**MeMYB26, a drought-responsive transcription factor in cassava (Manihot esculenta Crantz)**

Bin Wang¹, Xin Guo², Pingjuan Zhao², Wenbin Liao², Changying Zeng³, Kai Li¹, Yuxun Zhou¹, Junhua Xiao¹*, Mengbin Ruan², Ming Peng², Yang Bai⁴ and Ye Chen⁴

**Abstract:** Cassava (Manihot esculenta Crantz) is worldwide the sixth-largest food crop and highly drought resistant. The expression of the myeloblastosis (MYB) family protein MeMYB26 in cassava was previously reported to be significantly up-regulated under water deficit. To study the basic characterization, natural variations and potential functions of MeMYB26 for drought resistance in cassava, the protein was subjected to bioinformatics analysis, multiplex-PCR with next-generation sequencing and candidate association studies. The results indicated that MeMYB26 is a typical transcription factor, with two MYB DNA-binding and transcriptional activation domains. One stop-gained and five nonsynonymous variations in the genomic region of MeMYB26 were significantly associated with drought resistance traits. The results of the scale-free coexpression network showed that the MeMYB26 gene plays a critical role in plant stress resistance, growth and biomass development. It was concluded that MeMYB26 is a reliable candidate gene associated with drought tolerance and biomass storage in cassava.

**Keywords:** MYB, drought resistance, natural variation, candidate association study, coexpression interaction network.

**INTRODUCTION**

Cassava (Manihot esculenta Crantz) is an important food and energy crop (Vieira et al. 2020) that is mainly planted in tropical and subtropical areas. Just as other crops suffer from drought stress and lose a lot of yield (Santos et al. 2020), cassava is often affected by seasonal drought and is highly tolerant to intermittent drought. Therefore, drought resistance has been a key focus of cassava breeding studies.

The myeloblastosis (MYB) protein family is one of the largest transcription factor families in plants. Many MYB proteins have been reported to be involved in the resistance mechanism to dehydration stress in plants (Tan et al. 2019). In cassava, MYB proteins have also been thoroughly studied and resulted in the identification of 299 MYB proteins, consisting of 150 R2R3-MYB and 146 R1-MYB and MYB analogs, and only 2 R1R2R3-MYB and 1 4R-MYB, distributed on 18 cassava chromosomes (Ruan et al. 2017). A total of 26 R2R3-MYB genes were responsive to both ethylene and drought stress signals (Liao et al. 2016). Other studies reported 26 cassava MYB or MYB-related proteins that responded to drought and cold stress signals, 14 of which were up-regulated and 12 down-
regulated (Li et al. 2017a). In addition, MeMYB26 was reported to be located in the linkage disequilibrium (LD) decay region of the peak SNP of 25.19 Mbp on chromosome 3, and described as a candidate gene significantly associated with cassava leaf width (Zhang et al. 2018). Moreover, MeMYB26 was found to be specifically expressed in roots and significantly responsive to water deficit signals in roots (Ruan et al. 2017).

In this sense, this study investigated MeMYB26 in more detail, with a view to determining its relationship with drought tolerance and identifying functional genes and potential markers for drought resistance research and breeding. Basic characterization and candidate association studies were performed and the natural variations and coexpression networks were analyzed to understand the potential functions of MeMYB26. The final results provided useful information about the potential functional markers of MeMYB26 for marker-assisted breeding (MAB) in the future.

MATERIAL AND METHODS

Material

A total of 100 cassava accessions used for resequencing and candidate association studies from 12 countries and areas were used (Table S1).

Analysis of basic characteristics

The genomic and coding sequences (CDS) of MeMYB26 were downloaded from the cassava genome version 6.1 (Bredeson et al. 2016). The Splign tool of the National Center for Biotechnology Information (NCBI) was used to analyze the gene structure. A diagram was drawn using Gene Structure Display Server (GSDS) and the conserved domain database (CDD) of NCBI for conserved domain analysis.

