Genome-Wide Analysis of R2R3-MYB Transcription Factors Family in The Autopolyploid Saccharum Spontaneum: An Exploration of Dominance Expression and Stress Response

Yuan Yuan
Guangxi University
Xiping Yang
Guangxi University
Mengfang Feng
Guangxi University
Hongyan Ding
Guangxi University
Khan Muhammad Tahir
Nuclear Institute of Agriculture
Jisen Zhang
Fujian Agriculture and Forestry University
Muqing Zhang (zmuqing@163.com)
Guangxi University

Research Article

Keywords: MYB, Sugarcane, Expression analysis of stress, Allelic diversity

Posted Date: February 16th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-199103/v1

Version of Record: A version of this preprint was published at BMC Genomics on August 18th, 2021. See the published version at https://doi.org/10.1186/s12864-021-07689-w.

Abstract

Background: Sugarcane (*Saccharum*) is the most important sugar crop in the world. As one of the most enriched transcription factor families in plants, MYB genes display a great potential to contribute to sugarcane improvement by trait modification. We have identified the sugarcane MYB gene family at a whole-genome level through systematic evolution analyses and expression profiling. R2R3-MYB is a large subfamily involved in many plant-specific processes.

Results: A total of 202 R2R3-MYB genes (356 alleles) were identified in the polyploid *Saccharum spontaneum* genome and classified into 15 subgroups by phylogenetic analysis. The sugarcane MYB family had more members by a comparative analysis in sorghum and significant advantages among most plants, especially grasses. Collinearity analysis revealed that 70% of the SsR2R3-MYB genes had experienced duplication events, logically suggesting the contributors to the MYB gene family expansion. Functional characterization was performed to identify 56 SsR2R3-MYB genes involved in various plant bioprocesses with expression profiling analysis on 60 RNA-seq databases. We identified 22 MYB genes specifically expressed in the stem, of which *MYB43*, *MYB53*, *MYB65*, *MYB78*, and *MYB99* were validated by qPCR. Allelic expression dominance in the stem was more significant than that in the leaf, implying the differential expression of alleles may be responsible for the high expression of MYB in the stem. *MYB169*, *MYB181*, *MYB192* were identified as candidate C_4 photosynthetic regulators by C_4 expression pattern and robust circadian oscillations. Furthermore, stress expression analysis showed that *MYB*36, *MYB*48, *MYB*54, *MYB*61 actively responded to drought treatment; 19 and 10 MYB genes were involved in response to the sugarcane pokkah boeng and mosaic disease, respectively.

Conclusions: A Genome-wide expression analysis demonstrated that *SsMYB* genes were involved in stem development and stress response. This study largely contributed to understanding the extent to which MYB transcription factors investigate regulatory mechanisms and functional divergence in sugarcane.

Background

Modern cultivated sugarcane (*Saccharum spp.*) is the major source of sugar for the world. It is the topmost crop concerning total biomass production and is listed among the ten most valuable crops [1]. Sugarcane, having a complex genetic background resulting from polyploid interspecific hybrids, was first domesticated approximately10,000 years ago in New Guinea. *Saccharum spontaneum* contributes about 10%-15% genome to the modern sugarcane cultivars, endowing the characteristics such as disease resistance and ratooning capacity [2]. The genome of haploid *S. spontaneum* has been assembled to the chromosome level and used as the reference genome of sugarcane [3]. Because of the development of multiple transcriptome models in recent times, including those for different tissues, developmental stages, and under various stress treatments, huge RNA-seq data has become available and provides detailed insights and rich resources for studying gene functions of sugarcane.

Transcription factors recognize specific DNA motifs in upstream regions of the genes to regulate their expression. *MYB* genes constitute one of the largest families of plant transcription factors and characteristically possess highly conserved Myb DNA-binding domains, forming a helix-turn-helix structure of about 52 amino acids [4]. *MYB* genes can be divided into four categories, including *MYB*-related, *R2R3-MYB*, *R1R2R3-MYB*, and atypical *MYB*, depending on the number of adjacent *MYB* repeats (R). Proteins with a single or a partial MYB repeat, generally located at either ends or middle of the peptide chain, are MYB-related.

MYB-related proteins include important telomere binding proteins in maintaining the integrity of the chromosome structure [5]. Moreover, they also play an important role in regulating gene transcription, *e.g.*, the GARP family of plant Myb-related DNA binding motifs is involved in organ polarity in Arabidopsis[6]. Further, *CIRCADIAN CLOCK ASSOCIATED*1 (*CCA*1) and *LATE ELONGATED HYPOCOTYL* (*LHY*) genes regulate the plant circadian clock [7]. A small number of members of *R1R2R3-MYB* genes are found in higher plants. Interestingly, plant *R1R2R3-MYB* genes share a similar function of regulating the cell cycle control with the animals [8]. *3R-MYB* has also been involved in cell differentiation [9] and plant stress tolerance [10].

Atypical MYB proteins contain four or more adjacent MYB repeats (R). These proteins have been found to encode in a few plants, e.g., *Arabidopsis thaliana, Oryza sativa*, *Vitis vinifera*, *Glycine max*, *Physcomitrella patens* (data sources displayed in Materials and Methods 2.1), as shown in Figure 1. Only a few reports have been published about atypical MYB proteins by now, and the role of these proteins in the plant bioprocesses is largely unknown. MYB transcription factors binding specific DNA sequence (CAACG/TG) result from domain structure that is formed by two closely packed amino acid sequence repeats(R) [11]. When the MYB gene contains at least two MYB repeats (R), it has transcription factor characteristics and specifically recognizes the DNA motifs to regulate the gene transcription. R2R3-MYB proteins are the largest subfamily of MYB transcription factors in plants, as well as in *S. spontaneum* (Figure 1). R2R3-MYB is characterized by two MYB repeats and the presence of a single amino acid (Leu) in the first (R2) repeat [12]. R2R3-MYB has two MYB repeats and a single amino acid (Leu) inserted in the first (R2) repeat. The R2R3-MYB family's expansion originated from the R1R2R3-MYB gene ancestor when losing the R1 repeat sequences during evolution [13] and benefiting from gene duplication events [14].

MYB genes are widely involved in plant-specific processes, such as differentiation [15], hormone response [16], secondary metabolism [17], environmental stress tolerance [18], and diseases resistance [19][20]. At least four MYB genes are involved in lignin biosynthesis in *Arabidopsis* by activating key regulator genes related to secondary cell wall formation [21-23]. Under environmental stress, MYB genes have been reported to function in response to adverse stress in *Arabidopsis*. Moreover, *AtMYB2* and *AtMYB96* function as transcriptional activators in ABA-inducible gene expression under drought stress [24]. *AtMYB96* mediates abscisic acid signaling, induces pathogen resistance response by promoting salicylic acid biosynthesis, and provides drought tolerance *via* controlling the cuticular wax biosynthesis [20, 25].

This study focused on the R2R3-MYB gene family in the *S. spontaneum* published sugarcane genome. We provided a detailed overview of phylogenetic relationship, gene structure, regulatory elements, expression profiles, allelic evolution, and functional characterization based on abundant transcriptome data. Taken together, our study systematically explored the evolutionary dynamics and functional diversification of SsR2R3-MYB genes and could hence facilitate future research on sugarcane MYB transcription factors.

Materials And Methods

Obtainment of MYB genes

The autopolyploid sugarcane *Saccharum spontaneum* L. genome was published in 2018 and is available online (http://www.life.illinois.edu/ming/downloads/Spontaneum_genome/). The Hidden Markov Model (HMM) profile of the MYB DNA-binding domain (PF00249) downloaded from Pfam database (http://pfam.xfam.org/) [62] was used to search protein sequences containing MYB domain by hmmsearch program (HMM3.0) [63]. Then, putative MYB proteins were further screened through the NCBI-CDD database to investigate the former protein sequences and delete the proteins with incomplete domains. SbR2R3-MYB genes were obtained by performing the same sugarcane method without publicly available data for sorghum MYB genes. Data for sorghum protein sequences (the newest version of Sbicolor_454_v3.1.1.) were downloaded from the plant genome website Phytozome (https://phytozome.jgi.doe.gov/). Finally, we identified 418 (695) *SsMYB* genes and 252 *SbMYB* genes (Table S1), including 202 *SsR2R3-MYB* genes and 125 *SbR2R3-MYB* genes, belonging to haplotype genes. A plant phylogeny tree was constructed by the TimeTree Database (http://www.timetree.org)[64]. The distribution of MYB family genes in 19 plant species were demonstrated on the previously published reports: *Ostreococcus lucimarinus*, *Volvox carteri* and *Chlamydomonas reinhardtii* from PlantTFDB (http://planttfdb.cbi.pku.edu.cn/), and a public plant transcription factor database [65], including *Physcomitrella patens* [66], *Oryza sativa* [29], *Brachypodium distachyon* [28], *Zea mays* [45], *Ananas comosus* [33], *Vitis vinifera* [68], *Arabidopsis thaliana*[29], *Brassica napus* [69], *Glycine max* [70], *Medicago truncatula* [71], *Pyrus bretschneideri* [72], *Rosa chinensis* [73], *Populus trichocarpa* [74], *Beta vulgaris* [75], *Solanum tuberosum* [76], and *Solanum lycopersicum* [77].

Phylogenetic analysis

To generate the phylogenetic trees of MYB transcription factor family genes, multiple protein sequence alignment was performed through ClustalW [78] program using the reported 88 rice MYB proteins [29], and further phylogenetic trees were constructed *via* the neighbor-joining (NJ) method using software MEGA7.0 [79]. The consistency of the phylogenetic estimates was evaluated through several models as well as pairwise deletion treatment. NJ based phylogenetic tree of sugarcane and sorghum was performed as the same method.

