



Identification and Analysis of R2R3-MYB Genes in Sweet Potato Genome

Huang Xiaofang^{1,2*}, Bi Chuyun^{1,2*}, Huang Bifang³, Xu Ming^{1,2}, Yang Zhijian^{1,2}, Lin Shiqiang^{1,3*}✉, Chen Xuanyang^{1,2,4}✉

1 Key Laboratory of Crop Biotechnology, Fujian Agriculture and Forestry University, Fujian Province Universities, Fuzhou, 350002, China

2 College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou, 350002, China

3 College of Life Science, Fujian Agriculture and Forestry University, Fuzhou, 350002, China

4 Key Lab of Genetics, Breeding and Multiple Application of Crops, Ministry of Education, Fujian Agriculture and Forestry University, Fuzhou, 350002, China

* These authors contributed equally to this work

✉ Corresponding author email: linshiqiang@fafu.edu.cn; cxy@fafu.edu.cn

Molecular Plant Breeding, 2020, Vol.11, No.22 doi: [10.5376/mpb.2020.11.0022](https://doi.org/10.5376/mpb.2020.11.0022)

Received: 01 Sep., 2020

Accepted: 04 Sep., 2020

Published: 18 Sep., 2020

Copyright © 2020 Huang et al., This article was first published in Molecular Plant Breeding in Chinese, and here was authorized to translate and publish the paper in English under the terms of Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Huang X.F., Bi C.Y., Huang B.F., Xu M., Yang Z.J., Lin S.Q., and Chen X.Y., 2020, Identification and analysis of R2R3-MYB genes in sweet potato genome, Molecular Plant Breeding, 11(22): 1-12 (doi: [10.5376/mpb.2020.11.0022](https://doi.org/10.5376/mpb.2020.11.0022))

Abstract The plant MYB is a transcription factor family large in number and with important functions. In this study, the MYB family genes were screened and identified via bioinformatic methods from the raw sequence of sweet potato (*Ipomoea batatas*) genome and the gene structure and function of R2R3-MYB were analyzed. The results showed that there were 88 R2R3-MYB genes with intact R2 and R3 conservative domains, which contained 8 and 9 highly conserved basic amino acids. The results of MEME analysis showed that there were 10 conserved motifs within the *I. batatas* R2R3-MYB protein sequences. For the *I. batatas* R2R3-MYB protein sequences, over 80% contained motif 1, motif 2, motif 3, motif 4, motif 5 and motif 7. The R2R3-MYB genes were distributed unevenly across the 15 chromosomes. The number of R2R3-MYB genes in chromosome No.5 was 15, which was the largest; the numbers of R2R3-MYB genes in chromosome No. 4 and 13 were both only 2, which was the smallest. Analysis of the sequence alignment showed that there were 6 pairs of interchromosomal duplication and there were 20 pairs of intrachromosomal duplication, 19 of which existed in clusters. The function prediction and categorization via sequence analysis showed that 44 R2R3-MYB genes of the *I. batatas* could be categorized to the 13 subgroups of the *A. thaliana* R2R3-MYB genes, which were involved in the responses to biotic stress and abiotic stress, anthocyanin biosynthesis, anther development, etc. Further analysis showed that 36 R2R3-MYB genes might play important roles in dealing with biotic stress and abiotic stress, 9 of which showed significant up/down-regulation under *Fusarium oxysporum* f. sp. *batatas* stress and 27 of which showed significant up/down-regulation under low temperature stress. The domains of *I. batatas* R2R3-MYB transcription factors were highly conservative, which contained highly conserved motifs within R2 and R3 domains. The phylogenetic tree and transcriptomics data analysis showed that some R2R3-MYB genes might play roles in growth and development, metabolism regulation, biotic stress and abiotic stress, which lent support to *I. batatas* breeding.

Keywords *Ipomoea batatas*; R2R3-MYB; Transcription factor; Bioinformatics

MYB transcription factor is an important transcription factor that covers a large number of genes that play important roles in plant growth and development. MYB proteins contain highly conserved DNA binding domains, the repetitive numbers of which are taken for categorization into four types, i.e., 1R-MYB, 2R-MYB (R2R3), 3R-MYB (R1R2R3) and 4R-MYB. In plant, the main MYB transcription factor is R2R3 type (Millard et al., 2019), participating in various kinds physiological activities such as the regulations of primary and secondary metabolisms, the biosynthesis of anthocyanin (Zhu et al., 2016), cell differentiation regulation and leaf morphological genesis, cell cycle control and biotic and abiotic stresses (Cominelli and Tonelli, 2009).

Research on the plant MYB transcription factors has been reported in recent years. *Fagopyrum tataricum* (Tartary buckwheat) *FtMYB21* took part in the response to abiotic stresses of drought and salinity (Huang et al., 2015). *Astragalus membranaceus* (Astragalus) *AmMYB44* might be involved in the regulation during abiotic stresses such as drought and low temperature (Li et al., 2019). Dubos et al. (2010) classified the R2R3-MYB gene family of *Arabidopsis thaliana* (Arabidopsis) into 23 subgroups and the genes were functional annotated.

Sweet potato is an important food crop and serves as a novel energy crop as well. Studies have been conducted at different levels including breeding, cell culture, gene cloning and functional analysis, et al. However, little has been reported on the number, structure and functional analysis of MYB transcription factors in the sweet potato genome. Based on the genome data of sweet potato cultivar Taizhong 6, which was presented jointly by Shanghai Chenshan Plant Science Research Center, Institute of Plant Physiology and Ecology, Germany Max Planck Institute for Molecular Genetics and Max Planck Institute of Molecular Plant Physiology, we utilized bioinformatic methods to predict the sequences of R2R3-MYB and performed analysis on the gene structure, conservative domain, chromosomal localization and phylogenetic tree. Further, the differential expressions of R2R3-MYB genes under *Fusarium oxysporum* f. sp. *batatas* (Fob) stress and under low temperature stress were analyzed. This study lent support to the investigation of response mechanism of sweet potato to biotic and abiotic stresses and provided scientific reference for resistance breeding.

