

# Natural variations of MeMYB2 in cassava varieties covering a wide range of the reactive oxygen scavenging enzymes under drought

**Bin Wang**

Institute of Tropical Bioscience and Biotechnology

**Xin Guo**

Institute of Tropical Bioscience and Biotechnology

**Shuxia Li**

Institute of Tropical Bioscience and Biotechnology

**Jiixin Liang**

Heilongjiang University

**Wenbin Liao**

Institute of Tropical Bioscience and Biotechnology

**Ming Peng** (✉ [mingpengcatas@gmail.com](mailto:mingpengcatas@gmail.com))

Institute of Tropical Bioscience and Biotechnology

---

## Research article

**Keywords:** Natural variations of MeMYB2, cassava varieties, reactive oxygen scavenging enzymes under drought

**Posted Date:** October 25th, 2019

**DOI:** <https://doi.org/10.21203/rs.2.16449/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background** Cassava (*Manihot esculenta* Crantz) is an important tropical crop with excellent stress tolerance and stress resistance genes. The antioxidant enzymes play an important role in regulating the cassava plants respond to stress. To date, little is known the relationship between natural variations of MYB genes that have been proven to regulate stress tolerance in model plants and the diversity of reactive oxygen scavenging enzymes under drought in cassava varieties. Therefore, we conducted this study to discover natural variations in MYB2 gene in association with the reactive oxygen scavenging enzymes.

**Results** Here, the relationship between the natural variation of MYB2 gene and the reactive oxygen scavenging enzymes under drought in cassava were analyzed. Natural variation of antioxidant enzymes in leaf and roots of 97 cassava varieties under drought analyses indicated that the variation of the various antioxidant enzymes is very different among the cassava varieties. MeMYB2 expression profiling analysis indicated that MeMYB2 was up-regulated obviously when induced by drought stress. Re-sequencing analysis of the 97 cassava varieties indicated that the MeMYB2 gene region included 87 single nucleotide variants (SNPs) and two insertion/deletion variants (Insert/ Deletion, Indel), the SNP locus of the coding region of MeMYB2 gene can be divided into 29 haplotypes, including 7 major haplotypes. These 7 major haplotypes can be classified into two categories, one of which is more closely related to wild cassava genotype W14, which indicates that MeMYB2 is positively selected during cassava breeding. Correlation analysis between MeMYB2 and drought tolerance showed that 12 functional SNPs in the MeMYB2 coding region were significantly associated with CAT activity, proline content, SOD activity and other traits, and the prediction of interaction with key genes of drought tolerance was further confirmed that MeMYB2 and key genes involved in CAT, proline and SOD can interact through intermediate metabolic pathways.

**Conclusions** Some natural variations exist in the MeMYB2, which might be involved in regulating the reactive oxygen scavenging enzymes (SOD and CAT) changes under stress. This study will facilitate the understanding of the roles of MYBs in regulating the reactive oxygen scavenging enzymes.

## Background

Cassava is an important tropical crop with drought tolerance, high light efficiency and high yield. The main growing areas of cassava are tropical and subtropical, how to deal with the seasonal drought in these areas is one of the main problems in improving cassava production. The cultivation of drought-resistant and high-yield cassava varieties through molecular-assisted breeding is an important way to solve this problem. During drought stress, reactive oxygen species (ROS) and proline play the important roles in the regulation of cassava drought stress, effective ROS clearance mechanism can delay leaf senescence and slow the formation abscission zones in leaf petioles to control leaf abscission (Liao et al. 2016a). Active oxygen scavenging enzymes such as proline, catalase (CAT), superoxide dismutase (SOD), and ascorbate peroxidase (APX) play the important roles in cassava drought tolerance (Liao et al.

2016a; Liao et al. 2016b; Xu et al. 2013a). Xu et al. overexpressed two ROS scavenging enzyme *MeCu/ZnSOD* and *MeCAT1* in cassava, which greatly enhanced the ability of cassava to scavenge reactive oxygen species, resisting drought stress, leaf greening, senescence in transgenic cassava when compared with the wild-type controls (Xu et al. 2013a, Xu et al. 2013b). Therefore, ROS scavenging system and proline content have important regulatory effects on the drought tolerance in cassava.

There is a significant difference in drought tolerance between cassava genotypes (Oliveira et al. 2017), different cassava genotypes may take different countermeasures to resist drought stress. For example, cassava cultivar SC124 (usually used as a drought-tolerant genotype) and Arg7 (usually used as a drought-sensitive genotype) adopt two distinct strategies of stopping growth and maintaining life to fall off old leaves (Zhao et al. 2015).). This significant phenotypic difference should be closely related to its genotype.

In model plants, many *MYB* members have been reported to be involved in mechanisms to resist drought stress. For example, overexpression of *MYB12*, one of the regulatory genes of the flavonoid pathway, reduces water loss (Nakabayashi et al. 2014). *MYB60* is a negative regulator of drought stress (Oh et al. 2011). *MYB41*, *MYB88*, *MYB44* and *MYB102* are also positive regulators of drought stress (Xie et al. 2010, Denekamp et al. 2003, Mengiste et al. 2003, Cominelli et al. 2008, Jung et al. 2008) *MYB2*, *MYB96* and *MYB15* are positive regulators of drought stress bind to the MYB binding site of the key genes such as *RD22*, thereby regulating ABA-dependent signaling pathways (Urao et al. 1996; Abe et al. 1997; Ding et al. 2009; Seo et al. 2009). *MYB2* is involved in the regulation of abiotic stresses such as drought stress, salt stress and phosphorus deficiency. Abe et al. reported that *AtMYB2* regulates *rd22* under drought or salt stress conditions. Drought and salt stress lead to ABA synthesis, ABA as a signaling molecule induces expression of bHLH-like proteins *rd22BP1* and *AtMYB2*, *rd22BP1* and *AtMYB2* bind to drought stress response gene *rd22*, the expression of the *rd22* gene was regulated by the *MYC* and *MYB* binding sites of the promoter region (Abe 1997). Yoo et al. reported that *AtMYB2* directly regulates calmodulin and enhances *Arabidopsis* resistance to salt stress (Yoo et al. 2005). *AtMYB2* responds to a phosphorus deficiency signal, and *AtMYB2* binds directly to the promoter region of the *miR399f* gene, regulating the expression of *miR399f*, thereby increasing the *Arabidopsis* tolerance to phosphorus deficiency (Baek et al. 2013). In addition, *AtMYB2* regulates senescence in whole plants by inhibiting cytokinin-mediated branch growth at the late growth stage (Guo and Gan 2011).