To mediate the transcriptional activation, vector MeMYB26::pGBKT7 was constructed using the forward primer M26F (sequence 5' GGAATTCCATGTTGTTAGTAGAGGAG 3') with the enzyme-digestion site NdeI and reverse primer M26R (sequence 5' CGGGATCCAATTGAATGCTAGCTTGGTA 3') with the enzyme-digestion site BamHI. Yeast self-activation was experimentally tested as described by Ruan et al. (2017)

Natural variation analysis

Young leaves of 100 cassava accessions were used to extract genomic DNA by methods described elsewhere (Kunkeaw et al. 2011). Multiplex-PCR with next-generation sequencing was used to re-sequence the genomic region of MeMYB26 (Chen et al. 2016). The primers for multiplex-PCR are listed in Table S2. Multiplex-PCR with NGS for targeted re-sequencing was carried out by the company Shanghai Biowing Applied Biotechnology Co. LTD. The SNPs were annotated by software snpEff and the linkage disequilibrium (LD) and haplotypes by Haploview version 4.2 (Barrett et al. 2005).

Candidate association study

To associate MeMYB26 to drought resistance phenotypes, the candidate gene association study was performed. The drought tolerance coefficients (DTCs) of 100 cassava accessions, including DTCs of biomass-related and physiological traits were chosen from previously published data (Wang et al. 2017) were used as phenotypes and the results from target resequencing as genotypes. The Q and kinship (K) matrices of 100 cassava accessions were inferred from 100 of the previously reported EST-SSR profiles (Wang et al. 2017). Thereafter, the model Q+K+MLM of software TASSEL 2.1 was applied in the candidate association study (Yu et al. 2006). The P-value was controlled based on the false discovery rate (FDR), while SNP loci were considered significant associations at P-values lower than 0.05 after FDR control.

Coexpression network analysis

The expressions of the gene group regulating the same process were often significantly correlated. So, the possible functions of MeMYB26 could be inferred from the gene interactions. Previously reported drought stress-related RNA-seq data (Hu et al. 2016) were used to construct scale-free coexpression networks by the weighted correlation network analysis (WGCNA) package of R software. To calculate GO and KEGG enrichment analysis of the module members TBtools software was used (Chen et al. 2020). The expression heat map and correlation coefficients were calculated by the pheatmap package and the COR function of R software.
MeMYB26, a drought-responsive transcription factor in cassava (Manihot esculenta Crantz)

RESULTS

MeMYB26 is a typical MYB family transcription factor

To study the basic characterizations of MeMYB26, gene structure was analyzed by Splign. The MeMYB26 gene region included three exons and two introns (Figure 1A). The MeMYB26 contained 253 amino acids including two MYB DNA-binding domains. As shown in the neighbor-joining phylogenetic tree (Figure 1B), MeMYB26 was clustered in the same branch with AtMYB79 (At4G13480), AtMYB305 (At3G24310), and AtMYB121 (At3G30210). To examine the transcriptional activation potential, vector MeMYB26:pGBKT7 was constructed and transformed in yeast Y187. Three transformants of MeMYB26::pGBKT7/Y187 formed obvious blue plaque, while the plaque on the line of pGBKT7/Y187, the negative control, was milky white (Figure 1C). This indicates that MeMYB26 has a transcription activation domain (AD). In other words, MeMYB26 is a typical R2R3-MYB transcription factor that has two MYB DNA-binding and one transcription activation domain.

Natural variations of MeMYB26

Sequence variations in cassava germplasms might account for the functional differences of MeMYB26 among cassava accessions. Multiplex-PCR with next-generation sequencing was applied to the re-sequencing genomic sequence of MeMYB26. A total of 21 high-quality SNVs were discovered, including variations of 19 exons and 2 introns in genomic regions that did not contain the UTR regions of MeMYB26 (Table 1). Among them were 10 missense variations, 1 single base insertion/deletion frameshift mutation, and 1 stop-gained and 7 synonymous variations. Further, 1 nonsynonymous

![Figure 1. Basic characterizations of MeMYB26. A. Gene structure. B. Phylogenic tree of MeMYB26 and MYB transcription factor family of Arabidopsis thaliana. Peptide sequences of MeMYB26 and a total of 131 Arabidopsis MYB proteins were used to construct the neighbor-joining (NJ) phylogenic tree. C. Transcriptional activation analysis of MeMYB26. Transformed yeast Y187 strains of MeMYB26:pGBKT7 DNA construct were grown on an SD/–Trp plate containing 20 μg mL⁻¹ X-α-al.](image-url)
and 4 synonymous variations were discovered in the conserved MYB domain. The minor allele frequency (MAF) of these 30 SNP loci, consisting of 21 SNPs in the genomic sequence and 9 SNPs in the UTR region, was on average 24.34%, ranging from 0.49% to 50%.