Naming R2R3-MYB genes and gene structure

Because of the autopolyploid nature of sugarcane (*S. spontaneum*), the identified SsR2R3-MYB genes partly possessed several alleles. The representative gene models for different alleles were screened by comparing the phylogenetic relationship and protein identity with sorghum homology protein and paralogs. Tandem replication genes and paralogs were regarded as new, which gene IDs were followed by P and T, respectively. The 202 representative SsR2R3-MYB genes were named from SsMYB1 to SsMYB202 according to their physical position on the chromosomes. Subsequently, allele names were supplemented with numbers (e.g., The Sspon.01G0002470-1A gene located at the top of chromosome 1A is MYB1-1, and Sspon.01G0002470-2D is named as *MYB1-2*). In general, *MYB1-1* as a representative gene model was directly regarded as *MYB1*. The naming method of sorghum MYB genes was also treated like that of *S. spontaneum.*

SsR2R3-MYB genes and CDS sequences come from the newest version of Sspon.v20190103. The domain location was derived from the previous hmmsearch results. Gene structures were displayed using the Gene Structure Display Server (GSDS2.0) [80], consisting of the CDS region, intron region, and MYB domain. Each gene structure was arranged according to the phylogenetic location.

Collinearity analysis

Utilizing MCScanX analysis [78], collinearity relationships of *SsR2R3-MYB* genes and classifier program were used to sort gene duplication types. The identified collinear gene pairs were mapped to their respective locus in the *S. spontaneum* genome in a circular diagram using Circos 0.69 [81].

Regulatory element of upstream sequences

The 2000 bp upstream sequences were extracted from *SsR2R3-MYB* genes to the PlantCARE website, plant promoter, and cis-element database [82]. Then, we used them to predict regulatory motifs and estimate potentially related functions.

Abundant RNA-seq data showing gene expression

To analyze *SsR2R3-MYB* gene expression profiles thoroughly, 60 RNA-seq data were conducted to decipher their expressions from our lab and cooperative labs. Tissue and development transcriptome contained RNA-seq data of 16 samples, including leaf, stem, three different development stages *viz*. seeding (35-day-old), pre-maturity (9-month-old), and maturity (12-month-old) stages in *S.spontaneum* [83]. The leaf development transcriptome was derived from the second leaf alone, the ligule on 11-day-old seedlings; 15 cm leaves were selected and cut into 15 pieces with one segment per centimeter [84]. Mature leaves corresponding to ligule in *S.spontaneum*, over 12-month-old, were selected to supply

circadian rhythm transcriptome using 19-time points, *i.e.*, 2 hours apart from 6:00 am to the second day 4:00 am, and 4 hours apart from 6:00 am to the third day 6:00 am.

RNA-seq were extracted from the drought-treatment sugarcane of FN95-1702, a new sugarcane variety for both sugar and energy, bred by Fujian Agriculture and Forestry University. Sugarcane grown to 4-5 leaves was subjected to the natural drought stress treatment in the greenhouse. The mild drought was characterized by soil relative water content of about 55%~60% after six days, and severe drought by 25%~30% after twelve days. After a severe drought, rehydration was done, and relative water content was kept around 75%□85%, and then leave samples were retaken (5 days later). The *R2R3-MYB* gene expression profiles were obtained by Blast mapping to express data with transcripts of unreferenced genomes. RNA-seq for pokkah boeng disease were extracted from hybrid sugarcane ZZ1, which is highly resistant to smut disease but highly susceptible to pokkah boeng disease. According to the severity of the diseased leaves, pokkah boeng disease was divided into five grades from 0-5. The mildly diseased leaves (1 or 2 grades) and severely diseased leaves (4 or 5 grades) were selected for analysis, while healthy leaves were used as control (CK). Three samples were extracted for RNA-seq for sugarcane mosaic disease transcriptome analysis. For the infection experiment, sugarcane grown through virus-free tissue culture was used, and then leaves corresponding to ligule were collected one month after the infection, while the control plants were not infected. An expression ratio >2 (adjusted p-value<0.05) was considered statistically significant for evaluating differentially expressed genes.

Quantitative RT-PCR

8. *spontaneum* was planted in Multifunctional Specimen Garden, Institute of Agriculture, Guangxi University. The stem-3 at the third internode and mature leaves were collected for comparing the difference of relative expression between stem and leaf. Pro-stem is short for prophase stem, in which the samples were taken from the stem precursor tissue wrapped in the leaf sheath and is located on the upper part of the stem with obvious stem nodes. Combining with stem-3, stem-6, stem-9, and mature leaves were used to verify the expression during the prophase of stem formation. The extractions of samples total RNA was carried out using TRIZOL reagent (Takara), employing the corresponding protocol. The qualified RNA was reverse transcribed to produce cDNA using PrimeScript^{TM} RT reagent Kit with gDNA Eraser reagent (Takara, Japan). Primers were designed by qPCR-PrimerQuest Tool, and qPCR primers were shown in Table S8. Glyceraldehyde-3-phosphate dehydrogenase gene (GAPDH) was selected as a reference gene [85]. The real-time qPCR with three biological replications were performed with SYBR green on Roche Lightcyler® 480 instrument using 2×TB Green Mix (Takara). The reaction profile was as follows: 95°C for 30 s, followed by 40 cycles of 95°C for 10 s, 60°C for 30s, and 95°C for 10s. The relative expression levels were calculated by the $2^{-\triangle\triangle}$CT method.

Results

Genome-wide identification of R2R3-MYB genes and classification in S. spontaneum genome

Based on the functional annotation of the *Myb_DNA-binding* domain (PF00249), a total of 418 *MYB* genes (695 alleles) were identified in *the S. spontaneum* genome by combining the HMMER program and NCBI-CDD database (Figure 1). The SsMYB gene family was classified into four distinct subfamilies, including 207 MYB-related (329 alleles), 202 R2R3-MYB (356 alleles), 3 R1R2R3-MYB (3 alleles), and 5 Atypical MYB (7 alleles) genes (detailed data presented in supplementary Table S1). Total 122 SbMYB-related, 125 SbR2R3-MYB, 3 SbR1R2R3-MYB, and 2 Atypical MYB genes were also identified to increase the understanding of SsR2R3-MYB genes (Table S3).

To analyze the plant MYB genes thoroughly, twenty species in 11 lineages were screened to construct a plant phylogenetic tree with *S. spontaneum*, including Green algae, Bryophyta, Gramineae, Cruciferous, Leguminous, Rosaceae, Solanaceae, and others. The tree topology reflected the phylogenetic relationship of these species and divergence time (Figure 1). Plant phylogeny showed that the higher plants possessed more MYB genes than the lower plants, such as green algae (e.g., *Ostreococcus lucimarinus, Volvox carteri,* and *Chlamydomonas reinhardtii*). A significant expansion of MYB genes was observed after the Cambrian (about 540~480MYA), demonstrating an explosive biological diversification episode near the early period [26]. Most of the phylogenetic nodes of plant species were observed in the Cretaceous, a geological period when a typical global warming climate contributed to the diversity of the terrestrial species [27]. Compared with the other four kinds of grasses, *S. spontaneum* had one of the largest MYB genes as predicted by PlantTFDB. One reason is the tetraploid nature of the autopolyploid *S. spontaneum* (mainly octoploid). However, when corrected for ploidy level, the number of SsR2R3-MYB genes in *S. spontaneum* was still significantly higher than most of the species, including *Arabidopsis* and other grass species. From green algae to bryophyte and land plants, the number of MYB genes increased. The phylogenetic analysis of the plant species using the number of MYB genes indicated the extending of MYB genes from lower to higher plants, consistent with previous reports [28].

A neighbor-joining phylogenetic tree of *R2R3-MYB* genes from *O. sativa* and *S. spontaneum* showed that the sugarcane genome contained 15 subgroups (G1-G15) (Figure 2, Table S2) with *OsR2R3-MYB* genes [29]. Sugarcane and rice diverged in the Paleogene (67-26MYA) (Figure 1); the short divergence time indicated relative conservatism of the ortholog genes. As expected, two species of R2R3-MYB genes were evenly distributed in the tree, and most genes in rice clustered with sugarcane, except for *LOC_Os03g14100*. However, the number of genes in each clade varied greatly; for instance, the biggest group, G4, contained 26 genes while the group G13 comprised just one SsMYB gene *Sspon.02G0044740-1B*. Twenty *SsMYB* genes from three unique subgroups, G7, G10, and G15, did not contain rice genes, indicating the species' genetic divergence. Besides, the clusters depicted that the sugarcane MYB family exhibited a greater number of genes than that in rice, showing a significant expansion of the SsMYB family.

Analysis of genomic location, gene structure, and regulatory elements

A total of 202 *SsR2R3-MYB* genes were named in turn according to their physical position on the chromosomes. *MYB* genes were distributed throughout all 32 chromosomes (Figure 3C); the autopolyploid *S. spontaneum* genome comprised of 8 homologous groups of 4 members each [3]. The chromosome distribution map showed that the location of the *MYB* genes was not evenly distributed. Most of the *SsMYB* genes were located on Chr3A and Chr7A, encompassing 19 and 16 genes, respectively. About 11 enrichment clusters, tiny fragments on genomic regions containing 3 *MYB* genes, were detected, and half of these genes contained MYB-binding sites (MBS) depicting potential interaction among each cluster. However, some chromosomes only contained a few *MYB* genes. For instance, five chromosomes, including Chr2C, Chr2D, Chr6C, Chr8B, and Chr8D, had only one *MYB* gene.

S. bicolor is one of the closest lineages of sugarcane, possessing relatively perfect genome data [30-31]. Total 125 *SbR2R3-MYB* genes were identified from the available sorghum genome using a similar method (Figure 1, Table S3). The diversity of the gene structure might be a shred of evidence regarding the evolution of gene families. The phylogenetically and gene structure analysis were performed by the Neighbor-Joining method using diverse gene information (Figure 3A and Figure S1). The distribution of the tree branches was basically consistent with the structural features of the genes. In many clusters, various sorghum genes were clustered with highly similar *SsR2R3-MYB* genes, *e.g.*, *SbMYB92* clustered with *SsMYB149* and *SsMYB156* while *SbMYB27* was clustered with *SsMYB30* and *SsMYB44*. These results sharpened our understanding of the evolution of gene events during sugarcane polyploidization. A total of 19 *SsR2R3-MYB* genes did not show the presence of intron, including *SsMYB154, SsMYB188, SsMYB194, SsMYB170*, *SsMYB122*, *SsMYB182*, and *SsMYB189*. Many MYB genes demonstrated a domain with a cross-intron structure.