In our previous research, we reported the whole *MYB* superfamily genes in cassava genome, many *MYB* family members respond to drought stress and ethylene related pathways. In this study, We detailed the relationship between the natural variation of *MeMYB2* gene and reactive oxygen scavenging enzymes under drought, and found that there are 12 functional SNPs in the coding region of *MeMYB2* gene and significant correlation with CAT activity, proline content, SOD activity, sugar content, fresh weight of roots and storage root dry matter rate, suggesting that this gene may play an important regulatory role in the response of cassava drought stress.

## Results

## Natural variation of antioxidant enzymes in leaf and roots of different cassava varieties under drought

In order to investigate whether there are differences in the natural variation of antioxidant enzymes in different cassava varieties under the same drought stress, we detected antioxidant enzyme levels (SOD, CAT and proline) in different cassava varieties for two consecutive years (in the year of 2014 and 2015). Among the 97 cassava varieties (Additional file 1: Table S1), the highest level of SOD was in roots in 2015, which was 33.81, the minimum expression level of SOD was in leaves in 2014, and its value was 0.10; the highest level of CAT was in leaves in 2015, the value was 512.81, the minimum expression level of CAT is in root in 2014, and its value is 0.00; the highest level of proline was in roots in 2014, which was 66.34, the minimum expression level of proline was in root in 2015, and its value was 0.26. The results indicated that the natural variations of antioxidant enzymes in different cassava varieties under the same drought stress are very different (Additional file 2: Table S2). The statistical results of drought tolerance coefficient of various antioxidant enzymes are shown in Fig.1 and Table S2, the drought tolerance coefficients of various antioxidant enzymes are continuous quantitative traits, and the drought tolerance coefficient of antioxidant enzymes varies obviously. Among them, the largest drought tolerance coefficient of antioxidant enzymes was CAT-R-2014, the smallest was SOD-R-2014; the largest coefficient of variation was CAT-L-2015, and the smallest coefficient of variation was proline-L-2015 (Fig.1 and Additional file 2: Table S2). This indicates that the drought tolerance coefficient of various antioxidant enzymes is very different between the cassava varieties, and the variation is very rich.

### *MeMYB2* responds to drought stress in cassava

In our previous studies, we confirmed that the cassava leaf abscission zones were extremely easy to shed under drought conditions and the cassava abscission zones can be used to evaluate and screen the drought tolerance genes (Liao et al. 2016a; Liao et al. 2016b). Therefore, in order to understand the response of *MeMYB2* to drought in cassava, we used cassava leaf abscission zones to evaluate the *MeMYB2* response under drought. Specifically, we analyzed the sampling of the cassava leaf abscission zones at six time points under drought and analyzed by quantitative PCR. The results shown as the figure 1, *MeMYB2* was not detected to express at the early stage during drought (T2), however, *MeMYB2* was significantly induced to express from T3 to T6 time point under drought, moreover, the natural variation of *MeMYB2* were obviously increased as the drought progresses from T3 to T6 time point, the expression ratio increased from 13.762 at T3 time point to 30.1347 at T6 time point (Fig. 2). It is indicated that *MeMYB2* responded significantly to drought stress, and the induced natural variation of *MeMYB2* is also getting higher and higher as the drought becomes more and more serious.

### *MeMYB2* gene structure and conserved domain

The gene sequences of cassava germplasm AM560-2 were obtained from the JGI cassava genome database. The full length of the *MeMYB2* gene region is 1027bp long, including two introns, which are 148bp and 98bp in length respectively, and three coding exons, which are 160bp, 130bp and 517bp respectively. The CDS sequence of the *MeMYB2* gene is 807bp in length and encodes 268 amino acids. The CDS sequences were aligned in the conserved domain database of NCBI, two MYB binding domains were discovered in the N-terminus of *MeMYB2* that indicated the *MeMYB2* gene may belong to the R2R3 type MYB family members.

### The nucleotide diversity analysis of the cassava *MeMYB2*

The nucleotide diversity analysis of the cassava *MeMYB2* gene was carried out. The results are shown in Additional file 3: Table S3. The gene region includes 87 single nucleotide polymorphisms (SNPs) and two insertion/deletion polymorphisms (Insert/Deletion, Indel), the exon region has a total of 59 SNPs, including 17 synonymous mutations and 42 non-synonymous substitutions, 42 non-synonymous substitutions resulted in 40 amino acid differences, the conserved domain includes 2 synonymous substitutions and 12 non-synonymous substitutions. The DNA nucleotide diversity was analyzed by sliding window with a window size of 100bp and a step size of 25bp using DnaSP5 software, the results are shown in Fig. 3, the nucleotide diversity  $\pi$  value averaged is 0.01322, in detail, the highest acid diversity was discovered in the first intron, a total of 87 SNPs and 2 Indels (including 18-base SSR insertion deletion and 1 single-base insertion deletion) had 38 sites with MAF greater than 5%.

#### **Linkage disequilibrium and haplotype analysis of *MeMYB2***

The exon region of the *MeMYB2* gene includes 29 haplotypes and 7 major haplotypes (Fig. 4). All of these haplotypes can be divided into two categories three amino acid differences were discovered between the two categories and positive selection was indicated in the third exon region (Fig. 5),. Category II may be derived from wild type cassava W14, which may be introgressed into cassava cultivars during long-term breeding and may be produced by mutation in different directions. The linkage disequilibrium analysis of cultivars revealed that there were two haplotypes in the nucleotide variation of the exon region of *MeMYB2* gene, which were separated by recombination hotspots (Fig. 6). The three SNPs that divide the haplotype into two major categories are located in the haplotype 2, which can result in two amino acid variations (Fig. 6).