The association analysis was based on linkage disequilibrium (LD). So, LD analysis was performed to evaluate LD strength between every pair of SNPs. With MAF>0.05, 25 SNPs were obtained including exons, introns and UTRs. There were 300 combinations, consisting of 9 combinations with 1.0>R\textsuperscript{2}>0.7; 27 with 0.7>R\textsuperscript{2}>0.2; 39 with 0.2>R\textsuperscript{2}>0.1 and 225 with R\textsuperscript{2}<0.1. Four LD blocks, including six, four, two, and three SNP loci (Figure 2A) were found, which contained 7, 4, 2 and 2 haplotypes, respectively (Table S3). This indicated that the recombination rate was relatively high among the polymorphic loci in the whole gene region, which can explain the small size of each LD block.

**MeMYB26** was significantly associated with drought-resistance traits

To discover the relationship between natural variations of **MeMYB26** and cassava drought resistance, the candidate gene association study was performed by using previously reported drought resistance traits and EST-SSR profiles of the same cassava accessions. The Q+K+MLM model was applied at MAF>0.05 for the candidate association study. As shown in Figure 2B, a total of 20 SNPs of **MeMYB26** were significantly associated with DTCs of catalase (CAT), malondialdehyde (MDA), peroxidase (POD), superoxide dismutase activity (SOD), soluble reducing sugar (SRS), proline content, above-ground fresh weight (AGFW), storage root dry matter percentage (SRDMP), storage root dry weight (SRDW), storage root number (SRN) and storage root fresh weight (SRFW), at P<0.05. Eleven of them were located in exons, including 4 synonymous, 6 nonsynonymous, and 1 stop-gained variation. Seven variations, containing the 6 nonsynonymous and 1 stop-gained variations mentioned above, were significantly associated with the DTCs of CAT, MDA, SRN, proline, SOD, and AGFW. The mean contribution rate of phenotypic variation was 8.80%, ranging from 3.45% to 42.32%. Meanwhile, the variations of **MeMYB26** were significantly associated with DTCs of CAT, SOD, SRN and AGFW, at P<0.01.

For a better understanding of the associations of **MeMYB26** with drought resistance at the gene level, a total of 7 SNPs, which were annotated as functional variations and significantly associated with DTCs in the single marker association tests, were used to construct haplotypes. As shown in Figure 2C, six of them could be assigned to one LD block with six haplotypes (frequency > 0.01). These 7 SNPs were meaningful for marker-assisted selection in breeding of drought-tolerant cassava without loss of biomass storage. However, no significant associations between these haplotypes and drought resistance traits were detected. This might be due to the multiplexity of drought tolerance and contrasting functions of SNPs.

**Coexpression networks related to MeMYB26**

The possible interaction genes of **MeMYB26** could be used to draw conclusions about the potential functions of **MeMYB26**. The coexpression networks were constructed based on the drought-related RNA-seq data in cassava. The results showed that all of the 30,665 genes could be divided into 97 modules, among which **MeMYB26** was clustered into the green module, which included 1041 genes (Figure 3A). Gene ontology (GO) enrichment analysis showed that the molecular functions (MF) included transcription regulator activity, binding, and catalytic activity. The analysis of the enriched cell component (CC) detected cellular anatomical and intracellular entities. The biological process (BP) included 12 GO terms, of which the most abundant three were biological regulation, development process, and response to stimulus. The results of GO enrichment analysis showed that genes of the green module might be mainly involved in the biological processes of catalytic activity and transcription factor regulation in cells, such as gene expression, growth and development, and stimulus-response (Figure 3B). The green module was also enriched in some important pathways. As shown in Figure 3C, the green module was significantly enriched in nine pathways (P<0.05), and the three most significant pathways were transcription factor, phylopanoid biosynthesis and biosynthesis of other secondary metabolites.
MeMYB26, a drought-responsive transcription factor in cassava (*Manihot esculenta* Crantz)

Otherwise, the green module may be related to the growth and respond to abiotic stress through these pathways. Besides, it was also enriched in plant hormone signal transduction, glycolysis/gluconeogenesis and starch and sucrose metabolism, which were related to stress signal transduction, biomass development, and starch accumulation under stress. Glycerophospholipid metabolism and metabolism of terpenoids and polyketides played roles in plant antioxidants.