1 Results and Analysis

1.1 Identification of R2R3-MYB genes in *I. batatas* genome

The snap program is applied to predict the CDS within the *I. batatas* genome sequence and 178 458 CDS sequences were obtained (Huang et al., 2020) (Figure 1). With the MYB_DNA-binding HMM model, the hmmsearch program is used to search the whole protein sequences of *I. batatas* genome and 270 sequences with MYB domains were picked up. The sequences with high specificity ($E\text{-value} \leq 1e\text{-30}$) were selected to rebuild the HMM model specific to *I. batatas* for searching the whole protein sequences of *I. batatas* genome. The found 181 sequences were then verified with Interproscan. The 56 R2R3-MYB protein sequences after Interproscan verification were used to build the database of the R2R3-MYB transcription factors of the *I. batatas*. The database was then aligned to the whole protein sequences of *I. batatas* genome and 186 sequences ($E\text{-value} \leq 1e\text{-5}$) were returned. The sequences were merged and sorted and 186 sequences containing MYB domains were obtained. According to the number of MYB domains within the sequence, 44 2R-MYB transcription factors with intact R2R3 domains were screened. For the remaining 146 sequences of 1R MYB or with incomplete R2/R3 domains, the location information within the chromosome corresponding to the ZFF file was used to define the interval within the chromosome for a new search of the CDS. The consecutive sequences with an overlapped interval were considered one single gene and thus combined, and 44 2R-MYB transcription factors with intact R2R3 domains were supplemented. After sequence classification and count, the sequences without MYB domain or with severely incomplete MYB domain were excluded and 88 R2R3-MYB transcription factors were obtained eventually.

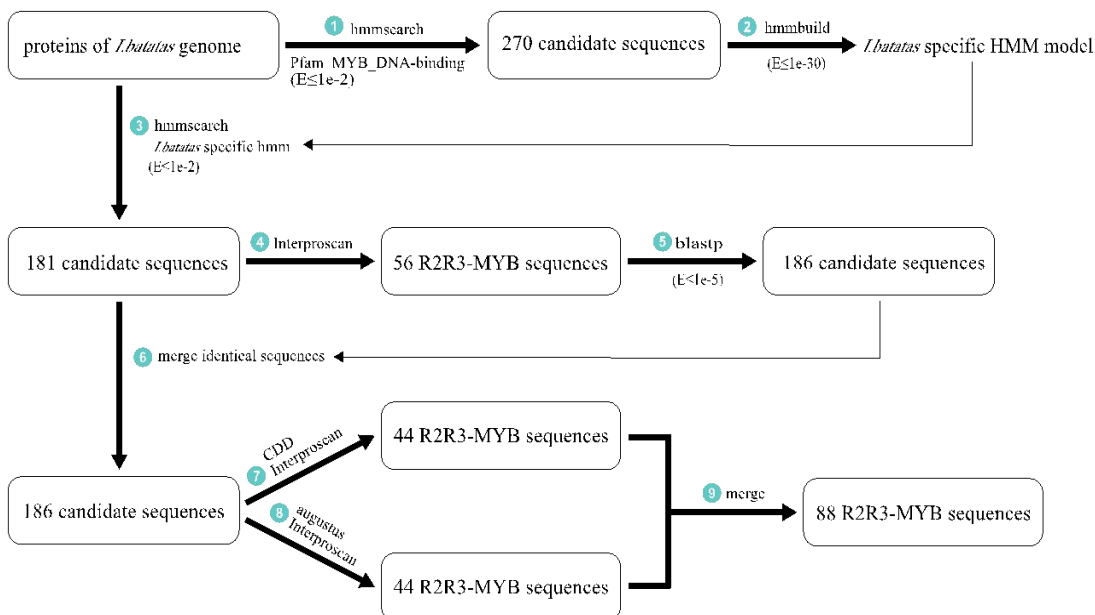


Figure 1 Screening the R2R3-MYB genes of *I. batatas*

1.2 Conservative domain analysis of R2R3-MYB family

Multiple sequence alignment was carried out for the 88 R2R3-MYB transcription factor sequences, with reference to the amino acid sequence information of the *A. thaliana* R2 and R3 domains for defining the *I. batatas* R2 and R3 conservative domains (Figure 2). The amino acid sequence of the most likely *I. batatas* R2 domain was KGPWTPEEDEKLINYI+KHGEGNWRALPKKAGLQRCGKSCRLRW+NYLRPDI (the underlines indicated the highly conserved amino acids), with amino acid sequence length 52 and number of highly conserved tryptophan residues 3. The amino acid sequence of the most likely *I. batatas* R3 domain is KRGNFSPEEEELIIRLHALLGNRWSAIAARLPGRTDNEIKNYWNTHLKKKI, with amino acid sequence length 51 and number of highly conserved tryptophan residues 2. In addition, we found that there were highly conserved basic amino acids (arginine R, lysine K and histidine H) in the R2 and R3 domains, with 8 in R2 domain and 9 in R3 domain (Figure 2).