#### **The correlation analyses between *MeMYB2* and the drought tolerance in cassava**

The correlation analyses between candidate gene *MeMYB2* and the drought tolerance were carried out with the rare allelic variation filtered out under the condition of  $MAF > 0.05$ . The results are shown in Fig. 7. A total of 38 SNP/Indel marker sites in the *MeMYB2* gene region were involved in drought-related phenotypic analysis. *MeMYB2* has significant correlated with the drought tolerance coefficients of CAT, proline, SOD, fresh weight in the ground, root dry matter rate, root dry weight per plant, fresh weight per plant, root number per plant, reducing sugar, malondialdehyde and POD under the threshold conditions of  $P < 0.05$  and  $FDR < 0.05$ , When  $P < 0.01$ , *MeMYB2* was significantly correlated with CAT, proline, stored root dry matter rate, fresh weight per plant, root dry weight per plant, number of roots per plant and reducing sugar. When the screening conditions were  $P < 0.05$ , a total of 30 SNP loci and 1 intron SSR marker were significantly associated with drought tolerance traits, 21 of which were located in exons, 9 synonymous mutations, 12 non-synonymous mutation. The average interpretation rate of these SNPs for phenotypic variation was 10.13%, ranging from 3.99% to 39.14%. Non-synonymous mutations were significantly associated with phenotypic drought tolerance coefficients such as CAT, proline, SOD, reducing sugar, fresh weight of individual plants, number of roots per plant, and root dry matter rate per plant (Fig. 7 and Additional file 4: Table S4). Two groups of three SNPs with high linkage disequilibrium were indicated in the third exon region, which are located in the haplotype 2 interval. The first three SNPs are SNP1\_31465734, SNP1\_31465773 and SNP1\_31465774, which have significantly associated with the drought tolerance coefficient of CAT and SOD, the second group of 3 SNPs were

SNP1\_31465762, SNP1\_31465765 and SNP1\_31465804, which were significantly associated with reducing sugar and CAT (Additional file 4: Table S4).

In order to further understand the relationship between *MeMYB2* and drought tolerance at the genetic level, we used the Haploview 4.2 software to analyze the missense mutations in the related sites of the *MeMYB2* gene region. The block was divided into confidence intervals, statistical genes and traits. The haplotype consisting of functional SNP loci associated with drought tolerance coefficient, and the haplotype was correlated with the drought tolerance.

The *MeMYB2* gene region includes 12 non-synonymous replacement SNPs that are significantly associated with drought tolerance traits. The haplotype analysis of these 12 functional mutation sites were carried out, among which 9 SNPs constitute 1 monolithic block, including 5 haplotypes (frequency >0.01). The cassava materials were grouped according to the five haplotypes, and a total of 96 haplotype groupings and drought tolerance groupings were obtained. According to ANOVA analysis of drought tolerance and drought tolerance coefficient of single traits among haplotype groups, *MeMYB2* and CAT-L-2014 ( $P=2.0E-03$ ), Proline-L-2015 ( $P=0.038$ ) and POD-R-2015 ( $P = 9.0E-03$ ) was significantly associated (Table 1).

**Table 1** ANOVA analysis of drought tolerance and drought tolerance coefficient of single traits among haplotype groups

Traits		Sum of Square	F value	Significant Level
CAT-L-2014	Among	763.000	4.420	.002
	Within	5610.448		
	Total	6373.449		
Proline-L-2015	Among	20.613	2.643	.038
	Within	196.906		
	Total	217.520		
POD-R-2015	Among	162511.283	3.632	.009
	Within	1040164.070		
	Total	1202675.353		

#### Interaction prediction of *MeMYB2* with key genes of drought tolerance in cassava

In order to understand the functional interactions between drought-tolerant genes and *MeMYB2* gene, we used the *Arabidopsis* homologous gene *AtMYB2* of cassava *MeMYB2* and other drought-tolerant candidate genes we got, including the gene from the EST-SSR and SNP analysis, drought-tolerant related enzyme encoding genes, *Arabidopsis* homologs of all these cassava genes were introduced into AraNet for analysis. The results are shown in Table 2. *MeMYB2* and 15G040700 (*MYB15*) may have direct interaction relationship, 15G040700 (*MYB15*) based on SNP correlation analysis were significantly associated with CAT activity drought tolerance coefficient, *MeMYB2* can form a functional interaction relationship with the genes of CAT, SOD, proline, and the EST-SSR genes. Three new candidate pathway genes responding to drought stress and two SNP-associated CAT activity candidate genes establish a

functional interaction relationship, and then interact with the CAT enzyme coding gene to form a functional network; *MeMYB2* responds to the new metabolic pathway gene AT1G07400 through drought signal. Two cassava homologous genes 02G124800 (SNP-associated genes) and 02G125200 interacted with three SOD key genes; two new drought pathway responses to SNP-associated proline-containing drought tolerance coefficient. The candidate gene 13G095700 (*MeDHS1*) constitutes a functional interaction relationship, through an intermediate gene 09G024500 (*MeCPK32*), and functional interaction with *MeMYB2* and 03G119400, it is significantly associated with the drought tolerance coefficient of CAT enzyme activity, and also exists with 15G040700 and 18G061500 (Additional file 5: Table S5).

## Discussion

Transcription factors play an important role in regulation the physiological mechanism of plant drought tolerance. Several members of the MYB family in cassava have been reported to be involved in drought stress and chilling signal transduction pathways (Liao et al. 2016b; Li et al. 2017). The *MeMYB2* is a typical member of the MYB family and includes two MYB family DNA binding domains in its protein amino acid sequence. The *MeMYB2* gene is classified into the R2R3-MYB subclass (Liao et al. 2016b). The *Arabidopsis* homologous gene *AtMYB2* of *MeMYB2*, together with *AtMYC2*, is involved in the regulation of the drought stress response gene *rd22* in the ABA-dependent signaling pathway (Abe et al. 1997). According to our analysis, *MeMYB2* is up-regulated when cassava is suffered from drought stress, which indicates that this gene may play a positive role to resist drought stress.

The nucleotide sequence diversity of the *MeMYB2* gene region is very high, and an average of one SNP marker is included every 11.80bp, and the average value of the nucleotide diversity  $\pi$  value of the nucleotide in the gene region of *MeMYB2* is higher than that of the *Arabidopsis AtMYB2* gene (Kamiya et al. 2002). There are 14 SNPs in the conserved domain of *MeMYB2*, and there are 12 SNPs that cause amino acid variation. These SNPs may affect the binding activity of *MeMYB2*, and may cause the downstream genes of their regulation to produce expression differences between different cassava genotypes.

There is strong linkage disequilibrium between the SNPs in the *MeMYB2* gene region, and the SNP locus in the exon region constitutes a variety of haplotypes. There is at least one error leading to amino acid differences between different haplotypes. Moreover, haplotypes can be divided into two categories, one of which has the highest similarity to the *MeMYB2* sequence derived from the wild type cassava W14. According to the analysis of Nzuki et al. on the resequencing of cassava cultivars and W14 whole genome, a large fragment of W14 origin was introduced at 31 MB at the end of chromosome 1 of some cultivars (Nzuki et al. 2017). This indicates that *MeMYB2* in some cassava cultivars is derived from wild type cassava W14.

