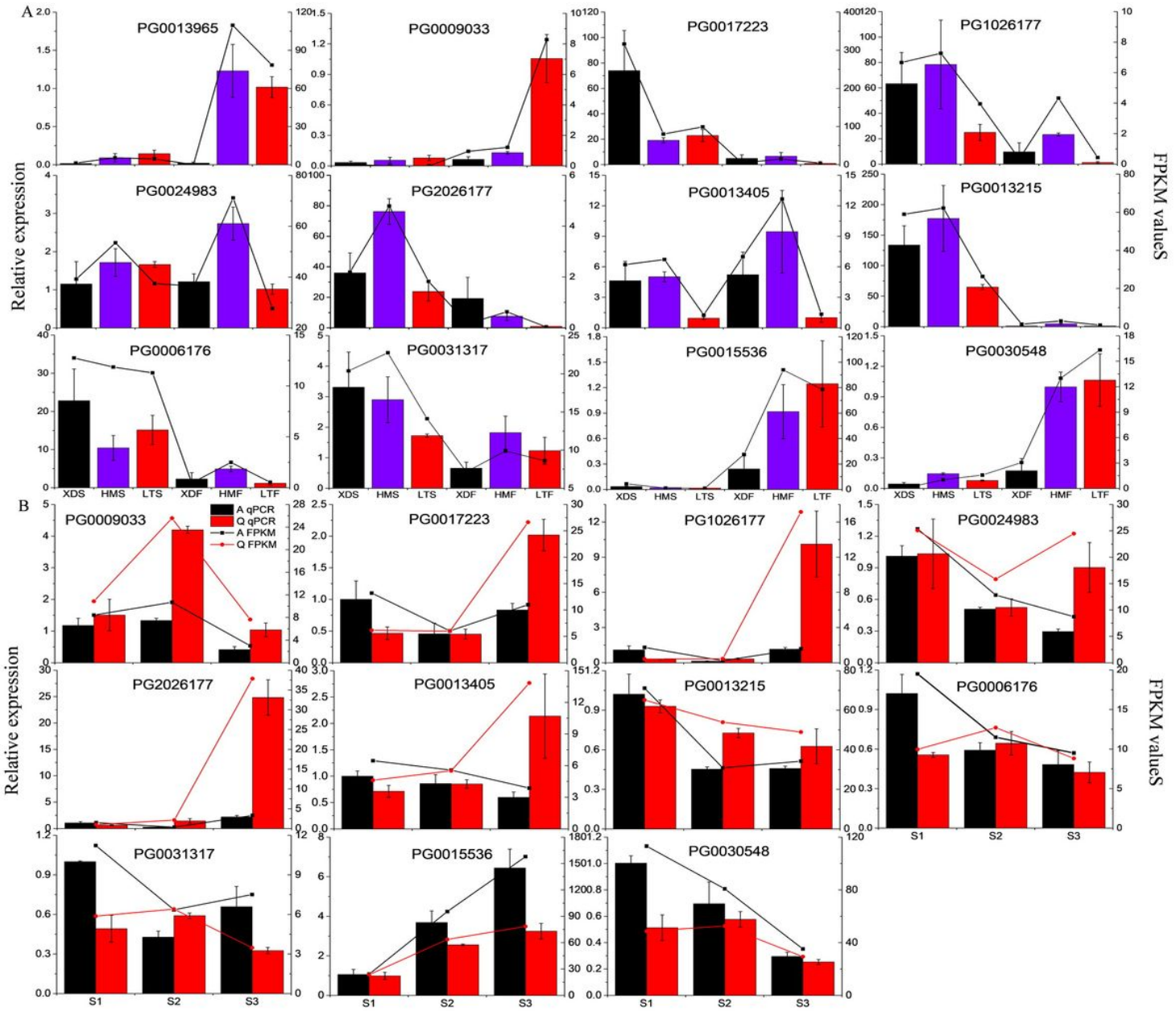

Figure 9

Quantitative RT-PCR expression analyses of 12 StR2R3-MYB genes in white and pigmented skin and flesh (A) and in drought sensitive cultivar (Atlantic, A) and drought-tolerant cultivar (Qingshu NO.9, Q) cultivar under drought stress at early flowering stage (S1), full-blooming stage (S2) and flower-falling stage (S3) (B).