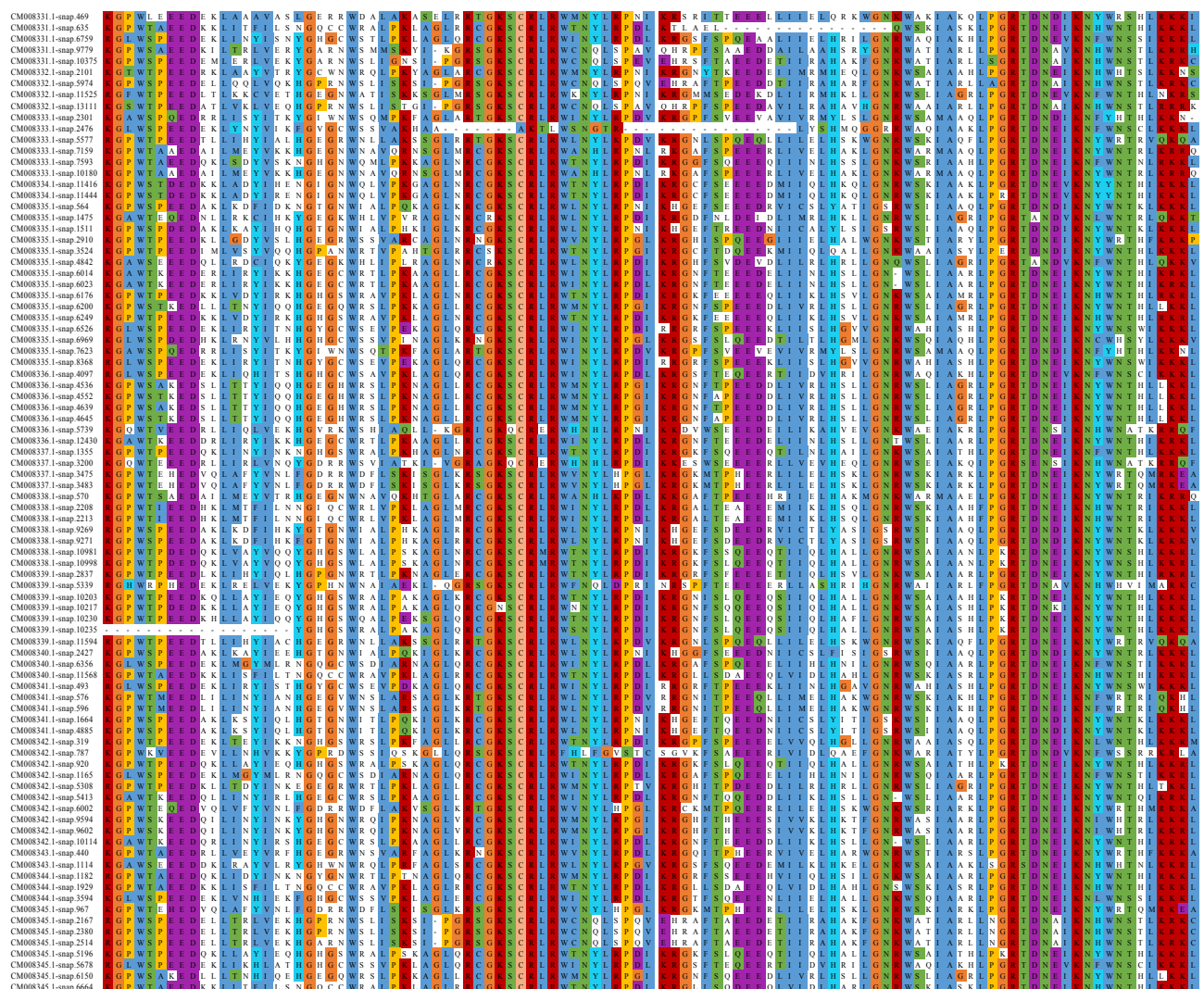


Figure 2 Conservation analysis of *I. batatas* R2R3-MYB transcription factors

1.3 Domain and motif analysis of the *I. batatas* MYB genes

The conservative motifs of the *I. batatas* R2R3-MYB protein sequences were analyzed with MEME program. Ten conservative motifs were obtained and labelled corresponding to the locations in the phylogenetic tree (Figure 3). The results showed that among the domains of the 88 R2R3-MYB transcription factors, the pattern of the motif arrangement was motif 6-motif 3-motif 7-motif 1-motif 5-motif 4-motif 2-motif 8-motif 9-motif 10. Of these motifs, the motif 10 was the least in number and existed in only 6 sequences that belonged to

the same branch of the phylogenetic tree. Also, the motif 9 was found only in 9 sequences that fell into the same branch of the phylogenetic tree.

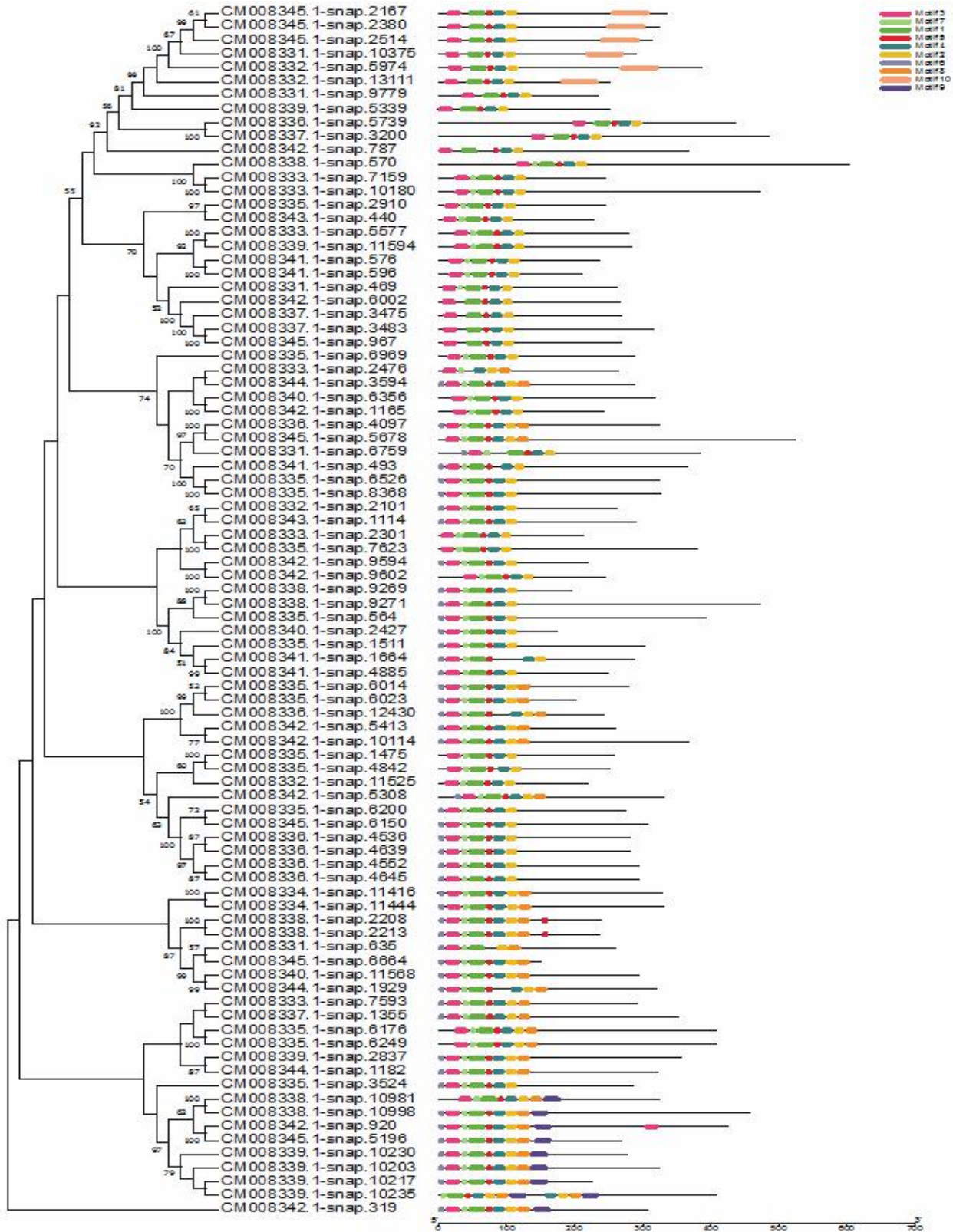









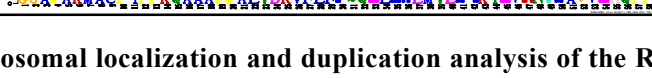


Figure 3 Motif distribution of *I. batatas* R2R3-MYB transcription factors

The amino acid sequences of the 10 motifs were aligned to the amino acid sequences of the R2 and R3 domains (Table 1). The results showed that motif 1 and motif 3 belonged to R2 domain and motif 2, motif 4, motif 5 and motif 7 resided in R3 domain. All of the 6 motifs were distributed among over 80% MYB protein sequences, suggesting a high domain conservation for the identified *I. batatas* R2R3-MYB transcription factors. Besides, the motif 9 existed in only 9 sequences that were in the same branch of the phylogenetic tree, similar to the situation of motif 10, which were only found in 6 sequences that stayed in the same branch of the phylogenetic tree, thus the motif information could serve as the reference for classifying the subgroups of R2R3-MYB transcription factors to a certain extent.