Analysis of the  $K_a/K_s$  of the *MeMYB2* gene region revealed that the third exon was strongly positively selected between the two major haplotypes, possibly due to the nucleotide variation of *MeMYB2* in this region against cassava, the contribution of agronomic traits is in line with breeding objectives and is thus positively chosen. The presence of three SNPs in the two haplotypes resulted in a difference of two

amino acids, and that these three SNP loci were significantly associated with the drought phenotype, suggesting that the nucleotide variation of *MeMYB2* may result in different cassava genotypes with differences in drought tolerance. The *MYB2* gene of interest has a very different evolutionary feature in the gene region. *MeMYB2* has a relatively long haplotype at the 5' and 3' ends of the gene region, including most polymorphic sites; it corresponds to the DNA binding domain of MYB and the transcriptional self-activation domain.

In this study, the *MeMYB2* gene was significantly associated with drought tolerance coefficients of multiple drought tolerance traits. To further analyze the relationship between *MeMYB2* gene and drought tolerance traits, we also attempted to perform haplotype association analysis on the functional SNP-related sites of *MeMYB2* gene. The monomer domains consisting of functionally related sites are located in the transcriptional autoactivation domain of the MYB gene, indicating that the *MeMYB2* gene may have an allele of transcriptional autoactivation activity in natural variation. Moreover, the haplotype association analysis between *MeMYB2* and individual traits showed significant differences between one and three drought tolerance traits among different haplotypes.

AraNet functional interaction analysis found that *MeMYB2* and CAT, SOD, Proline, POD may be in the same functional network, with the help of some intermediate genes such as *Hop3*, *GBF3*, *MYB102*, and *CPK32* and so on. The gene has functional interactions and directly interacts with the SOD enzyme-encoding gene to participate in the formation of these physiological phenotypes, which plays a regulatory role in the active oxygen scavenging mechanism and osmotic adjustment during drought stress.

## Conclusions

Some natural variations exist in the *MeMYB2*, which might be involved in regulating the reactive oxygen scavenging enzymes (SOD and CAT) changes under stress. Natural variation of antioxidant enzymes in leaf and roots of 97 cassava varieties under drought analyses indicated that the variation of the various antioxidant enzymes is very different among the cassava varieties. *MeMYB2* expression profiling analysis indicated that *MeMYB2* was up-regulated obviously when induced by drought stress. Re-sequencing analysis of the 97 cassava varieties indicated that the *MeMYB2* gene region included 87 single nucleotide variants (SNPs) and two insertion/deletion variants (Insert/ Deletion, Indel), the SNP locus of the coding region of *MeMYB2* gene can be divided into 29 haplotypes, including 7 major haplotypes. These 7 major haplotypes can be classified into two categories, one of which is more closely related to wild cassava genotype W14, which indicates that *MeMYB2* is positively selected during cassava breeding. Correlation analysis between *MeMYB2* and drought tolerance showed that 12 functional SNPs in the *MeMYB2* coding region were significantly associated with CAT activity, proline content, SOD activity and other traits, and the prediction of interaction with key genes of drought tolerance was further confirmed that *MeMYB2* and key genes involved in CAT, proline and SOD can interact through intermediate metabolic pathways. This study will facilitate the understanding of the roles of *MYBs* in regulating the reactive oxygen scavenging enzymes.

## Methods

# Analysis of gene structure and conserved domain

The DNA sequences and CDS sequences of the cassava *MeMYB2* gene was obtained from the JGI cassava genome database (<https://phytozome.jgi.doe.gov>) and imported into NCBI's Salign tool (<https://www.ncbi.nlm.nih.gov/sutils/splign/splign.cgi?textpage=online&level=form>) to analyze the length, number and starting position of its exons and introns, and use GSDS (<http://gsds.cbi.pku.edu.Cn/>) diagram of the gene structure. The CDS sequences of *MeMYB2* gene was introduced into NCBI's conserved domain database CDD (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) for sequence alignment and analysis of its conserved structure.

## Re-sequencing of the gene region

97 cassava genotypes were used to re-sequencing analysis the gene natural variations of *MeMYB2* gene and drought tolerance. The gene region of *MeMYB2* was re-sequenced using TA clone sequencing. Primers covering the entire length of the gene region were designed based on the DNA sequence of the cassava *MeMYB2* gene, and the primer sequences are shown in as the following: M2DF: ATGGACACTCAAGTAAGAAACCACG; M2DR: AATCAGTTGCTGCTGTAGAAACCAAA. The genomic DNA of the population consisting of cassava materials was used as a template. PCR amplification was performed using TAKARA's PrimerStar high-fidelity enzyme amplification, amplification protocols and procedures. The PCR product was subjected to agarose gel electrophoresis, gelatinization, and purification. Connect the T vector and transform E. coli, pick a single clone for PCR detection, send it to Shanghai for sequencing, and finally get 116 DNA sequence of the *MeMYB2* gene of the cassava genotype. The BWA software aligns the amplicon sequences to the cassava genome Cassava Genome Version 6.1 ([https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org\\_Mesculenta](https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Mesculenta), Bredeson et al. 2016), SNP for GATK software (V3.8-0, <https://software.broadinstitute.org/gatk/>) calling and low quality SNP filtering. The multiplex PCR-targeted resequencing work was completed in Shanghai Wing and Applied Biotechnology. The SNP information is imported into the snpEff software ([http://snpeff.sourceforge.net/SnpEff\\_manual.html](http://snpeff.sourceforge.net/SnpEff_manual.html)), and the obtained SNPs are functionally annotated using the cassava genome Version 6.1 sequence and annotation information.

## Nucleotide diversity analysis

The DNA sequence of the *MeMYB2* gene of cassava material was introduced into MEGA4.0 software (Tamura et al. 2007) for sequence alignment, the.MEG format file was saved, and the.MEG format file was opened using DnaSP5 software (Librado et al. 2009) to analyze the nucleoside. Acid diversity, statistical type and number of SNP variants.

# Haplotype analysis

The *MeMYB2* haplotype was extracted using the DnaSP5 software and saved as an NRF format file. The haplotype network of the *MeMYB2* gene was mapped using NETWORK5 (<http://www.fluxus-engineering.com>) and the sequence of the major haplotype was introduced into MEGA4.0 draws a phylogenetic tree. Linkage disequilibrium between the polymorphic sites of the two MYB genes was analyzed using Haploview version 4.2 software (Barrett et al. 2005).

## Candidate gene association analysis

According to the sequencing results, the SNP locus and genotyping of cassava material were extracted and imported into TASSEL2.1 soft (Yu et al. 2006), combined with drought tolerance-related phenotypic drought tolerance coefficient as a phenotype, genome-wide EST-SSR Marking the population structure Q-value matrix and family effect values as covariance for candidate gene association analysis, the significance association threshold was recorded as  $P < 0.05$ , and the P value was controlled by FDR (Benjamini and Hochberg 1995), and  $P_{FDR} < 0.05$  was considered to be significantly associated with the phenotype.