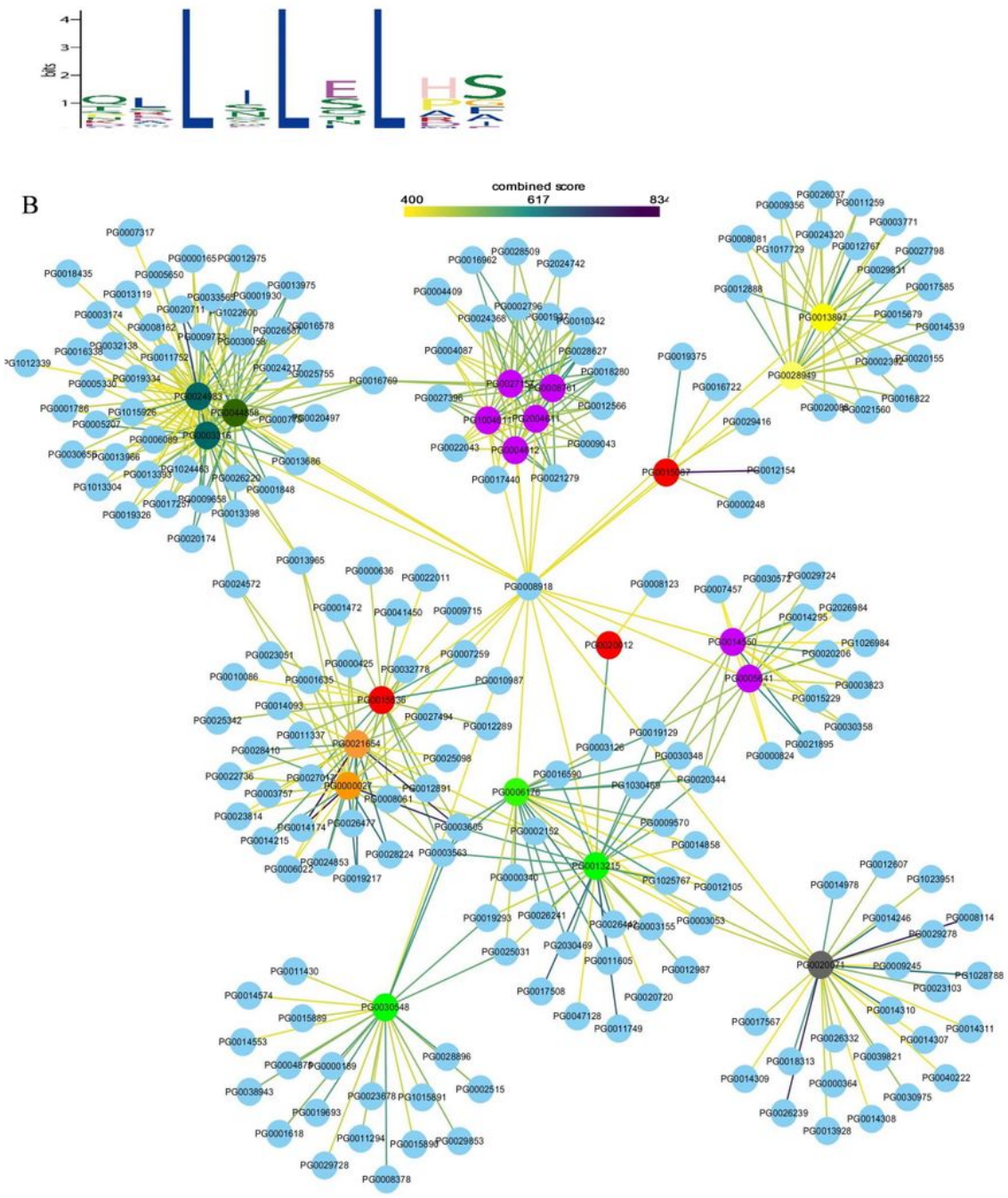

Figure 10

Protein Sequence logo of 21 EAR (LxLxL) Motifs and protein interaction network predicted for EAR motif-containing StR2R3-MYB TFs with interacting potato proteins (combined score > 400, medium confidence). The edge color was shown from yellow to purple in accordance with the combined score. PG0020071 in S2 was highlighted in dark grey; PG0006176, PG0013215 and PG0030548 in S4 were highlighted in green, PG0000027 and PG0021654 in S9 were highlighted in orange; PG0028949 and

PG0013897 in S18 were highlighted in yellow; PG0008761, PG0014550, PG0027157, PG2004611, PG1004611, PG0004612 and PG0005641 in S20 were highlighted in violet; PG0044858, PG0003316 and PG0024983 in S22 were highlighted in dark green; PG0020012, PG0015087 and PG0015536 were highlighted in red.

Supplementary Files

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

- TableS9.xlsx
- TableS2.xlsx
- TableS1.xlsx
- FigureS1.jpg
- TableS4.xlsx
- TableS5.xlsx
- TableS8.xlsx
- FigureS2.jpg
- TableS6.xlsx
- TableS12.xlsx
- TableS11.xlsx
- TableS10.xlsx
- TableS7.xlsx
- TableS3.xlsx