Table 1 The conserved motifs of *I. batatas* R2R3-MYB transcription factors and their corresponding domains

Name	Motif	E-value	Width	Predicted domain
motif 1		8.7e-1792	26	R2
motif 2		1.1e-1073	17	R3
motif 3		2.4e-1101	22	R2
motif 4		3.7e-1008	19	R3
motif 5		3.1e-285	10	R3
motif 6		1.8e-270	10	-
motif 7		3.2e-250	10	R3
motif 8		1.6e-195	19	-
motif 9		9.7e-125	26	-
motif 10		8.1e-100	58	-

1.4 Chromosomal localization and duplication analysis of the R2R3-MYB transcription factors

The location of the 88 *I. batatas* R2R3-MYB sequences within the chromosomes were shown in Figure 4. It could be seen that the R2R3-MYB transcription factors were distributed among the 15 chromosomes in an uneven style. The No. 5 chromosome possessed the largest number of R2R3-MYB transcription factors, which was 15. The number of R2R3-MYB transcription factors in No. 12 chromosome was 10, while those in No. 4 and No. 13 chromosomes were the least, which were both 2.

Generally, gene duplication within the genome occurred during the evolution process of plants. The judgement of whether or not two genes had a duplication relationship relied on two conditions (Vatansever et al., 2016), the coverage of the aligned area exceeded 75% of the sequence length of the longer gene, and the similarity between the two aligned sequences should be no less than 75%. We conducted pair-wise alignment for the DNA sequences of 88 R2R3-MYB genes. The pairs with over 75% similarity were verified with needle program (Rice et al., 2000). The pairs that satisfied the two conditions were regarded as showing potential duplication relationships, which were linked with lines in Figure 4 (Blue lines represent inter-chromosomal duplication relationships and orange lines represent intra-chromosomal duplication relationships).

hips). The results showed that there were both intra-chromosomal duplication relationships and inter-chromosomal duplication relationships among the *I. batatas* R2R3-MYB genes. The number of gene pairs with potential inter-chromosomal duplication relationship was 6 and that with potential intra-chromosomal duplication relationship was 20. Furthermore, the gene pairs of the potential intra-chromosomal duplication relationship were arranged in clusters, such as *CM008334.1-snap.11416* and *CM008334.1-snap.11444* (in No. 4 chromosome), *CM008338.1-snap.2208* and *CM008338.1-snap.2213* (in No. 8 chromosome), and *CM008342.1-snap.9594* and *CM008342.1-snap.9602* (No. 12 chromosome). Each two of the four genes that were clustered in No. 8 chromosome, i.e., *CM008336.1-snap.4536*, *CM008336.1-snap.4552*, *CM008336.1-snap.4639* and *CM008336.1-snap.4645*, showed potential duplication relationships.

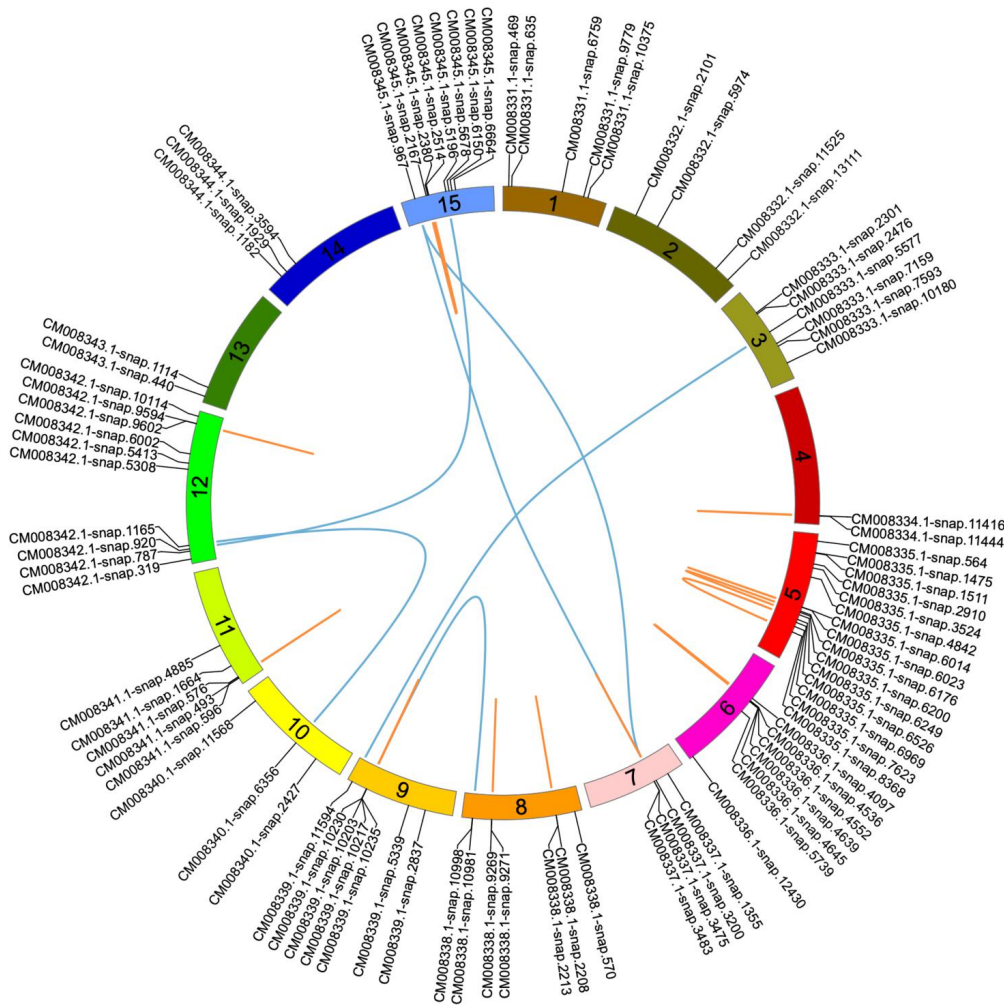


Figure 4 Chromosomal localization and duplication relationship of the R2R3-MYB genes in *I. batatas*