## Declarations

## Abbreviations

SNP, single nucleotide variant; Indel, insertion/deletion variant; CAT, Catalase; SOD, Superoxide dismutase; POD, Peroxidase; ROS, reactive oxygen species; APX, ascorbate peroxidase; MAF, Minor allele frequency; FDR, false discovery rate; ANOVA, Analysis of Variance; DTC, Drought tolerance coefficient.

## Funding information

This work was financially supported by National Key Research and Development Project (grant no. 2018YFD100003), and the National Natural Science Foundation of China (grant no. 31471551).

## Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files.

## Authors' contributions

BW, XG, JXL and WBL designed and performed the experiments, WBL and MP financed the experiments, SXL assisted in the experiments, BW and WBL analyzed the data and wrote the manuscript, SXL and MP edited the manuscript, and all authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

Not applicable.

## References

- Abe H, Yamaguchi-Shinozaki K, Urao T, Iwasaki T, Hosokawa D, Shinozaki K. Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell*. 2013;9:1859–1868.
- Baek D, Kim M-C, Chun H-J, Kang S-W, Park H-C, Shin G, Park J, Shen M-Z, Hong H, Kim W-Y, Kim D-H, Lee S-Y, Bressan R-A, Bohnert H-J and Yun D-J. Regulation of miR399f transcription by AtMYB2 affects phosphate starvation responses in Arabidopsis. *Plant Physiol*. 2013; 161:362–373.
- Bredeson J-V, Lyons J-B, Prochnik S-E, Wu G-A, Ha C-M, Edsinger-Gonzales E, Grimwood J, Schmutz J, Rabbi I-Y, Egesi C, Nauluvula P, Lebot V, Ndunguru J, Mkamilo G, Bart R-S, Setter T-L, Gleadow R-M, Kulakow P, Ferguson M-E, Rounsley S, Rokhsar D-S. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat Biotechnol*. 2016;34(5):562–57.
- Cominelli E, Sala T, Calvi D, Gusmaroli G and Tonelli C. Over-expression of the Arabidopsis AtMYB41 gene alter cell expansion and leaf surface permeability. *Plant J*. 2008;53:53–64.
- Denekamp M and Smeekens S-C. Integration of wounding and osmotic stress signals determine the expression of the AtMYB102 transcription factor gene. *Plant Physiol*. 2003;132:1415–1423.
- Ding Z-H, Li S-M, An X-L, Liu X, Qin H-J and Wang D-W. Transgenic expression of MYB15 confers enhanced sensitivity to abscisic acid and improved drought tolerance in Arabidopsis thaliana. *J Genet*

Gen. 2009;36:17–29.

Guo Y-F and Gan S-S. AtMYB2 regulate whole plant senescence by inhibiting cytokinin-mediated branching at late stages of development in *Arabidopsis*. *Plant Physiol.* 2011;156:1612–1619.

Jung C, Seo J-S, Han S-W, Koo Y-J, Kim C-H, Song S-I, Nahm B-H, Choi Y-D, Cheong J-J. Overexpression of AtMYB44 enhances stomatal closure to confer abiotic stress tolerance in transgenic *Arabidopsis*. *Plant Physiol.* 2008;146:623–635.

Kamiya T, Kawabe A and Miyashita N-T. Nucleotide polymorphism at the AtMYB2 locus of the wild plant *Arabidopsis thaliana*. *Genet Res.* 2002;80(2):89–98.

Li S-X, Yu X, Cheng Z-H, Yu X-L, Ruan M-B, Li W-B and Peng M. Global gene expression analysis reveals crosstalk between response mechanisms to cold and drought stresses in cassava seedling. *Front Plant Sci.* 2017;8:1259.

Liao W-B, Wang G, Li Y-Y, Wang B, Zhang P, Peng M. Reactive oxygen species regulate leaf pulvinus abscission zone cell separation in response to water-deficit stress in cassava. *Sci Rep.* 2016a;6:21542.

Liao W-B, Yang Y-Y, Li Y-Y, Wang G, Peng M. Genome-wide identification of cassava R2R3 MYB family genes related to abscission zone separation after environmental-stress-induced abscission. *Sci Rep.* 2016b;6:32006.

Librado P and Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009;25, 1451–1452.

Mengiste T, Chen X, Salmeron J and Dietrich R. The BOTRYTIS SUSCEPTIBLE1 gene encodes an R2R3MYB transcription factor protein that is required for biotic and abiotic stress responses in *Arabidopsis*. *Plant Cell.* 2003;15:2551–2565.

Nakabayashi, R, Yonekura-Sakakibara, K, Urano K, Suzuki M, Yamada Y, Nishizawa T, Matsuda F, Kojima M, Sakakibara H, Shinozaki K, Michael A-J, Tohge T, Yamazaki M and Saito K. Enhancement of oxidative and drought tolerance in *Arabidopsis* by overaccumulation of antioxidant flavonoids. *Plant J.* 2014;77:367–379.

Nzuki I, Katari M-S, Bredeson J-V, Masumba E, Kapinga F, Salum K, Mkamilo G-S, Shah T, Lyons J-B, Rokhsar D-S, Rounsley S, Myburg A-A, Ferguson M-E. QTL mapping for pest and disease resistance in cassava and coincidence of some QTL with introgression regions derived from *Manihot glaziovii*. *Front Plant Sci.* 2017;8:1168.

Oh J-E, Kwon Y, Kim J-H, Noh H, Hong S-W, Lee H. A dual role for MYB60 in stomatal regulation and root growth of *Arabidopsis thaliana* under drought stress. *Plant Mol Biol.* 2011;7:91–103.