Cis-elements in promoter regions play an essential role in controlling transcription and expression, and hence they can deepen the understanding of the regulatory function of MYB genes. Total 2000 bp upstream of transcription initiation site (ATG) was regarded as MYB gene promoters and submitted to the PlantCARE for predicting the motifs. Various motifs from 202 *SsR2R3-MYB* gene promoters were involved in various plant bioprocesses (Figure 3B). These diversified cis-regulatory elements could be divided into four main categories in terms of function: stress response, hormone response, light response, and plant growth and metabolism. A high percentage of *MYB* genes in the anaerobic induction (92%) and drought elements (58.9%) indicated that the MYB genes were more likely to function under these stresses. Moreover, a notable gene, *MYB88,* was found to have 10 LTR motifs, which is a cis-acting element involved in low-temperature responsiveness. The significantly enriched LTR elements (5'-CCG AAA-3') suggested that the *MYB88* gene might be involved in plant metabolic response to cold stress. Many of the *MYB* genes regulate the plant hormone response, especially methyl jasmonate (MeJA) and abscisic acid (ABA) responsiveness. A total of 75 genes promoters enriched regulatory elements TGACG-motif (5'-TGACG-3') and CGTCA-motif (5'-CGTCA-3') involved in MeJA-responsiveness, while 38 gene promoters

enriched regulatory elements ABRE involved in abscisic acid responsiveness. These MYB genes were predicted to regulate MeJA and ABA signaling in plants and function in plant defense and leaf abscission. Furthermore, more than thirty light response-related elements were predicted; for instance, conservative light element G-box was widely present in the upstream sequence of genes. Several regulatory elements were also associated with other functions in plant growth and development and regulation of seed growth and meristem development. Genes involved in seed-specific regulation contained the same RY-element (5'-CATGCATG-3'), and the elements involved in meristem expression demonstrated CAT-box (5'-GCC ACT-3') and NON-box (5'-AGATCGACG-3') in promoter regions. Finally, 119 genes were detected to be scattered on MYB binding sites, and 49 genes showed more than one binding site, suggesting that these genes probably interacted with other *MYB* genes. Four MYB binding elements were found in 202 *SsR2R3-MYB* promoters, including CCAAT-box (5'-CAACGG-3'), MBS (5'-CAACTG-3'), MBSI (5'-aaaAaaC(G/C)GTTA-3'), and MRE (5'-AACCTAA-3'). There was only one base difference between the former two elements, which accounted for 80% of the total MYB binding elements, suggesting the conservative nature of the sequence CAACG/TG of the MYB binding site. The autoregulation of plant transcription factors is common in one family, which showed sequence-specific interactions of the family [32-33]. Dof1 binds the PEPC1 promoter, but Dof2 blocks the transactivation of Dof1 [34]. Hence, these MYB genes with MYB binding site indicated the potential interaction effects.

Pervasive gene duplications

Duplication is a striking feature of the plant genome. Gene duplication in the *R2R3-MYB* gene family occurred during earlier evolution in land plants and contributed to its amplification [35]. We estimated gene duplication events in the *S. spontaneum* genome by collinearity analysis. A total of 274 collinearity pairs of SsR2R3-MYB genes were identified by Blastp for all protein sequences and evaluated with MCScanX, including 144 allelic pairs and 130 non-allelic pairs (Figure 4, Table S4). The collinearity relationships revealed that over half of the collinearity genes were concentrated in Chr 3 and Chr 7. The duplication events for MYB genes were predicted. Total 91 (25.84%) genes were tandem repeats, of which one-quarter of genes were located on Chr 7. Furthermore, 146 (39.88%) genes were identified to derive from segmental duplication events; 28.1% genes on Chr 2 and 33.5% on Chr 3 evolved from segmental duplication (Figure 4, Table S5). Segmental duplication played a critical role in the evolution of *S. spontaneumMYB* genes, similar as in the other species. Totally, 66.5% of the *R2R3-MYB* genes derived from gene duplication events, driving the MYB gene family expansion.

Temporal and spatial expression of the R2R3-MYB gene family

To characterize the expression profiles of MYB transcription factors, the temporally and spatially expression profiles of 202 *SsR2R3-MYB* genes were analyzed using a total of 50 RNA-seq data among

three transcriptome models, including tissue and developmental stages, leaf developmental gradient, and circadian rhythm. The expression heatmap showed that most of the MYB genes had low expression levels, but 71% of gene expression values were greater than 1 (FPKM) in at least one RNA-seq sample (Figure 5A, Table S6). Expression values of 15 MYB groups were presented in Table S6, and G14 genes seemed to be expressed greater than the other groups.

Five different expression patterns, i.e., C1-C5, were investigated on the tissue and developmental stages transcriptome by K-means (Figure 5B). A total of 85 *SsR2R3-MYB* genes belonging to the C1 and C3 clusters had low expression value, particularly C1 genes with almost no expression. On the contrary, C2 cluster genes displayed a relatively higher expression level in all developmental periods of leaf and stem. Interestingly, 37 genes of the C4 cluster were highly expressed in the stem during the seedling stage, the early stage of the stem formation (Figure S2A). Moreover, in the C5 cluster, 35 genes were highly expressed in the stem during each period, probably playing a regulatory role in the stem development (Figure S2B). The clusters indicated that the gene expression levels in the stem as a whole were significantly higher than those in the leaves, suggesting *SsR2R3-MYB* genes might play an important role in stem tissue. The relative expression of *SsMYB43*, *SsMYB52*, *SsMYB65*, *SsMYB78*, and *SsMYB99* were quantified by qPCR, verifying the results of RNA-seq data (Figure S3B); additionally, *SsMYB3, SsMYB15, and SsMYB157* predominant expressed in the early stage of stem formation depicted as prophase of the stem (Pro-stem), which was much higher than other stem nodes and leaf tissues (Figure S3A).

Sugarcane is a typical C_4 plant with high light use efficiency. The developmental gradient model of grass leaves could be used to study C_4 photosynthesis and its regulatory factors [36-38]. The regulatory role of *SsMYB* genes on C_4 photosynthesis was investigated on the developmental dynamical transcriptome of sugarcane leaf. As suggested by the C_4 photosynthetic development model, leaves are gradually differentiated for active photosynthesis [36]. A total of 27 differentially expressed *SsR2R3-MYB* genes were detected by the leaf developmental gradient alone, and most of the genes (class I) showed an expression profile, illustrating high value in the early stage of leaf development (Figure S4). Only three genes *SsMYB169*, *SsMYB181*, and *SsMYB192* in class II (Figure 5C, Figure S4), were identified as putative C_4-related transcription factors using the method that associated the co-expression pattern with the photosynthetic activity [37]. The expression increased with the development of C_4 photosynthesis and displayed the highest accumulation at the leaf mature zone. Interestingly, *SsMYB181* and *SsMYB192* shared one haplotype gene Sspon.07G0015250 with *SsMYB169*, as the tandem genes *SsMYB181* and *SsMYB192* derived from a gene duplication event. Circadian rhythm is another module to study photosynthesis, in which previously identified C_4-related regulators could also be verified. Nine *SsR2R3-MYB* genes showed a significant association of expression profile with the light-dark cycle (Figure 5D). These genes were divided into three types, containing three genes each type. The expression level of *SsMYB169*, *SsMYB159*, and *SsMYB153* tailed off during the daytime until around 6:00 pm, and then it gradually recovered till the next cycle. However, the expression profiles of *SsMYB48*, *SsMYB57*, and *SsMYB158* were just opposite to the expression pattern of the former, rising during the day and falling at night. Unexpected but reasonable, the preliminarily identified three C_4-related regulators, *SsMYB169*,

SsMYB181, and *SsMYB192*, also showed daylight expression pattern, hinting at their involvement in the regulation of circadian rhythm. This strong evidence showed that the three candidate MYB transcription factors were associated with C_4 photosynthesis.

MYB genes involved in response to drought and disease-induced stress.

The expression patterns of SsR2R3-MYB genes were evaluated under environmental stress (biotic and abiotic stress). Six *SsR2R3-MYB* genes with significantly differentially expressed genes (SDEGs) were responsive to drought induction (Figure 6A, Table S7). The transcripts of four genes, *SsMYB54*, *SsMYB36*, *SsMYB61*, and *SsMYB48*, rapidly accumulated after drought treatment, but their expression reduced to normal after rewatered. On the other hand, *SsMYB29* and *SsMYB166* showed the opposite trend. Further, the upstream regulatory elements of these six genes contained the MBS element (5'-CAACTG-3'), which was identified as MYB binding site involved in drought-inducibility. Half of these genes retained more than one MBS.

Pokkah boeng disease of sugarcane (PBD) is one of the most serious and devastating diseases caused by the *Fusarium* species complex, a fungal pathogen [39-40]. Nineteen19 different MYBs were associated with sugarcane PBD-infection and response (Figure 6B, Table S7). According to the gene expression trends, these genes could be divided into 14 genes with increased expression in defense response and the other 5 genes with reduced expression.

Sugarcane mosaic disease is a highly transmissible viral disease present in the cane-growing regions worldwide. Sugarcane mosaic virus (SCMV), belonging to the positive-sense single-stranded RNA viruses, reduces yields by damaging chloroplast and blocking photosynthesis [41-42]. After SCMV infection, 10 *SsR2R3-MYB* genes expression increased, and one gene, *MYB176*, decreased, suggesting that these MYB genes were involved in defense against SCMV infection (Figure 6C, Table S7). We discovered that these MYB genes were unique to sugarcane diseases, indicating the defense specificity of MYB genes for conferring the resistance of sugarcane pokkah boeng and mosaic disease.