Note: Blue lines: Inter-chromosomal duplication relationship; Orange lines: Intra-chromosomal duplication relationship

1.5 Phylogenetic analysis of the R2R3-MYB gene families of *I. batatas* and *A. thaliana*

The R2R3-MYB protein sequences of the *I. batatas* were used to build the phylogenetic tree together with those of the *A. thaliana* (Figure 5) and the categorization and functional prediction of the *I. batatas* R2R3-MYB sequences were performed in the light of functionally annotated subgroups of *A. thaliana* R2R3-MYB genes. It has been demonstrated that the *A. thaliana* R2R3-MYB family could be classified into 23 subgroups (Dubos et al., 2010). Each subgroup of the *A. thaliana* R2R3-MYB family was marked with one color and it was found that 44 of the *I. batatas* R2R3-MYB transcription factor family were listed to the 13 subgroups of the *A. thaliana* R2R3-MYB family including S1 (response to biotic and abiotic stresses), S4 (control of

metabolic pathway), S6 (control of anthocyanin biosynthesis), S9 (control of early inflorescence development and seed germination), S13 (influencing lignin deposition, mucilage production and stomatal aperture), S14 (partially redundant regulators of axillary meristem formation and root development), S18 (involved in anther/pollen development), S19 (control of anther development), S20 (related to stress responses), S21 (control of axillary meristem formation), S22 (modulating auxin, implicated in stress responses), S24 (lateral root development and fatty acid synthesis) (Gibbs et al., 2014; To et al., 2020) and S25 (play roles in embryogenesis). Moreover, in accordance with the motifs shown in Fig. 3, the S9 subgroup and S22 subgroup both had a unique motif, which were motif 9 for S9 subgroup and motif 10 for subgroup S22.

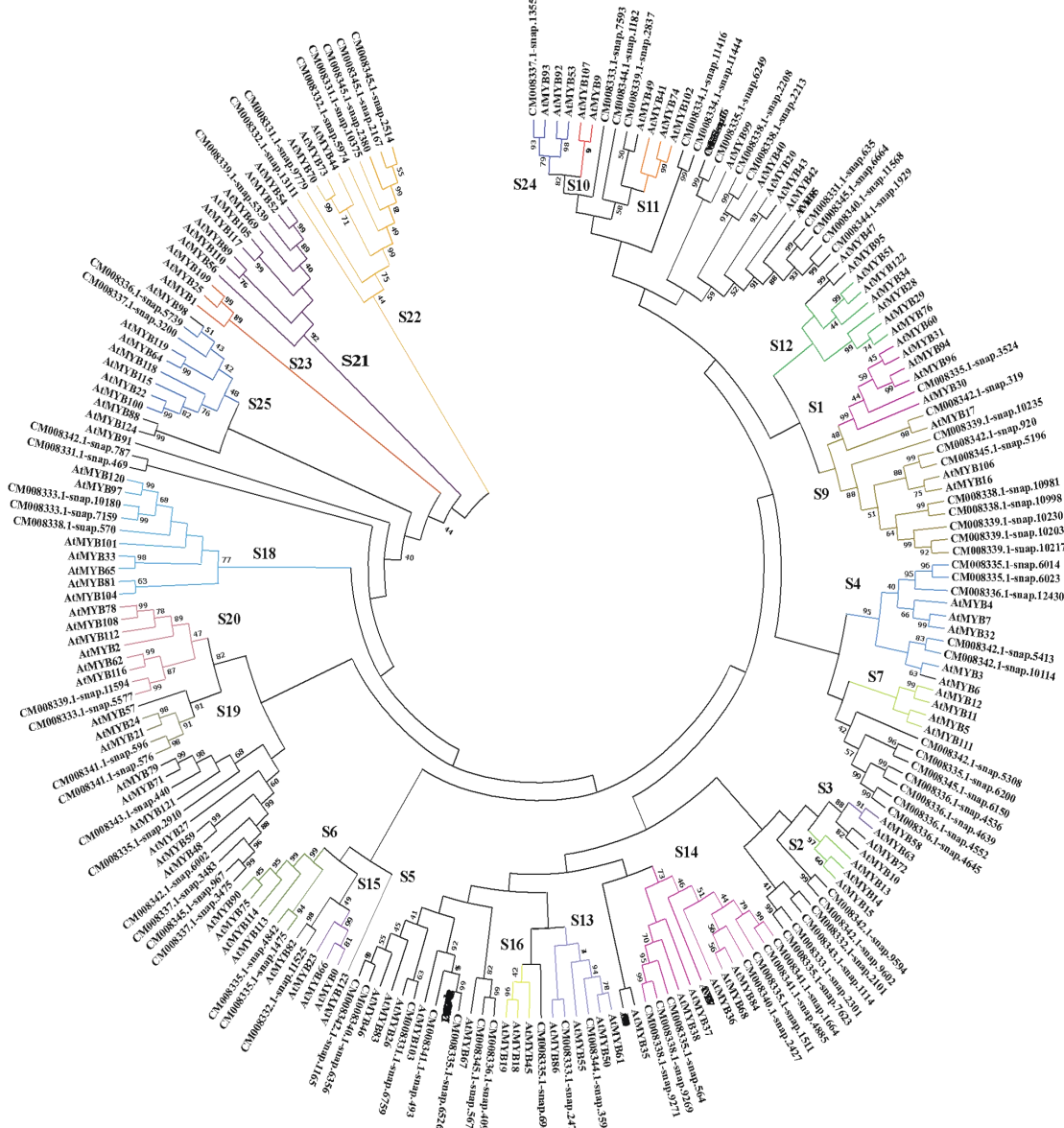


Figure 5 Phylogenetic tree of R2R3-MYB transcription factors in *I. batatas* and *A. thaliana*

1.6 Differential expression analysis of *I. batatas* R2R3-MYB genes under Fob stress and under low temperature stress

Fusarium wilt is an important fungal disease of sweet potato during plantation and the main pathogen is the *I. batatas* Fob. Previously, we isolated the strain F04 and strain F07 of the *I. batatas* Fob, which were used to infect the highly resistant cultivar JS57 and the highly susceptible cultivar XZH. The stems of the seedlings after 24h of infection were used for transcriptome sequencing. A total of 101 988 unigenes were obtained and 62 605 of them were functionally annotated (Lin et al., 2017). At that time, the genome sequence of *I. batatas*

has not been reported yet, thus it was not convenient for us to conduct in-depth analysis on the transcription factors including MYB.