- Oliverira E-J-D, Morgante C-V, Aidar S-D-T, Chaves A-R-D-M, Antonio R-P, Cruz J-L, Filho M-A-C. Evaluation of cassava germplasm for drought tolerance under field conditions. *Euphytica*. 2017;213:188.
- Seo P-J, Xiang F, Qiao M, Park J-Y, Lee Y-N, Kim S-G, Lee Y-H, Park W-J, Park C-M. The MYB96 transcription factor mediates abscisic acid signaling during drought stress response in *Arabidopsis*. *Plant Physiol*. 2009;151:275–289.
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4.0: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol*. 2007;24, 1596–1599.
- Urao T, Noji M, Yamaguchi-Shinozaki K and Shinozaki K. A transcriptional activation domain of ATMYB2, a drought-inducible *Arabidopsis* Myb-related protein. *Plant J*. 1996;10:1145–1148.
- Xie Z, Li D, Wang L, Sack F-D, Grotewold E. Role of the stomatal development regulators FLP/MYB88 in abiotic stress responses. *Plant J*. 2010; 64:731–739.
- Xu J, Duan X-G, Yang J, Beeching J-R, Zhang P. Coupled expression of Cu/Zn-superoxide dismutase and catalase in cassava improves tolerance against cold and drought stresses. *Plant Signal Behav*. 2013a; 8:e24525.
- Xu J, Duan X-G, Yang J, Beeching J-R and Zhang P. Enhanced reactive oxygen species scavenging by overproduction of superoxide dismutase and catalase delays postharvest physiological deterioration of cassava storage roots. *Plant Physiol*. 2013b; 161:1517–1528.
- Yoo J-H, Park C-Y, Kim J-C, Heo W-D, Cheong M-S, Park H-C, Kim M-C, Moon B-C, Choi M-S, Kang Y-H, Lee J-H, Kim H-S, Lee S-M, Yoon H-W, Lim CO, Yun D-J, Lee S-Y, Chung W-S and Cho MJ. Direct interaction of a divergent CaM isoform and the transcription factor, MYB2, enhances salt tolerance in *Arabidopsis*. *J Biol Chem*. 2005;280(5):3697–3706.
- Yu J, Pressoir G, Briggs W-H, Bi I-V, Yamasaki M, Doebley J-F, McMullen M-D, Gaut B-S, Nielsen D-M, Holland J-B, Kresovich S, Buckler E-S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*. 2006;38:203–208.
- Zhao P-J, Liu P, Shao J-F, Li C-Q, Wang B, Guo X, Yan B, Xia Y-J, Peng M. Analysis of different strategies adapted by two cassava cultivars in response to drought stress: ensuring survival or continuing growth. *J Exp Bot*. 2015;66(5):1477–1488

## Figures

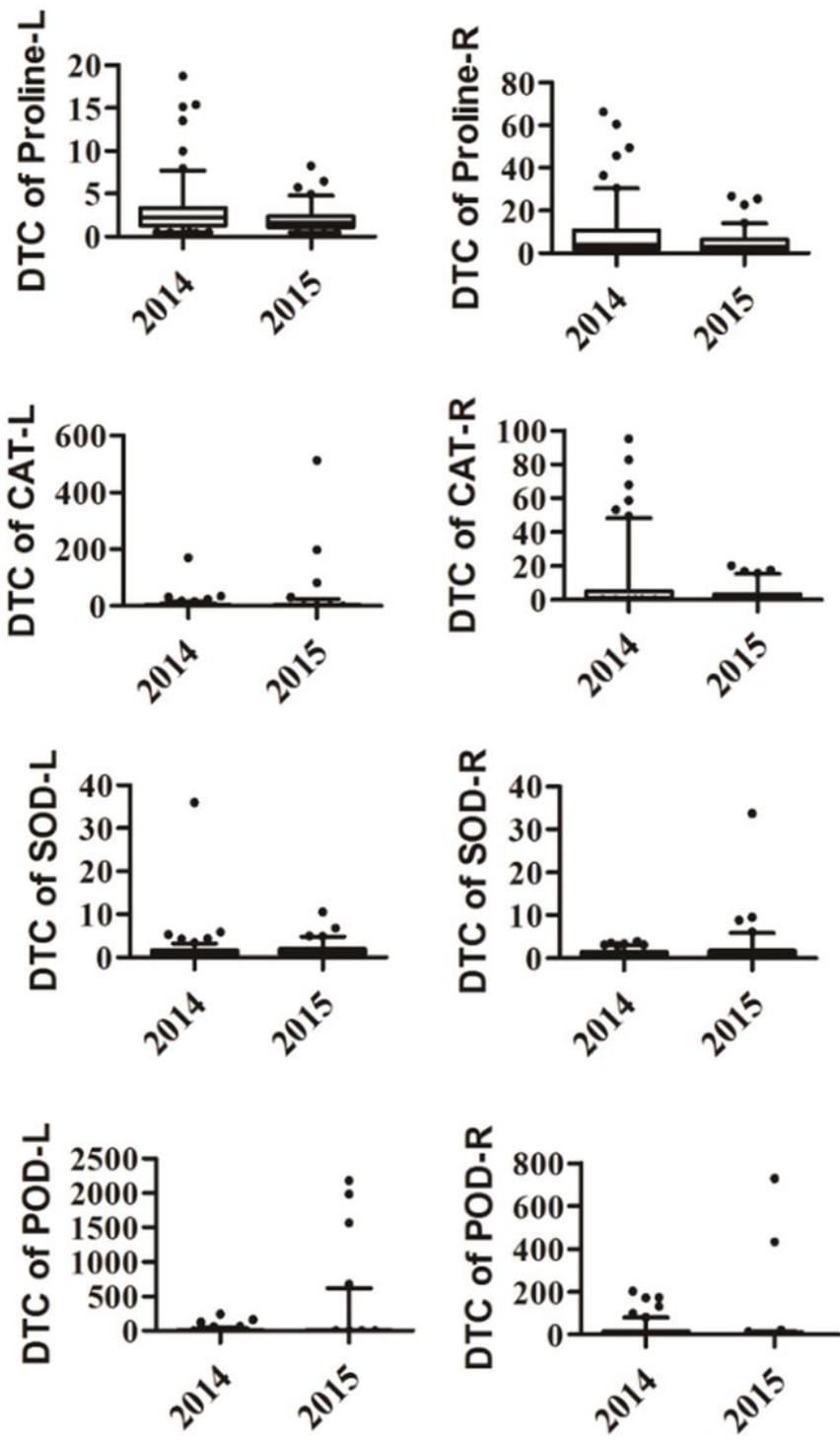
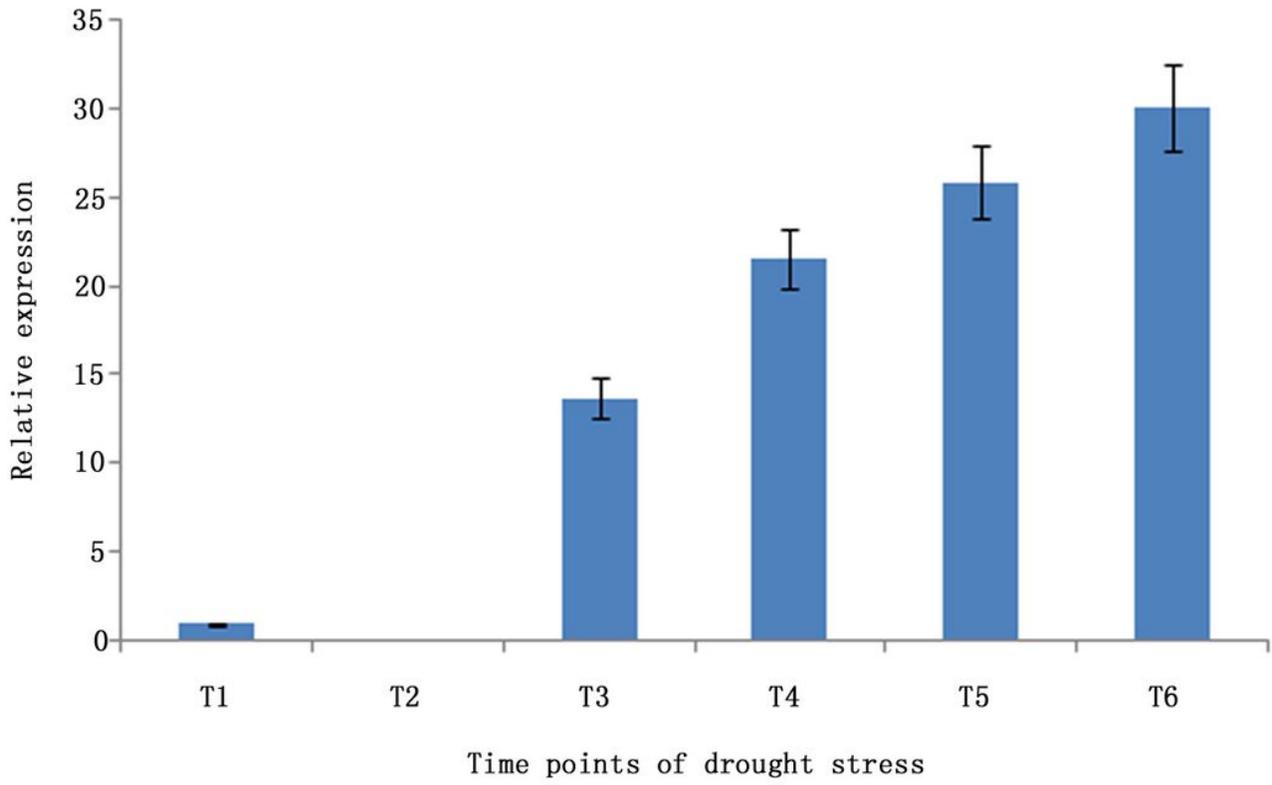