**Figure 2.** Natural variations and candidate association study. A. Linkage disequilibrium analysis of MeMYB26. Linkage disequilibrium plot of MeMYB26 was drawn by Haploview 4.2 software. Different colors were used to distinguish LD strength between each locus pair. B. Manhattan plot of MeMYB26 associated with drought tolerance-related traits by candidate gene association mapping. The horizontal dashed lines indicate the significance thresholds P<0.05 and P<0.01. C. Haplotypes of functional SNPs. Total of 7 functional SNPs located in the exons of MeMYB and associated with DTCs. Six SNPs were assigned to one LD block with 6 haplotypes whose frequency was larger than 0.01.
The FPKM value of these 1041 genes was extracted from the cassava drought stress RNA-seq data. After data filtering and conversion, the expression values of the remaining 723 genes were obtained. The results showed that expression patterns of these genes responded similarly to drought stress, which was mainly reflected in the roots of cassava cv. Arg7 (Figure S1). Except for a small portion of genes (33/723), which were significantly down-regulated or did not differ, almost all genes (690/723) were up-regulated, including 529 genes that responded significantly to drought stress in cv. Arg7.

To focus on the genes most strongly correlated with MeMYB26, the 30 closest correlations with MeMYB26 were selected. The expression patterns of these 31 genes were similar (Figure 3D). All of them were significantly up-regulated in roots of cassava cv. Arg7, but significantly down-regulated or without significant difference in the leaves of the three cassava genotypes. Among the top 30 correlated genes (Table 2), three others were MYB proteins, six were annotated as starch transporting and sugar metabolizing genes and four play a role in the redox reaction. Three other genes were related to biosynthesis and metabolism of the plant hormones ABA and IAA. The expression patterns of these 30 genes were similar to that of MeMYB26. Consequently, MeMYB26 is believed to be associated with drought resistance and biomass storage under water deficit.

Figure 3. Coexpression networks and expression pattern analysis. A. Coexpression networks that were constructed with the WGCNA package of R software based on previously reported RNA-seq data (Hu et al. 2016). Different colors stand for different modules. B. Gene ontology (GO) enrichment analysis of the green module. Only level 2 GO terms at p<0.05 were used in the bar plot. C. KEGG pathway enrichment analysis of the green module. D. Expression heat maps of MeMYB26 and the top 30 related genes. The three cassava genotypes Arg7, SC124, and W14 have frequently been used in drought treatment experiments. Heatmap drawn based on the log value (foldchange, base 2).
MeMYB26, a drought-responsive transcription factor in cassava (*Manihot esculenta* Crantz)

**Table 2.** Annotations of 16 genes and number of MYB-related elements in their promoter region

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>No. of MYB-related elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>01G074400</td>
<td>MYB domain protein 42</td>
<td>5</td>
</tr>
<tr>
<td>15G047000</td>
<td>MYB domain protein 79</td>
<td>2</td>
</tr>
<tr>
<td>02G058900</td>
<td>MYB transcription factor</td>
<td>4</td>
</tr>
<tr>
<td>10G149300</td>
<td>Beta-glucosidase (E3.2.1.21)</td>
<td>7</td>
</tr>
<tr>
<td>10G149400</td>
<td>Beta-glucosidase (E3.2.1.21)</td>
<td>3</td>
</tr>
<tr>
<td>07G001700</td>
<td>Beta-glucosidase (E3.2.1.21)</td>
<td>13</td>
</tr>
<tr>
<td>02G190300</td>
<td>Sucrose transport protein suc1-related</td>
<td>4</td>
</tr>
<tr>
<td>14G047700</td>
<td>Bidirectional sugar transporter sweet10</td>
<td>7</td>
</tr>
<tr>
<td>15G183200</td>
<td>Bidirectional sugar transporter sweet4-related</td>
<td>8</td>
</tr>
<tr>
<td>15G014000</td>
<td>Oxidoreductase, 2og-fe ii oxygenase family protein</td>
<td>3</td>
</tr>
<tr>
<td>02G019500</td>
<td>Oxidoreductase, 2og-fe(ii) oxygenase family protein</td>
<td>7</td>
</tr>
<tr>
<td>14G063900</td>
<td>Glutaredoxin-c1</td>
<td>6</td>
</tr>
<tr>
<td>10G055900</td>
<td>Isoflavone-7-O-beta-glucoside 6''-O-malonyltransferase / Flavone/flavonol 7-O-</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>-beta-D-glucoside malonyltransferase</td>
<td></td>
</tr>
<tr>
<td>03G150400</td>
<td>9-cis-epoxycarotenoid dioxygenase NCED6, chloroplastic</td>
<td>6</td>
</tr>
<tr>
<td>03G083500</td>
<td>9-cis-epoxycarotenoid dioxygenase NCED3, chloroplast-related</td>
<td>2</td>
</tr>
<tr>
<td>04G136800</td>
<td>IAA-amino acid hydrolase ilr1-like 1-related</td>
<td>13</td>
</tr>
</tbody>
</table>