Functional characterization

The potential function of *SsR2R3-MYB* genes was predicted on the identified genes with significantly specific expression. Fifty-six *SsMYB* genes were involved in seven plant bioprocesses (Figure 6D), of which six MYB genes only expressed during seeding stem and were possibly involved in stem differentiation and formation (Figure 6A). Three MYB genes were identified as candidate C_4 photosynthesis regulators, and nine genes responded in the circadian clock. Under diverse stresses, it was seen that six, nineteen, and ten SsR2R3-MYB genes responded to drought, pokkah boeng disease, and mosaic disease, respectively. Notably, *SsMYB51* and *SsMYB162* illustrated different expression changes

between two sugarcane diseases (pokkah boeng and mosaic disease). *SsMYB162* significantly accumulated, actively responding to the infection of two diseases (Table S7). However, *SsMYB51* showed a different expression pattern, negatively responding to pokkah boeng but positively answering SCMV. Moreover, 13 MYB genes had more than one putative function, indicating their role in diverse plant bioprocesses (Figure 6D).

Allelic expression dominance drove SsMYB to function in stem

The transcriptional levels of *R2R3-MYB* allelic genes were compared among different tissues and different developmental stages to investigate the transcriptome dynamics of *R2R3-MYB* genes in the allopolyploid across eight homoeologous chromosome pairs, of which 25% of the R2R3-MYB genes displayed allelic expression dominance in all samples. The number of expression dominant genes in the A, B, C, and D genomes was 84, 93, 82, and 79, respectively. Further, the allelic genes were compared in pairs, including A-B, A-C, A-D, B-C, B-D, and C-D (Figure 7a). Both the number of dominant genes in a single set of homoeologous chromosomes and the pairwise comparison of alleles showed no significant allelic dominance. Captivatingly, the number of dominant genes in the stem was more than that in the leaf in each allelic pair comparison. For four sets of homoeologous chromosomes, the percentages increase was 46.5%, 90.6%, 10.2%, 143.4%, corresponding to A, B, C, and D genomes, respectively, and the overall average rise was 64.5%. The transcriptional expression of allelic genes in the stem tissues showed significant differences among different alleles than those in the leaf tissues. Allelic expression dominant genes derived predominantly from stem transcriptomes. Selective pressure analysis demonstrated that Ka/Ks values of expression dominance *MYB* genes in the stem were higher than in the leaf, indicating tissue specificity (Figure 7B). In contrast with the neutral genes, the Ka/Ks values of differential expression genes were higher, while the subordinate genes exhibited top Ka/Ks values (Figure 7C).

Discussion

Gene duplication played an important role in gene expansion and functional diversification in the genetic revolution and phenotypic evolution [43]. A total of 202 *SsR2R3-MYB* genes were identified, the second-highest number of these genes among the 21 important plant species (displayed in Figure 1). The number of *R2R3-MYB* in sugarcane was far higher than the other members of the grass family. Nevertheless, sugarcane with octoploid nature had a higher number of MYB genes compared with the other species. The significant enrichment of *SsMYB* genes probably was affected by the two rounds of whole-genome duplication, including allopolyploidization followed by autopolyploidization [44], or two rounds of autopolyploidization [3]. In grasses, 11 (7.09%) genes in *O. sativa* were derived from tandem duplications, 26 (21.31%) in *B. distachyon*, and 24 (15%) in *Z. mays*, while 44 (28.38%) segmental gene pairs were derived from segmental duplications in *O. sativa*, 34 (45.08%) in *B. distachyon*, and 19 (24%)

in *Z. mays*, respectively [28-29, 45]. The duplication of genes distribution indicated that the MYB genes family expansion in *S. spontaneum* could be attributed to these duplication events.

The large *R2R3-MYB* gene family resulted from duplication events and autopolyploidization, demonstrated diverse functions in plant-specific processes. Some genes specially expressed in stem tissues were concentrated in the stem prophase, indicating that these *MYB* genes might regulate biological processes related to stem development. Stem morphogenesis is tightly associated with a secondary wall (the major mechanical tissue in the stems of grass species) formation and lignification [46]. Indeed, some MYB transcription factors are identified to be involved in sugarcane stem development. A previous study revealed that 7 ScMYB genes were correlated with lignin content and biosynthesis [47]. *ShMYB78* has been recognized as an activator of suberin biosynthesis and regulates suberin deposition [48]. In *Arabidopsis*, the asymmetric leaves1 (*as*1) gene encoding an MYB protein-mediated stem cell function and interacted with meristematic genes to regulate the shoot morphogenesis [49]. Furthermore, a group of rice and maize MYB genes (*OsMYB46* and *ZmMYB46*) activated the transcription of secondary cell wall biosynthesis and probably interacted with secondary wall-associated *NAC* genes[46]. The stem is the main storage organ of sugarcane. The role of these *SsMYB* genes in stem development might provide potential genetic resources for sugarcane breeding.

MYB genes also play an important role in leaf development in grasses. In maize, a group of MYB was recognized to be involved in leaf development, as indicated by expression gradients. *Myb-ZmRS2*, *MYB60,* and *MYB61* influence adaxial/abaxial polarity and stomata patterning [36, 50-51]. Moreover, some *ZmMYBs* are highly expressed in the transition zone, affecting secondary cell wall and lignin production [36]. The Class I containing 24 SsR2R3-MYB genes were also inferred to have similar functions. Furthermore, a few MYB genes were identified as C_4 regulators and correlated with C_4 photosynthetic cell type-specific gene expression. An *MYB* gene encoding GRMZM2G130149, which apparently regulates the transcription of phosphoenolpyruvate carboxykinase (PEPCK) in *Z. mays*, was categorized as a C_4 transcription factor [38]. Similarly, three putative C_4 transcription factors (*SsMYB169*, *SsMYB181*, and *SsMYB192*) identified in this study might play a potential role in forming photosynthetic organs and regulating the C_4 photosynthetic pathway. The *LATE ELONGATED HYPOCOTYL* (*LHY*) gene, encoding an *MYB* transcription factor, regulated circadian rhythms in *Arabidopsis*, and *MYB-LHY* was involved in circadian photoperiod [52]. In sugarcane, nine candidate MYB genes with high expression were associated with the circadian cycle and therefore performed similar functions. The genes associated with leaf development showed relatively low expression levels (FPKM<10) than those linked with the stem tissues, hinting at a stem-related expression dominance for most *SsMYB* genes.

Drought is one of the main factors restricting sugarcane growth and sugar production [53]. Identifying special and novel candidate genes is a great strategy to improve stress tolerance in sugarcane in this context. Certain MYB transcription factors, for instance, MYB_2 [54], SoMYB18 [55], ScMYB2S1 & 2 [56], and ScMYBAS1 [57], have been associated with the response to drought-induced stress in sugarcane. Six differentially expressed MYB genes were predicted in this study, helping understand the sugarcane drought tolerance mechanism.

Following pathogen invasions, plants turn on a series of plant defense mechanisms. MYB transcription factors play a facilitating role in disease resistance by regulating plant hormone metabolism and mediating systemic resistance [58]. *AtMYB30* [59], *AtMYB96* [20], and *SpMYB* [60] have already been reported to be involved in disease resistance. Several defense-related MYB candidate genes were identified against pokkah boeng disease and mosaic disease of sugarcane. Hence, MYB genes are a component of plant defense mechanisms against fungal and viral pathogens.

Polyploids are widely distributed among plants, and about 70% of the angiosperms have experienced one or more polyploidization events during their evolution. As a plant genome evolutionary force, polyploidization plays an essential role in speciation and genomic plasticity. In this study, homologous expression dominant genes Ka/Ks of autopolyploid sugarcane were higher than those of neutral genes, consistent with the allopolyploid of *B.juncea* [61]. Besides, MYB homologous expression of dominant genes was greater in number in stem tissues than those in leaves, and the Ka/Ks ratio was also higher, implying that MYB stems dominant genes intensified selection in sugarcane. Not surprisingly, the transcription level of MYB genes more significantly enriched in the stem. The transcriptional advantages of these MYB homologous expression dominant genes in stem tissues might provide new insights for facilitating polyploid crop breeding, including sugarcane.

Conclusions

It is the first time deciphering the phylogeny, gene structure, and expression of the MYB family in *S. spontaneum*. Genome-wide expression analysis demonstrated that *SsMYB* genes were involved in the stem development and stress response. The MYB genes might be engineered to adjust important sugarcane traits, and therefore, these genes would be a promising target for sugarcane genetic improvement.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in supplementary information files. Genomic data of sugarcane and sorghum for testing were obtained from the autopolyploid *Saccharum spontaneum* L. genome (http://www.life.illinois.edu/ming/downloads/Spontaneum_genome/) and *Sorghum bicolor* genome (https://phytozome-next.jgi.doe.gov/info/Sbicolor_v3_1_1). The domain

architecture of the MYB genes was downloaded from Pfam database (http://pfam.xfam.org/family/PF00249/hmm). The sequencing data of Sugarcane pokkah boeng disease: SRP127969 (https://www.ncbi.nlm.nih.gov/sra/SRP127969); and Sugarcane mosaic virus disease: SRR10058145, SRR10058144 in the GenBank database. RNA-seq of tissues and development stage, leaf segments and circadian rhythms were downloaded from sugarcane public database (http://sugarcane.zhangjisenlab.cn/sgd/html/mRNA.html).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by the National Natural Science Foundation of China (31660420), the Key Project of Science and Technology of Guangxi (AA17202042-7), and the earmarked fund for the Modern Agriculture Technology of China (CARS-170190) and the Innovation Project of Guangxi Graduate Education (YCBZ2020031). The funding body only provided the funds and didn't have any role in the study's design, sample collection, data analysis and interpretation, and writing of the manuscript.

Authors' contributions

YY and MQZ conceived and designed this study. XPY and JSZ made guidance during the experiment. YY performed the most analysis, including identifying the MYB family, phylogenetic analysis, collinearity analysis, and expression analysis. MFF assisted in allelic differential expression analysis and Ka/Ks calculation. HYD completed the qPCR experiment together. YY prepared the manuscript. MQZ, XPY and MTK advised on the revised manuscript, providing valuable comments. All authors reviewed and approved the final manuscript.

Acknowledgements

Not Applicable.

Author Details

[1] State Key Laboratory for Conservation and Utilization of Agro Bioresources; Guangxi Key Laboratory for Sugarcane Biology, Guangxi University, Nanning 530005, China. [2] Nuclear Institute of Agriculture (NIA), Tando Jam, 70060, Pakistan. [3] Fujian Agricultural and Forestry University, Fuzhou, 350002, China

References

1. FAO F. Food and Agriculture Organization of the United Nations - Statistic Divisionhttps://www.fao.org/faostat/en/#data/QC. 2019.