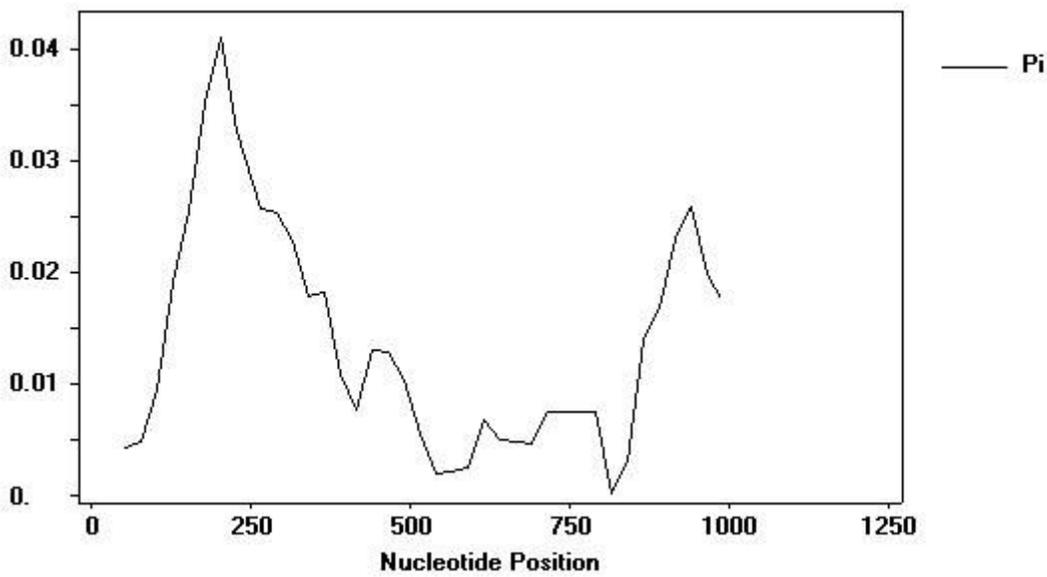
Figure 1

Boxplots of the antioxidant enzymes of drought tolerance coefficients (DTCs)