In this study, we plotted the clustering chart of the expression pattern for the identified R2R3-MYB genes by taking advantage of the transcriptomics sequencing data of *I. batatas* under Fob infection and analyzed the expression variations of these genes. The results showed that the expression variations of the 88 R2R3-MYB genes were significantly different with regard to the three treatments. Compared to the control, the JS57_F04 group and JS57_F07 group showed no apparent discrepancy, while there were 9 genes displaying significant differential expression for the XZH_F07 group (Figure 6A), with 3 down-regulated (green) and 6 up-regulated (red). Among the 9 genes, except the down-regulated *CM008335.1-snap.6249* and the up-regulated *CM008342.1-snap.6002* that were not categorized into the *A. thaliana* subgroups, all other 7 genes were included in the *A. thaliana* subgroups. The down-regulated *CM008337.1-snap.1355* and *CM008339.1-snap.5339* were in S24 and S21, respectively. *CM008333.1-snap.5577* and *CM008339.1-snap.11594* were in S20. *CM008342.1-snap.10114* was in S4. *CM008345.1-snap.2167* and *CM008345.1-snap.2380* were in S22 (Figure 5 and Figure 6A). The R2R3-MYB transcription factors in these groups were relevant to the growth, development and stress response, etc.

Ji et al. (2020) stored the tuberous roots of cultivars Xushu15-1 and Xushu15-4 at °C for 0, 2 and 6 weeks respectively and conducted transcriptomics analysis. A total of 27 636 unigenes were obtained, 525 of which were identified to be differentially expressed under low temperature stress. In this study, we exploited the transcriptomics data to analyze the differential expression of the R2R3-MYB genes (Figure 6B) and differential expressions were observed in 27 of the R2R3-MYB genes, 10 of which varied in 4 treatment groups. Under low temperature stress, most of the differentially expressed R2R3-MYB genes were down-regulated, however, the magnitudes of change for the up-regulated genes were greater than those of the down-regulated genes. It could be seen that several R2R3-MYB transcription factors increased expression under low temperature stress, such as *CM008333.1-snap.5577* and *CM008335.1-snap.3524*. The expression variation of *CM008335.1-snap.3524* in Xushu15-1 was greater than in Xushu15-4, whereas the expression was down-regulated in Xushu15-1 for 6 weeks compared to that for 2 weeks. Additionally, the expression of *CM008336.1-snap.12430* was significantly down-regulated under low temperature stress, which was taken in *A. thaliana* S4 subgroup that was involved in the regulation of plant metabolic pathways (Figure 5).

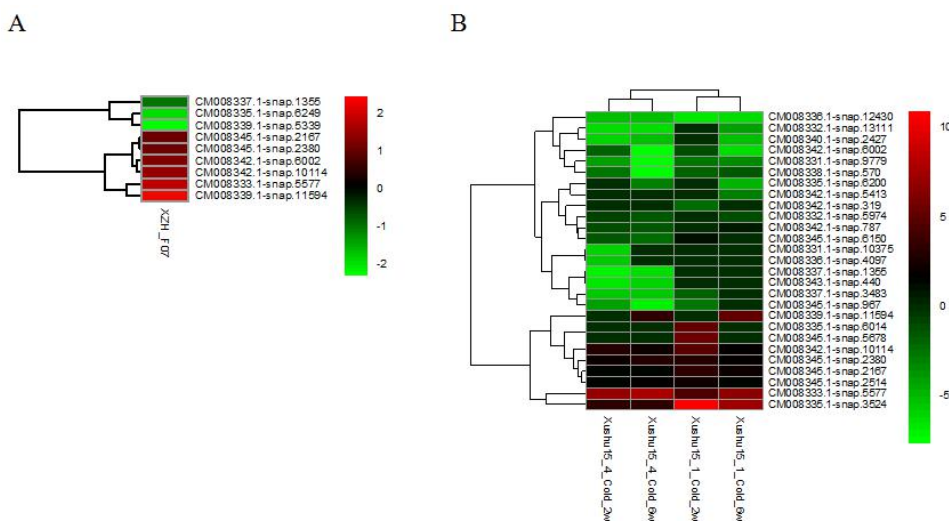


Figure 6 Differential expression of R2R3-MYB genes of *I. batatas* under Fob stress and cold stress

Note: A: Differential expression of R2R3-MYB genes of *I. batatas* under Fob stress; B: Differential expression of R2R3-MYB genes of *I. batatas* under cold stress

2 Discussion

MYB gene family is the largest transcription factor family in plant and is crucial for plant growth and development. In this study, the bioinformatic methods were applied to identify and obtain the 88 R2R3-MYB protein sequences in sweet potato. This number is relative less compared with those in other plants such as *A. thaliana* (Dubos et al., 2010) (126), *Solanum tuberosum L.* (Li et al., 2020) (111) and *Ananascomosus (L.) Merr* (Chen et al., 2019) (103). Sequence alignment of the 88 R2R3-MYB proteins of the *I. batatas* indicated that the R2 and R3 domains were highly conserved and the amino acid sequences of the R2 and R3 domains were very similar to those of the R2 and R3 domains in other plants.

In the present study, the classification of R2R3-MYB transcription factors of *A. thaliana* was used as a reference for the 88 R2R3-MYB transcription factors of *I. batatas* to build the phylogenetic tree together and yield the subgroup information for the *I. batatas* R2R3-MYB sequences. As for the R2R3-MYB transcription factors in the S9 subgroup of *A. thaliana*, AtMYB16 controlled the shape of petal epidermal cells; AtMYB17 regulated the early development of inflorescence and the germination of seeds; AtMYB106 was a negative regulator of trichome branching. There were 9 R2R3-MYB transcription factors of *I. batatas* enlisted to the S9 subgroup. These 9 *I. batatas* R2R3-MYB transcription factors all contained the motif 9, which was not found in other *I. batatas* R2R3-MYB sequences. Furthermore, the motif 10 was seen in only 6 *I. batatas* R2R3-MYB sequences, i.e., CM008345.1-snap.2167, CM008345.1-snap.2380, CM008345.1-snap.2514, CM008331.1-snap.10375, CM008332.1-snap.5974 and CM008332.1-snap.13111, which belonged to the S22 subgroup that were pertinent to stress responses.