2. D'Hont A, Grivet L, Feldmann P, Rao S, Berding N, Glaszmann JC. Characterisation of the double genome structure of modern sugarcane cultivars (. MGG Mol Gen Genet. 1996;250:405.
3. Saccharum L, Zhang J, Zhang X, Tang H, Zhang Q, Hua X, et al. Allele-defined genome of the autopolyploid. Nat Genet. 2018;50 November:1565–73.
4. Stracke R, Werber M, Weisshaar B. The R2R3-MYB gene family in Arabidopsis thaliana. Curr Opin Plant Biol. 2001;4:447–56.
5. Bilaud T, Koering CE, Binet-Brasselet E, Ancelin K, Pollice A, Gasser SM, et al. The Telobox, a Myb-Related Telomeric DNA Binding Motif Found in Proteins from Yeast, Plants, and Human. Nucleic Acids Res. 1996;24:1294–303.
6. Kerstetter RA, Bollman K, Taylor RA, Bomblies K, Poethig RS. KANADI regulates organ polarity in Arabidopsis. Nature. 2001;411:706–9.
7. Yakir E, Hilman D, Kron I, Hassidim M, Melamed-Book N, Green RM. Posttranslational Regulation of CIRCADIAN CLOCK ASSOCIATED1 in the Circadian Oscillator of Arabidopsis. Plant Physiol. 2009;150:844–57.
8. Ito M, Araki S, Matsunaga S, Itoh T, Nishihama R, Machida Y, et al. G2/M-Phase-Specific Transcription during the Plant Cell Cycle Is Mediated by c-Myb-Like Transcription Factors. Plant Cell. 2001;13:1891–905.
9. Haga N, Kato K, Murase M, Araki S, Kubo M, Demura T, et al. R1R2R3-Myb proteins positively regulate cytokinesis through activation of KNOLLE transcription in Arabidopsis thaliana Development. 2007;134:1101 LP – 1110.
10. Dai X, Xu Y, Ma Q, Xu W, Wang T, Xue Y, et al. Overexpression of an R1R2R3 MYB gene, OsMYB3R-2, increases tolerance to freezing, drought, and salt stress in transgenic Arabidopsis. Plant Physiol. 2007;143:1739–51.
11. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in Arabidopsis. Trends Plant Sci. 2010;15:573–81.
12. Dias AP, Braun EL, McMullen MD, Grotewold E. Recently Duplicated Maize R2R3 Myb Genes Provide Evidence for Distinct Mechanisms of Evolutionary Divergence after Duplication. Plant Physiol. 2003;131:610 LP – 620.
13. Rosinski JA, Atchley WR. Molecular Evolution of the Myb Family of Transcription Factors: Evidence for Polyphyletic Origin. J Mol Evol. 1998;46:74–83.
14. Zhang P, Chopra S, Peterson T. A segmental gene duplication generated differentially expressed myb-homologous genes in maize. Plant Cell. 2000;12:2311–22.
15. Waites R, Selvadurai HRN, Oliver IR, Hudson A. The PHANTASTICA Gene Encodes an MYB Transcription Factor Involved in Growth and Dorsoventrality of Lateral Organs in Antirrhinum. Cell. 1998;93:779–89.
16. Jin H, Martin C. Multifunctionality and diversity within the plant MYB-gene family. Plant Mol Biol. 1999;41:577–85.

17. Bender J, Fink GR. A Myb homolog, ATR1, activates tryptophan gene expression in Arabidopsis. Proc Natl Acad Sci U S A. 1998;95:5655–60.
18. Urao T, Yamaguchi-Shinozaki K, Urao S, Shinozaki K. An Arabidopsis myb homolog is induced by dehydration stress, and its gene product binds to the conserved MYB recognition sequence. Plant Cell. 1993;5:1529 LP – 1539.
19. Vos M De, Denekamp M, Dicke M, Vuylsteke M. The Arabidopsis thaliana Transcription Factor AtMYB102 Functions in Defense Against The Insect Herbivore Pieris rapae ND SC RIB. Plant Signal Behav. 2006; December:305–11.
20. Seo PJ, Park C-M. MYB96-mediated abscisic acid signals induce pathogen resistance response by promoting salicylic acid biosynthesis in Arabidopsis. New Phytol. 2010;186:471–83.
21. McCarthy RL, Zhong R, Ye Z-H. MYB83 Is a Direct Target of SND1 and Acts Redundantly with MYB46 in the Regulation of Secondary Cell Wall Biosynthesis in Arabidopsis. Plant Cell Physiol. 2009;50:1950–64.
22. Zhou J, Lee C, Zhong R, Ye Z-H. MYB58 and MYB63 Are Transcriptional Activators of the Lignin Biosynthetic Pathway during Secondary Cell Wall Formation in Arabidopsis Plant Cell. 2009;21:248 LP – 266.
23. Zhong R, Ye Z-H. MYB46 and MYB83 Bind to the SMRE sites and Directly Activate a Suite of Transcription Factors and Secondary Wall Biosynthetic Genes. Plant Cell Physiol. 2012;53:368–80.
24. Abe H, Urao T, Ito T, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) Function as Transcriptional Activators in Abscisic Acid Signaling. Plant Cell. 2003;15:63 LP – 78.
25. Seo PJ, Lee SB, Suh MC, Park M-J, Go YS, Park C-M. The MYB96 Transcription Factor Regulates Cuticular Wax Biosynthesis under Drought Conditions in Arabidopsis Plant Cell. 2011;23:1138 LP – 1152.
26. Bowring SA, Grotzinger JP, Isachsen CE, Knoll AH, Pelechaty SM, Kolosov P. Calibrating rates of early Cambrian evolution. Science. 1993;261:1293–8.
27. McInerney FA, Wing SL. The Paleocene-Eocene Thermal Maximum: A Perturbation of Carbon Cycle, Climate, and Biosphere with Implications for the Future. Annu Rev Earth Planet Sci. 2011;39:489–516.
28. Chen S, Niu X, Guan Y, Li H. Genome-Wide Analysis and Expression Profiles of the MYB Genes in Brachypodium distachyon. Plant & cell Physiol. 2017;58:1777–88.
29. Katiyar A, Smita S, Lenka SK, Rajwanshi R, Chinnusamy V, Bansal KC. Genome-wide classification and expression analysis of MYB transcription factor families in rice and Arabidopsis. BMC Genomics. 2012;13:544.
30. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, et al. The Sorghum bicolor genome and the diversification of grasses. Nature. 2009;457:551–6.
31. McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, et al. The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of

genome organization. Plant J. 2018;93:338–54.

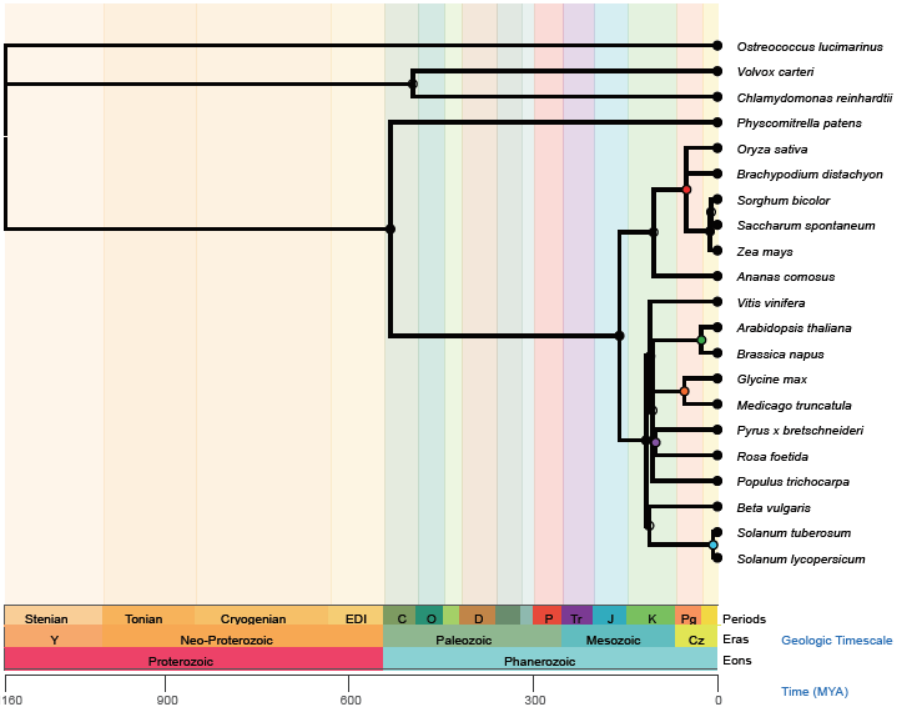
32. Kieffer M, Stern Y, Cook H, Clerici E, Maulbetsch C, Laux T, et al. Analysis of the transcription factor WUSCHEL and its functional homolog in Antirrhinum reveals a potential mechanism for their roles in meristem maintenance. Plant Cell. 2006;18:560–73.
33. Sharma P, Lin T, Grandellis C, Yu M, Hannapel DJ. The BEL1-like family of transcription factors in potato. J Exp Bot. 2014;65:709–23.
34. Yanagisawa S, Sheen J. Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. Plant Cell. 1998;10:75–89.
35. Rabinowicz PD, Braun EL, Wolfe AD, Bowen B, Grotewold E. Maize R2R3 Myb genes: Sequence analysis reveals amplification in the higher plants. Genetics. 1999;153:427–44.
36. Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, et al. The developmental dynamics of the maize leaf transcriptome. Nat Genet. 2010;42:1060–7.
37. Wang L, Czedik-Eysenberg A, Mertz RA, Si Y, Tohge T, Nunes-Nesi A, et al. Comparative analyses of C and C photosynthesis in developing leaves of maize and rice. Nat Biotechnol. 2014;32:1158–65.
38. Studer AJ, Schnable JC, Weissmann S, Kolbe AR, McKain MR, Shao Y, et al. The draft genome of the C3 panicoid grass species Dichanthelium oligosanthes. Genome Biol. 2016;17:223.
39. R Rott, RA Bailey, JC Comstock BC, and AS. A guide to sugarcane diseases. France: CIRAD; 2000.
40. Singh A, Chauhan SS, Singh A, Singh SB. Deterioration in sugarcane due to pokkah boeng disease. Sugar Tech. 2006;8:187–90.
41. Xu D-L, Park J-W, Mirkov TE, Zhou G-H. Viruses causing mosaic disease in sugarcane and their genetic diversity in southern China. Arch Virol. 2008;153:1031.
42. Chauhan RP, Rajakaruna P, Verchot J. Complete genome sequence of nine isolates of canna yellow streak virus reveals its relationship to the sugarcane mosaic virus (SCMV) subgroup of potyviruses. Arch Virol. 2015;160:837–44.
43. Flagel LE, Wendel JF. Gene duplication and evolutionary novelty in plants. New Phytol. 2009;183:557–64.
44. Kim C, Wang X, Lee T-H, Jakob K, Lee G-J, Paterson AH. Comparative Analysis of Miscanthus and Saccharum Reveals a Shared Whole-Genome Duplication but Different Evolutionary Fates. Plant Cell. 2014;26:2420–9.
45. Du H, Feng B-R, Yang S-S, Huang Y-B, Tang Y-X. The R2R3-MYB transcription factor gene family in maize. PLoS One. 2012;7:e37463.
46. Zhong R, Lee C, McCarthy RL, Reeves CK, Jones EG, Ye Z-H. Transcriptional activation of secondary wall biosynthesis by rice and maize NAC and MYB transcription factors. Plant & cell Physiol. 2011;52:1856–71.
47. Michael dos Santos Brito. Unraveling Two Ways of Lignin Biosynthesis in Sugarcane Culm: An Overview of Transcriptional Regulation. 2012; P0738.