**DISCUSSION**

Cassava is an important food crop, known as the underground granary and an important source of industrial starch. Even under drought stress, cassava still produces reasonable yields. Here, one MYB transcription factor, *MeMYB26*, previously reported as responsive to water deficit, was studied in detail (Ruan et al. 2017). According to these results, *MeMYB26* is a typical MYB protein. Its protein includes two MYB DNA-binding domains and was assigned to the R2R3-MYB subgroup, and the experiments showed that *MeMYB26* had the transcriptional activation domain (AD). *MeMYB26* was significantly up-regulated during drought stress, which indicated that it might play a positive role in cassava drought resistance. *MeMYB26* was closest to *AtMYB71*, *AtMYB79*, and *AtMYB121* in the same phylogenetic tree branch. Since reports on the three *Arabidopsis* MYB proteins are rare, it is impossible to draw conclusions about the potential functions of *MeMYB26* according to its homologous genes.

Natural variations are critical to study further variations in germplasm and select excellent alleles of the target gene. The heterozygosity of *MeMYB26* is high, which might be related with the domestication for human consumption in cassava breeding. Due to the impossibility of effective recombination, many harmful variations were masked by heterozygosity, and could not be effectively removed (Ramu et al. 2017). There were 4 LD blocks, but each block contained very few SNPs. The LD decayed very fast with physical distance, whereas the LD of SNP loci associated with DTCs was very strong. This indicated that excellent alleles might exist in the natural variations of *MeMYB26*. For significantly associated alleles with universal drought resistance-related traits, it was inferred that *MeMYB26* may play a role in a basic signal regulation process involved in drought resistance. In addition, a total of 7 functional SNPs could be used as markers in molecular-assisted selection (MAS) of cassava drought tolerance breeding with biomass storage.

The commonness of the module to which *MeMYB26* belongs could also be used to draw conclusions about its potential functions. Coexpression networks showed that the expression of the green module, in which *MeMYB26* was clustered, was obviously a common pattern in the roots of cassava cv. Arg7. The main functions of the green module are transcriptional regulation and catalytic activity. The major pathways of the green module were to regulate gene expression, phenylalanine biosynthesis and secondary metabolite biosynthesis by acting as transcription factors, which play an active role in plant resistance to drought stress (Ripoll et al. 2014, Takahashi et al. 2018, Liu et al. 2019). In addition, the cyanoamino acid metabolism pathway, plant hormone signal transduction and starch sugar metabolism pathway played an important role in drought resistance and yield potential (Li et al. 2017b, Ku et al. 2018, Prathap et al. 2019). The top 30 genes related to *MeMYB26*, including MYB transcription factors, starch metabolism pathway, antioxidant metabolism and so on, significantly responded to cassava drought stress signals. The regularity of the expression patterns of these
30 genes and \textit{MeMYB26} was more consistent and the same as that of the green module. Furthermore, this expression pattern was also very similar to the trend of starch metabolism genes, which were significantly down-regulated in rice leaves but had a significantly up-regulated expression for starch accumulation in the sink under drought stress (Prathap et al. 2019). The combination of coexpression network and candidate association analysis indicated that \textit{MeMYB26} is an important candidate gene for cassava drought resistance with maintenance or increase of biomass storage in cassava.

In conclusion, \textit{MeMYB26} is a typical drought-responsive MYB transcription factor, which may play a positive role in drought resistance and was significantly associated with multiple drought resistance phenotypes. It might play an important role in the expression of drought resistance and cassava storage roots, but more experimental evidence is needed to verify its functions. Excellent alleles might be available in natural variations, providing important evidence and direction for future research for functional verification and breeding for cassava drought resistance and high yields. Finally, our results also identified a series of potential polymorphism loci for the development of functional molecular markers.

**ACKNOWLEDGMENTS**

This research was supported by grants from the National Key R&D Program of China (2019YFD1000500) and the National Natural Science Foundation of China (Grant no. 31701509).

**REFERENCES**


MeMYB26, a drought-responsive transcription factor in cassava (Manihot esculenta Crantz)