In this study, we made use of the transcriptomics sequencing data to verify and analyze the R2R3-MYB transcription factors of *I. batatas* and found that of the 88 R2R3-MYB genes, 9 were differentially expressed after Fob infection and 27 were differentially expressed under low temperature stress. Lin et al. (2017) reported that the digital gene expression (DGE) of four genes, i.e., *c51550-g2*, *c52297-g1*, *c54420-g2* and *c56184-g1* were up-regulated by 1.62, 1.40, 1.71 and 3.10 folds, respectively. The result of the qRT-PCR verification showed that the four genes were up-regulated by 2.30, 4.88, 3.85 and 2.12 folds, which was consistent with the DGE data, suggesting the viability of the sequencing data and DGE data. Ji et al. (2020) selected 8 genes of the DGE results randomly and 4 potential candidate genes that might increase cold resistance for qRT-PCR verification and the results were in line with the expression patterns of these genes as revealed by DGE, indicating the reliability of the sequencing data. We reused these verified sequencing data for the verification and analysis of R2R3-MYB transcription factors and acquired useful results (Figure 6).

With respect to the R2R3-MYB transcription factors identified in this study, the number of differentially expressed genes varied considerably between under the Fob stress and under the low temperature stress. It might be that the infection of Fob into *I. batatas* was intrusive and some of the host pathways were likely to be blocked so that the number of response genes was less. In contrast, the *I. batatas* was capable of responding to low temperature stress at levels of tissue, cell and molecule, etc. The nonintrusive and systematic stress was less than Fob stress in terms of specificity, thus there were more R2R3-MYB genes in response to low temperature stress. Further, most of the response *I. batatas* R2R3-MYB genes were down-regulated, which was likely to be related to the metabolism down-regulation of *I. batatas* at low temperature. Nonetheless, there were a small number of up-regulated genes such as *CM008333.1-snap.5577* and *CM008335.1-snap.3524* (Fig. 6B), which might cope with low temperature stress. The number of up/down regulated R2R3-MYB transcription factors under Fob stress was relatively small, nevertheless, several genes showed comparative large magnitude of up/down regulation, such as *CM008339.1-snap.5339* (with log₂FoldChange -2.32), *CM008339.1-snap.11594* (with log₂FoldChange 2.45), which might be implicated in the mechanism of *I. batatas* resistance to Fob.

3 Materials and Methods

3.1 Genome sequence of *Ipomoea batatas*

The genome sequence of *I. batatas* was downloaded from <https://www.ncbi.nlm.nih.gov/genome/?term=Ipomoea+batatas> for the DNA sequences of the 15 chromosomes.

3.2 Gene annotation with snap

The snap program was utilized to predict the coding domain sequences (CDS) within the *I. batatas* chromosomes with the HMM models of *A. thaliana* (At.hmm), *Caenorhabditis elegans* (Ce.hmm), *Oryza sativa* (Os.hmm) and 178 458 protein sequences were obtained (Huang et al., 2020).

3.3 Prediction of genes with MYB domains

The HMM domain model of Myb_DNA-binding (PF00249) was downloaded from the Pfam website. The hmmsearch program was used to search the whole protein sequences of *I. batatas*. For the protein sequences with MYB domain, those with E-value \leq 1E-30 were selected to rebuild the Myb_DNA-binding HMM domain model specific to *I. batatas* with the hmmbuild program. The new HMM model was used to search the whole protein sequences of *I. batatas* to obtain the protein sequences with E-value \leq 0.01. With these protein sequences, the blastp program was applied to build the model database of the R2R3 type MYB domain of *I. batatas* for alignment to the whole protein sequences of the *I. batatas*. The MYB domain features of the obtained protein sequences were aligned and analyzed with NCBI Conserved Domains Tool, Interproscan and Clustal Omega. The augustus program was utilized for an alternative search of the CDS for the protein sequences without intact R2R3 domains. Finally, the R2R3 type MYB protein sequences were screened in compliance with the number of MYB domains within the protein sequence and the database of R2R3 MYB transcription factors was established.

3.4 Conservative analysis of R2R3-MYB transcription factors

The Clustal Omega was used for the multiple sequence alignment of the amino acid sequences of R2R3-MYB. The TBtools was utilized to select conservative sites and the domain feature sites were marked manually.

3.5 Conservative motif analysis

The conservative motifs of the *I. batatas* R2R3-MYB were analyzed with the MEME program, with the maximum number of searched motifs set as 10 and the search window set as 10-70 amino acids.

3.6 Chromosomal localization of *I. batatas* MYB genes

The sequence lengths of the 15 chromosomes were calculated and the information file was produced, which recorded the positions of the R2R3-MYB transcription factors within the chromosomes in line with the output ZFF file by snap program in section 3.2. The blastn program was applied to align each pair of the R2R3-MYB gene sequences to obtain the potential duplication relationship, which was verified by needle program (Rice et al., 2000). The positions of the R2R3-MYB genes in the chromosomes and the duplication relationships were drawn by circos program.

3.7 Construction of phylogenetic tree of R2R3-MYB

The phylogenetic tree of the R2R3-MYB protein sequences was constructed with the MEGA X program by referencing the 122 R2R3-MYB transcription factors of *A. thaliana*. The statistical method is neighbor-joining and the model/method is the p-distance. The gaps/missing data was set pairwise deletion and the No. of bootstrap replications was set 1 000.

3.8 Expression analysis of the R2R3-MYB transcription factors under disease stress and under abiotic stress

The transcriptomics data of the sweet potato under Fob stress (Lin et al., 2017) and under low temperature stress (Ji et al., 2020) were downloaded from EBI (<https://www.ebi.ac.uk/>) website. The sequencing samples of the Fob infection were JS57_F07, JS57_F04, JS57_CK, XZH_F07 and ZXH_CK, with three replicates for

each sample and 15 sequencing samples in total. The sequencing samples of low temperature stress were Xushu15_1_Cold_0w, Xushu15_1_Cold_2w, Xushu15_1_Cold_6w, Xushu15_4_Cold_0w, Xushu15_4_Cold_2w and Xushu15_4_Cold_6w, with three replicates for each sample and 18 sequencing samples in total. The Hisat2 (Kim et al., 2015) was used to build the index file of the *I. batatas* R2R3-MYB transcription factors for alignment to the transcriptomics sequencing samples. The aligned reads for each R2R3-MYB gene were counted and the DESeq2 program was utilized to standardize the read counts. The genes with fold change of experiment group vs control group ≥ 2 and P value < 0.05 were regarded as significantly up or down-regulated. The \log_2 FoldChange values ($\text{padj} < 0.05$) were used to draw the heatmaps of the R2R3-MYB transcription factors of *I. batatas* under Fob stress and under low temperature stress.