48. Figueiredo R, Portilla Llerena JP, Kiyota E, Ferreira SS, Cardeli BR, de Souza SCR, et al. The sugarcane ShMYB78 transcription factor activates suberin biosynthesis in Nicotiana benthamiana. Plant Mol Biol. 2020;104:411–27.
49. Byrne ME, Barley R, Curtis M, Arroyo JM, Dunham M, Hudson A, et al. Asymmetric leaves1 mediates leaf patterning and stem cell function in Arabidopsis. Nature. 2000;408:967–71.
50. Husbands AY, Chitwood DH, Plavskin Y, Timmermans MCP. Signals and prepatterns: new insights into organ polarity in plants. Genes Dev. 2009;23:1986–97.
51. Liu T, Ohashi-Ito K, Bergmann DC. Orthologs of Arabidopsis thaliana stomatal bHLH genes and regulation of stomatal development in grasses. Development. 2009;136:2265–76.
52. Carré IA, Kim J-Y. MYB transcription factors in the Arabidopsis circadian clock. J Exp Bot. 2002;53:1551–7.
53. Prabu G, Kawar PG, Pagariya MC, Prasad DT. Identification of Water Deficit Stress Upregulated Genes in Sugarcane. Plant Mol Biol Report. 2011;29:291–304.
54. Silvana Creste Souza. Characterization of Transcription Factors Differentially Expressed Under Drought Conditions in Sugarcane (Saccharum spp). 2015; P0180.
55. Shingote PR, Kawar PG, Pagariya MC, Kuhikar RS, Thorat AS, Babu KH. SoMYB18, a sugarcane MYB transcription factor improves salt and dehydration tolerance in tobacco. Acta Physiol Plant. 2015;37:217.
56. Guo J, Ling H, Ma J, Chen Y, Su Y, Lin Q, et al. A sugarcane R2R3-MYB transcription factor gene is alternatively spliced during drought stress. Sci Rep. 2017;7:41922.
57. Fávero Peixoto-Junior R, Mara de Andrade L, Dos Santos Brito M, Macedo Nobile P, Palma Boer Martins A, Domingues Carlin S, et al. Overexpression of ScMYBAS1 alternative splicing transcripts differentially impacts biomass accumulation and drought tolerance in rice transgenic plants. PLoS One. 2018;13:e0207534.
58. Raffaele S, Rivas S, Roby D. An essential role for salicylic acid in AtMYB30-mediated control of the hypersensitive cell death program in Arabidopsis. FEBS Lett. 2006;580:3498–504.
59. Vailleau F, Daniel X, Tronchet M, Montillet J-L, Triantaphylidès C, Roby D. A R2R3-MYB gene, AtMYB30, acts as a positive regulator of the hypersensitive cell death program in plants in response to pathogen attack. Proc Natl Acad Sci U S A. 2002;99:10179–84.
60. Liu Z, Luan Y, Li J, Yin Y. Expression of a tomato MYB gene in transgenic tobacco increases resistance to Fusarium oxysporum and Botrytis cinerea. Eur J Plant Pathol. 2016;144:607–17.
61. Yang J, Liu D, Wang X, Ji C, Cheng F, Liu B, et al. The genome sequence of allopolyploid Brassica juncea and analysis of differential homolog gene expression influencing selection. Nat Genet. 2016;48:1225–32.
62. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44:D279-85.

63. Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. Bioinformatics. 2013;29:2487–9.
64. Hedges SB, Dudley J, Kumar S. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics. 2006;22:2971–2.
65. Guo A-Y, Chen X, Gao G, Zhang H, Zhu Q-H, Liu X-C, et al. PlantTFDB: a comprehensive plant transcription factor database. Nucleic Acids Res. 2008;36 Database issue: D966-9.
66. Pu X, Yang L, Liu L, Dong X, Chen S, Chen Z, et al. Genome-Wide Analysis of the MYB Transcription Factor Superfamily in Physcomitrella patens. Int J Mol Sci. 2020;21.
67. Liu C, Xie T, Chen C, Luan A, Long J, Li C, et al. Genome-wide organization and expression profiling of the R2R3-MYB transcription factor family in pineapple (Ananas comosus). BMC genomics. 2017;18:503.
68. Matus JT, Aquea F, Arce-Johnson P. Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across Vitis and Arabidopsis genomes. BMC Plant Biol. 2008;8:83.
69. Chen B, Niu F, Liu W-Z, Yang B, Zhang J, Ma J, et al. Identification, cloning and characterization of R2R3-MYB gene family in canola (Brassica napus L.) identify a novel member modulating ROS accumulation and hypersensitive-like cell death. DNA Res. 2016;23:101–14.
70. Du H, Yang S-S, Liang Z, Feng B-R, Liu L, Huang Y-B, et al. Genome-wide analysis of the MYB transcription factor superfamily in soybean. BMC Plant Biol. 2012;12:106.
71. ZHENG X, YI D, SHAO L, LI C. In silico genome-wide identification, phylogeny and expression analysis of the R2R3-MYB gene family in Medicago truncatula. J Integr Agric. 2017;16:1576–91.
72. Feng S, Xu Y, Yang L, Sun S, Wang D, Chen X. Genome-wide identification and characterization of R2R3-MYB transcription factors in pear. Sci Hortic (Amsterdam). 2015;197:176–82.
73. Han Y, Yu J, Zhao T, Cheng T, Wang J, Yang W, et al. Dissecting the Genome-Wide Evolution and Function of R2R3-MYB Transcription Factor Family in Rosa chinensis. Genes (Basel). 2019;10:823.
74. Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM. Expansion and Diversification of the Populus R2R3-MYB Family of Transcription Factors. Plant Physiol. 2009;149:981 LP – 993.
75. Stracke R, Holtgräwe D, Schneider J, Pucker B, Rosleff Sörensen T, Weisshaar B. Genome-wide identification and characterisation of R2R3-MYB genes in sugar beet (Beta vulgaris). BMC Plant Biol. 2014;14:249.
76. Sun W, Ma Z, Chen H, Liu M. MYB Gene Family in Potato (Solanum tuberosum L.): Genome-Wide Identification of Hormone-Responsive Reveals Their Potential Functions in Growth and Development. Int J Mol Sci. 2019;20.
77. Zhao P, Li Q, Li J, Wang L, Ren Z. Genome-wide identification and characterization of R2R3MYB family in Solanum lycopersicum. Mol Genet Genomics. 2014;289:1183–207.

78. Thompson JD, Gibson TJ, Higgins DG. Multiple Sequence Alignment Using ClustalW and ClustalX. Curr Protoc Bioinforma. 2003;00:2.3.1-2.3.22.
79. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33:1870–4.
80. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31:1296–7.
81. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45.
82. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 2002;30:325–7.
83. Chen Y, Zhang Q, Hu W, Zhang X, Wang L, Hua X, et al. Evolution and expression of the fructokinase gene family in Saccharum. BMC Genomics. 2017;18:197.
84. Li Z, Hua X, Zhong W, Yuan Y, Wang Y, Wang Z, et al. Genome-Wide Identification and Expression Profile Analysis of WRKY Family Genes in the Autopolyploid Saccharum spontaneum. Plant Cell Physiol. 2020;61:616–30.
85. Ling H, Wu Q, Guo J, Xu L, Que Y. Comprehensive selection of reference genes for gene expression normalization in sugarcane by real-time quantitative rt-PCR. PLoS One. 2014;9:e97469.