**Figure 2**

Expression profiles of MeMYB2 under drought stress



**Figure 3**

Nucleotide diversity of MeMYB2 by sliding windows analysis

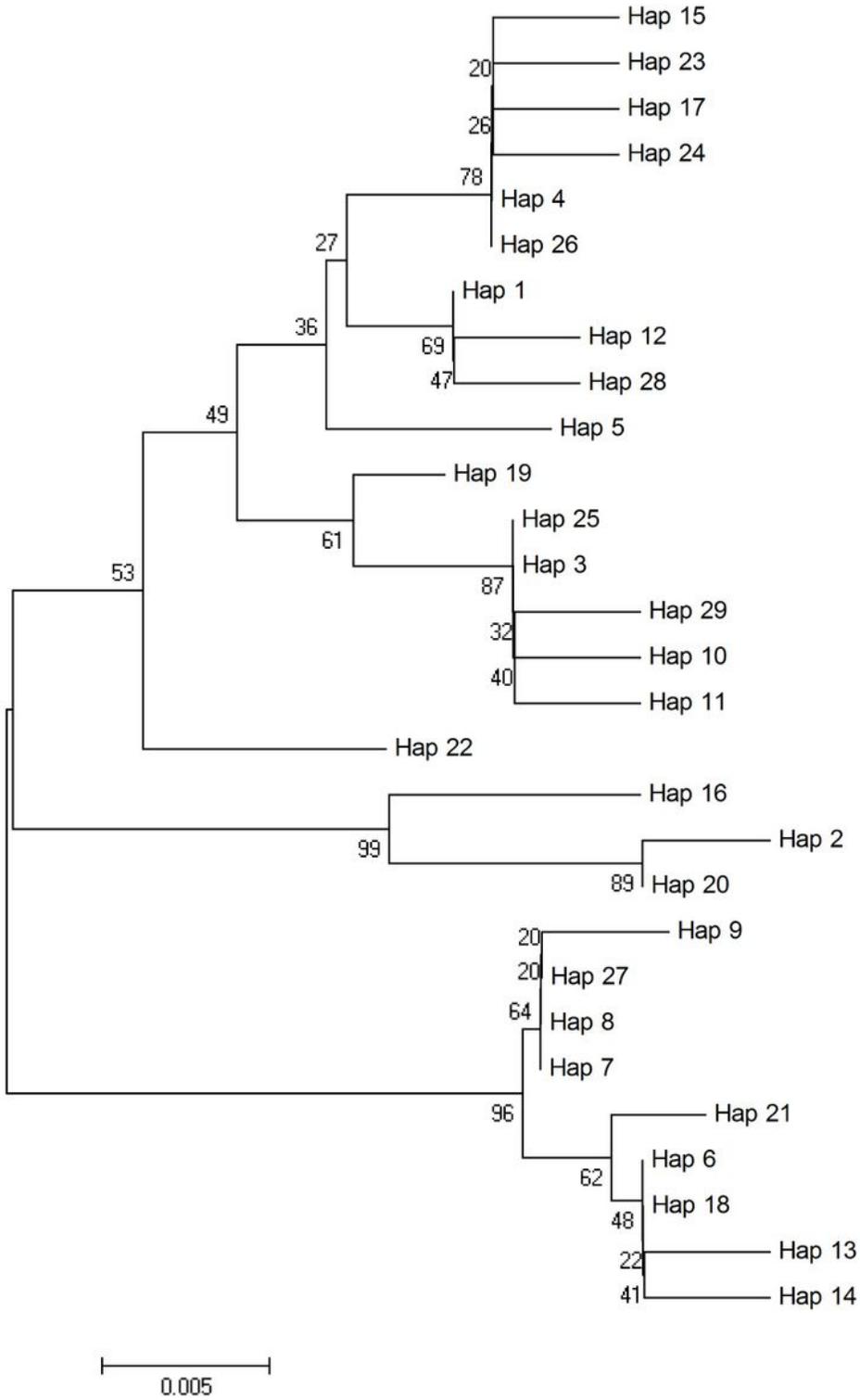
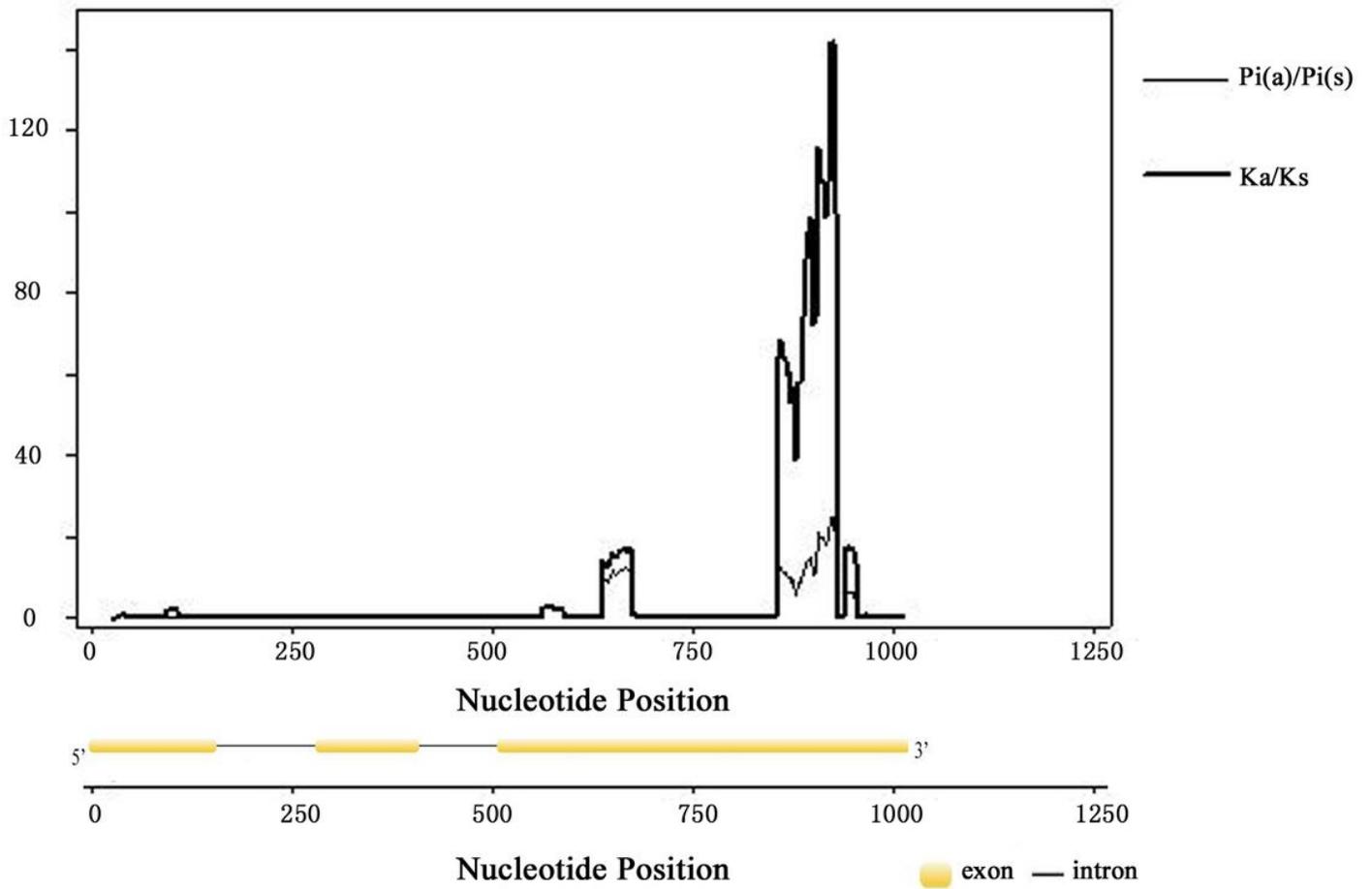


Figure 4

Haplotypes phylogenetic tree of MeMYB2 exons



**Figure 5**

Ka/Ks analysis of two cluster exon haplotypes of MeMYB2 by sliding windows method. Note Nucleotide position means the location which was relative to the first nucleotide A of the start codon.



Manhattan plot of MeMYB2 associated to drought tolerance related traits by candidate gene association mapping. Note: The horizontal dashed lines were the significance threshold  $P < 0.05$  and  $P < 0.01$  respectively.

## Supplementary Files

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

- [TableS3.docx](#)
- [TableS1.xlsx](#)
- [TableS4.docx](#)
- [TableS2.docx](#)
- [TableS5.docx](#)