Authors' contributions

HXF and BCY designed and conducted the experiments. HXF and BCY carried out the data analysis and wrote the manuscript. HBF, XM and YZJ participated in experimental design and data analysis. LSQ and CXY conceived, supervised and guided the project.

Acknowledgments

This project was supported by the Fujian Provincial Department of Science & Technology (2019N0050).

References

- Chen Z., Hu F.C., Ruan C.C., Fan H.Y., Guo L.J., and Zhang Z.L., 2019, Bioinformatics and gene expression analysis of pineapple R2R3- MYB gene family, Redai Zuowu Xuebao (Chinese Journal of Tropical Crops), 40(10): 1958-1971
- Cominelli E., and Tonelli C., 2009, A new role for plant R2R3-MYB transcription factors in cell cycle regulation, Cell Res., 19(11): 1231-1232
<https://doi.org/10.1038/cr.2009.123>
PMid:19881525
- Dubos C., Stracke R., Grotewold E., Weisshaar B., Martin C., and Lepiniec L., 2010, MYB transcription factors in *Arabidopsis*, Trends Plant Sci., 15(10): 573-581
<https://doi.org/10.1016/j.tplants.2010.06.005>
PMid:20674465
- Gibbs D.J., Voß U., Harding S.A., Fannon J., Moody L.A., Yamada E., Swarup K., Nibau C., Bassel G.W., Choudhary A., Lavenus J., Bradshaw S.J., Stekel D.J., Bennett M.J., and Coates J.C., 2014, AtMYB93 is a novel negative regulator of lateral root development in *Arabidopsis*, New Phytol., 203(4): 1194-1207
<https://doi.org/10.1111/nph.12879>
PMid:24902892 PMCID:PMC4286813
- Huang X.F., Bi C.Y., Shi Y.Y., Hu Y.Z., Zhou L.X., Liang C.X., Huang B.F., Xu M., Lin S.Q., and Chen X.Y., 2020, Discovery and analysis of NBS-LRR gene family in sweet potato genome, Zuowu Xuebao (Acta Agronomica Sinica), 46(8): 1195-1207
- Huang Y.J., Deng R.Y., Gao F., Lao X.P., and Wu Q., 2015, Cloning and expression analysis of transcription factor gene *FtMYB21* from tartary buckwheat under abiotic stress, Jiyinzuxue Yu Yingyong Shengwuxue (Genomics Applied Biology), 34(9): 1939-1945
- Ji C.Y., Kim H.S., Lee C.J., Kim S.E., Lee H.U., Nam S.S., Li Q., Ma D.F., and Kwak S.S., 2020, Comparative transcriptome profiling of tuberous roots of two sweetpotato lines with contrasting low temperature tolerance during storage, Gene, 727: 144244
<https://doi.org/10.1016/j.gene.2019.144244>
PMid:31715303
- Kim D., Langmead B., and Salzberg S.L., 2015, HISAT: a fast spliced aligner with low memory requirements, Nat. Methods, 12(4): 357-360
<https://doi.org/10.1038/nmeth.3317>
PMid:25751142 PMCID:PMC4655817
- Li Y.M., Kui L.W., Liu Z., Allan A.C., Qin S.H., Liu Y.H., and Zhang J.L., 2020, Genome-wide analysis and expression profiles of the StR2R3-MYB transcription factor superfamily in potato (*Solanum tuberosum* L.), Int. J. Biol. Macromol., 148: 817-832
<https://doi.org/10.1016/j.ijbiomac.2020.01.167>
PMid:31962068
- Li Z.Y., Jiao Z.X., Chen Y., Gao F., and Zhou Y.J., 2019, Cloning and expression analysis of the *AmMYB44* gene from astragalus membranaceus, Jiyinzuxue Yu Yingyong Shengwuxue (Genomics Applied Biology), 38(8): 3605-3613
- Lin Y.L., Zou W.K., Lin S.Q., Onofua D., Yang Z.J., Chen H.Z., Wang S.L., and Chen X.Y., 2017, Transcriptome profiling and digital gene expression analysis of sweet potato for the identification of putative genes involved in the defense response against *Fusarium oxysporum* f. sp. *batatas*, PLoS One, 12(11): e0187838
<https://doi.org/10.1371/journal.pone.0187838>
PMid:29131830 PMCID:PMC5683638



- Millard P.S., Kragelund B.B., and Burow M., 2019, R2R3 MYB transcription factors - functions outside the DNA-binding domain, *Trends Plant Sci.*, 24(10): 934-946
<https://doi.org/10.1016/j.tplants.2019.07.003>
PMid:31358471
- Rice P., Longden I., and Bleasby A., 2000, EMBOSS: the European molecular biology open software suite, *Trends Genet.*, 16(6): 276-277
[https://doi.org/10.1016/S0168-9525\(00\)02024-2](https://doi.org/10.1016/S0168-9525(00)02024-2)
- To A., Joubès J., Thueux J., Kazaz S., Lepiniec L., and Baud S., 2020, AtMYB92 enhances fatty acid synthesis and suberin deposition in leaves of *Nicotiana benthamiana*, *Plant J.*, 103(2): 660-676
<https://doi.org/10.1111/tpj.14759>
PMid:32246506
- Vatansever R., Koc I., Ozyigit I.I., Sen U., Uras M.E., Anjum N.A., Pereira E., and Filiz E., 2016, Genome-wide identification and expression analysis of sulfate transporter (*SULTR*) genes in potato (*Solanum tuberosum* L.), *Planta*, 244(6): 1167-1183
<https://doi.org/10.1007/s00425-016-2575-6>
PMid:27473680
- Yang J., Moeinzadeh M.H., Kuhl H., Helmuth J., Xiao P., Haas S., Liu G.L., Zheng J.L., Sun Z., Fan W.J., Deng G.F., Wang H.X., Hu F.H., Zhao S.S., Fernie A.R., Boerno S., Timmermann B., Zhang P., and Vingron M., 2017, Haplotype-resolved sweet potato genome traces back its hexaploidization history, *Nat. Plants*, 3(9): 696-703
<https://doi.org/10.1038/s41477-017-0002-z>
PMid:28827752
- Zhu T.T., Liang D., and Xia H., 2016, Progress in the study of the regulation of anthocyanin synthesis by R2R3-MYB in fruits, *Jiyinzuxue Yu Yingyong Shengwuxue (Genomics Applied Biology)*, 35(4): 985-991