Figures



MYB-related	R2R3-MYB	3R-MYB	Atypical MYB genes	Total
30	5	1	-	36
32	11	0	-	43
31	7	0	-	38
62	50	3	1	116
62	88	4	1	185
34	85	3	-	122
122	125	3	2	252
207(329)	202(356)	3(3)	5(7)	417(695)
-	157	-	-	157
-	94	-	-	94
-	108	5	1	114
52	138	5	2	196
-	76	-	-	76
-	244	6	2	252
-	155	-	-	155
-	184	-	-	184
-	121	-	-	121
-	192	5	-	192
-	70	-	-	70
-	155	3	-	158
-	121	-	-	121

Figure 1

Phylogenetic tree of diverse species showing the number of MYB family. The phylogenetic tree reflects the evolutionary relationship and divergence time of various species in plants through the TimeTree database (Hedges et al., 2006; http://www.timetree.org). Linear scale Time MYA (millions of years ago) and Geologic Timescale are shown at the tree's bottom. These species contain green algae, Grasses (red node), Cruciferous (green node), Leguminosae (orange node), Rosaceae (purple node), Chenopodiaceae (blue node), and others altogether 11 lineages. MYB gene family can be divided into four subfamilies according to the number of Myb domain. The available information of the MYB gene family was obtained from the reported literature, showing the table. The short line represents undetermined. The MYB families were estimated by performing profile searches using a combination of the HMMER3 program (PF00249) and NCBI-CDD database for S. spontaneum and S. bicolor. Myb_DNA-binding domain (PF00249) was downloaded from Pfam (Finn et al., 2010). S. spontaneum MYB gene in brackets are identified throughout the genome and contain alleles. A single set of genes are shown in outsides for each subfamily. Different alleles of one gene may be divided into different subfamilies. Thus, to better classification, a single set of genes does not distinguish this condition that one gene is in different subfamilies. Geologic Periods: C(Cambrian), O(Ordovician), D(Devonian), P(Permian), Tr(Triassic), J(Jurassic), K(Cretaceous), Pg(Paleogene).

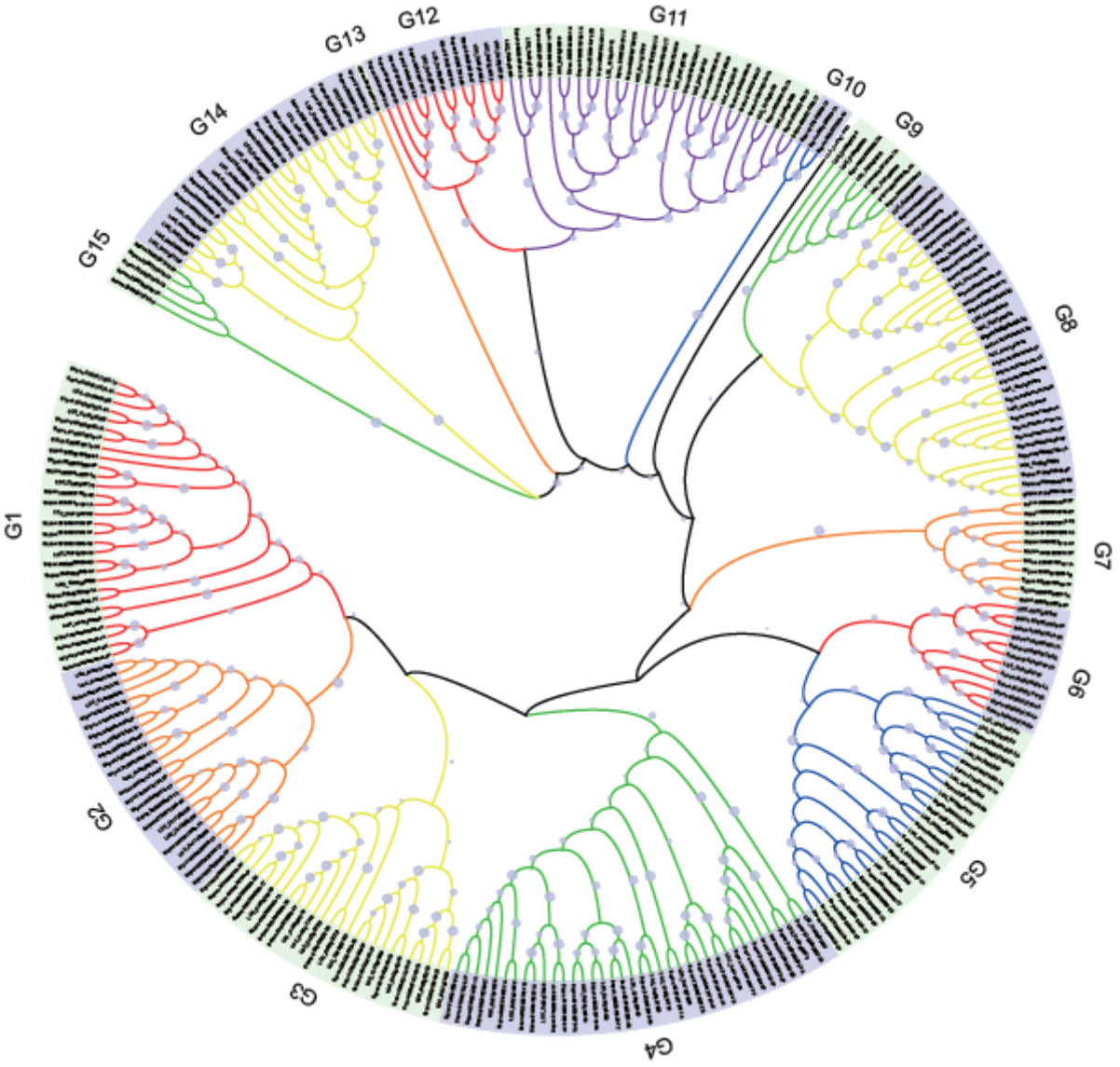


Figure 2

Phylogenetic relationships of R2R3-MYB subgroup members between S. spontaneum and O. sativa. A phylogenetic tree of R2R3-MYB proteins from Saccharum and rice was constructed using MEGA 7.0 with the Neighbor-Joining (NJ) method with the bootstrap test replicated 1000 times, the NO.of difference and Pairwise deletion. The sugarcane R2R3-MYB families are clustered into 18 subgroups (no containing the branches with rice gene only), named G1 to G15. The different clade of subgroups are marked as colorful lines, and their gene ID label was added with green and grey background in turn. The size of the grey point positively reflects phylogenetic bootstrap.

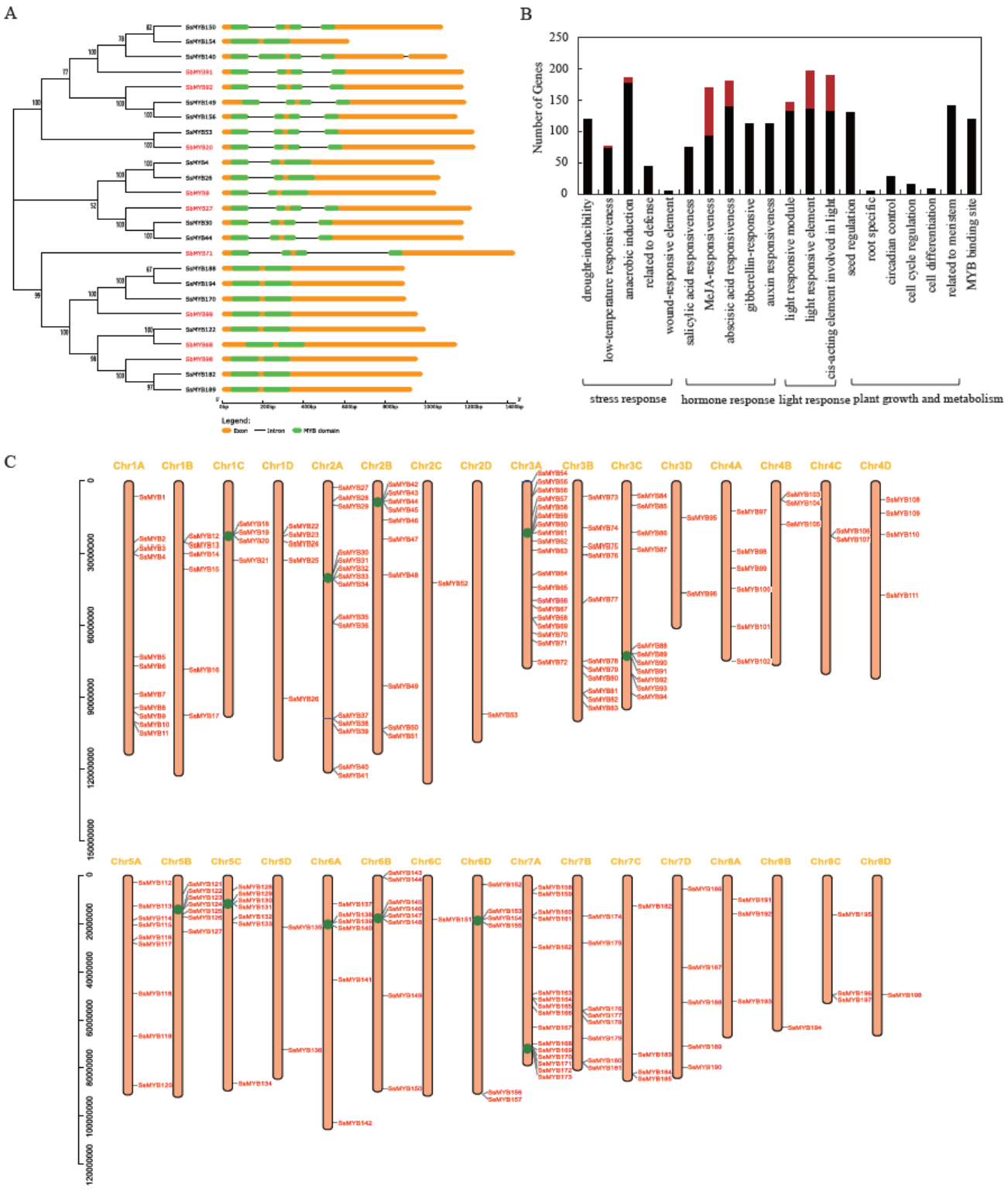


Figure 3

Collinearity relationships of SsR2R3-MYB genes on the S. spontaneum genome. SsR2R3-MYB collinear gene pairs were mapped to their respective locus in the S. spontaneum genome in a circular diagram. Genes located on the same chromosome (e.g., Chr1, Chr2, Chr3, Chr4, Chr5, Chr6, Chr7, Chr8) share one line, and the interchromosomal collinear genes pairs are linked with the former linear color.

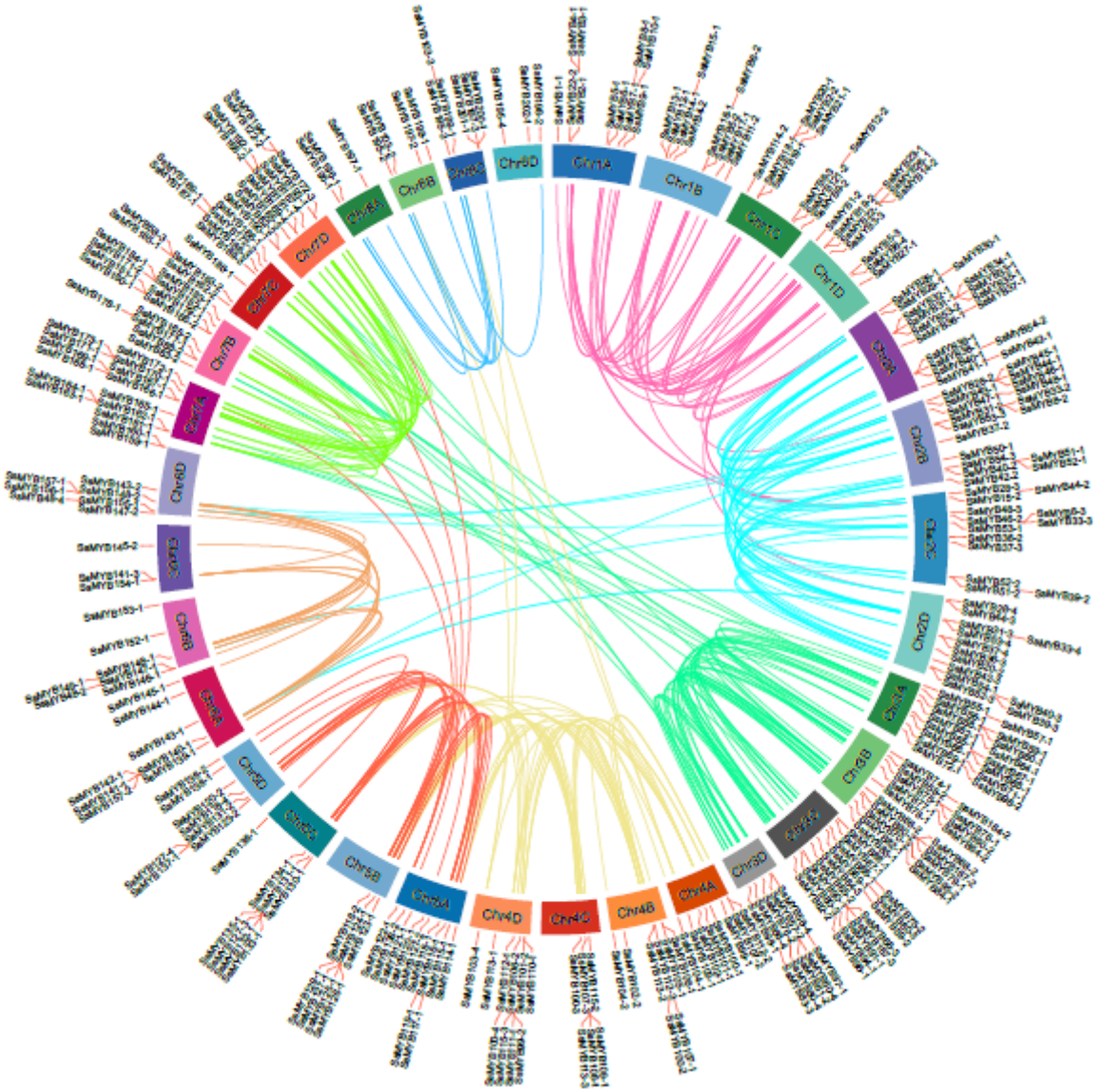


Figure 4

Structure, distribution, and regulatory elements of SsR2R3-MYB genes. (A) Comparison of gene structure between S. spontaneum and S. bicolor based on phylogenetic tree. The sequence alignment of SsR2R3-MYB and SbR2R3-MYB proteins were performed by ClustalX, and the Phylogenetic tree was constructed using MEGA 7.0 with Neighbor-Joining (NJ) method, 1,000 bootstrap replicates, Pairwise deletion, and Bootstrap values on the nodes. SsMYB gene names are marked black, and SbMYB gene names are marked red. Gene sequences were modified to start at the transcription initiation site (ATG), and gene structures were displayed using GSDS2.0 (http://gsds.cbi.pku.edu.cn/). The CDS sequence and intron are represented as fine lines and yellow cylinders, and the MYB domain were highlighted by green cylinders. One of the subgroups was showed in here when the estimated phylogenetic relationship of S. spontaneum and S. bicolor, and others were shown in Figure S1. (B) Cis-regulatory elements of SsR2R3-MYB gene promoters with diversified plant biological functions. The functions of the predicted cis-regulatory elements cover four main categories: stress response, hormone response, light response, plant growth, and metabolism. The x-axis shows divers plant biological functions, and the y-axis indicates the

number of a specific category of genes in that main category. The red rectangle represents the genes containing more than six elements involved in regulating a certain plant function. (C) Distribution of SsR2R3-MYB gene members in S. spontaneum genome. 202 SsR2R3-MYB genes were named according to their physical position on the chromosome and tagged in red font. Yellow font indicated chromosome name, and chromosome is represented as hollow cylinders with length scale (bp) on the left. The green spots displayed a gene enrichment cluster.

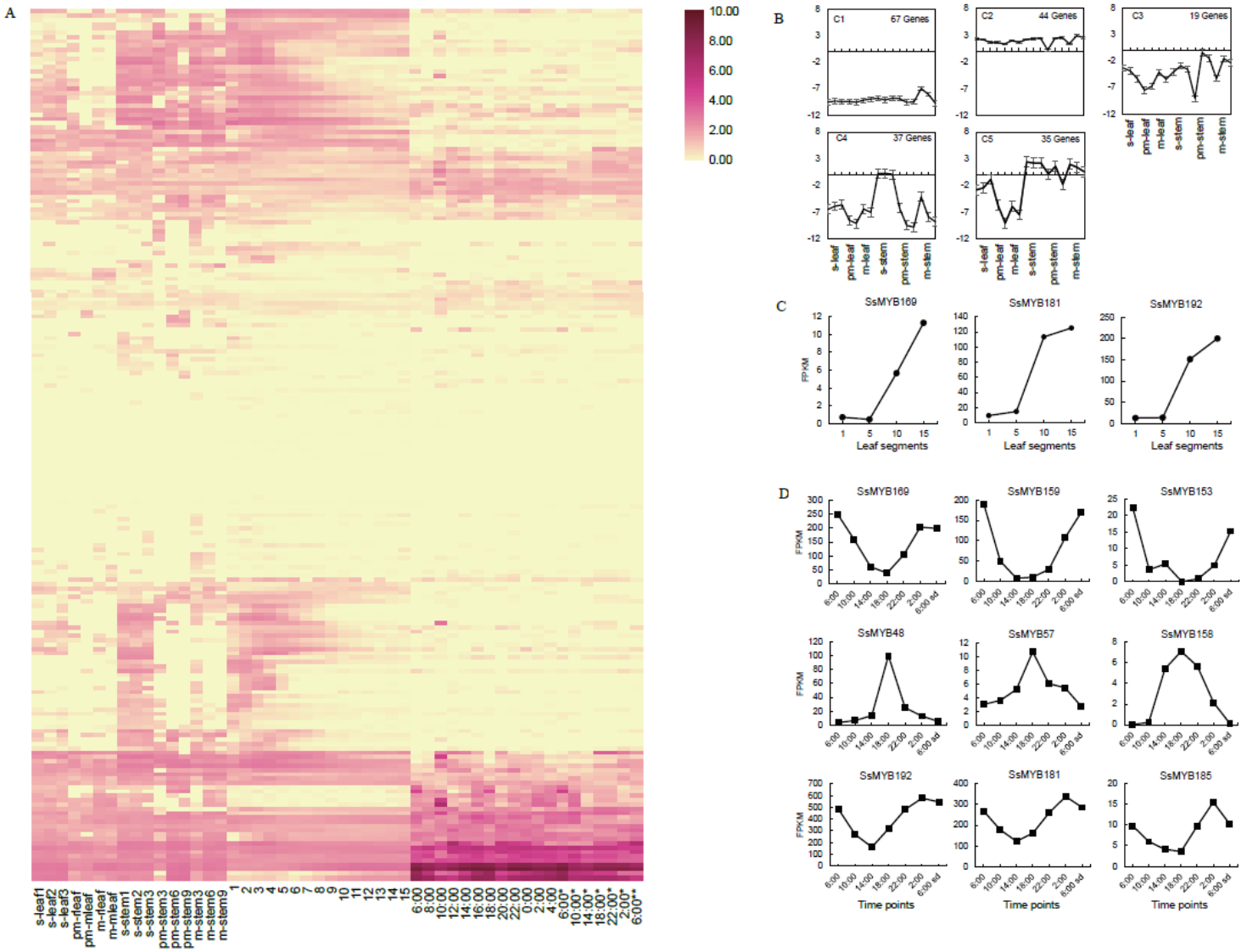


Figure 5

Temporal and spatial expression dynamics of SsR2R3-MYB genes. (A) A heatmap showed the expression profile of SsR2R3-MYB genes. Columns showed 18 subfamilies, and rows showed developmental stages and tissues, leaf developmental gradient, and circadian rhythm. (B) K-means clustering showing the expression profile of the developmental stages and tissue transcriptome. Five clusters were identified as C1-C5, error bars showing standard deviation. (C) The expression of three genes identified as C4 regulator was shown along with the leaf development. (D) The expression of 9 genes showed a circadian cycle, and the x-axis indicates different time points on the second day. Developmental stages and tissues: s,

seeding stage; pm, pre-mature stage; m, mature stage; r leaf, roll leaf; m leaf, mature leaf. Leaf developmental gradient showed 1-15 segment in one leaf blade from base to tip. Circadian rhythm showed 19-time point including first day 2h time span and second day 4h time span. Asterisk was used to distinguish the same time point on different days.

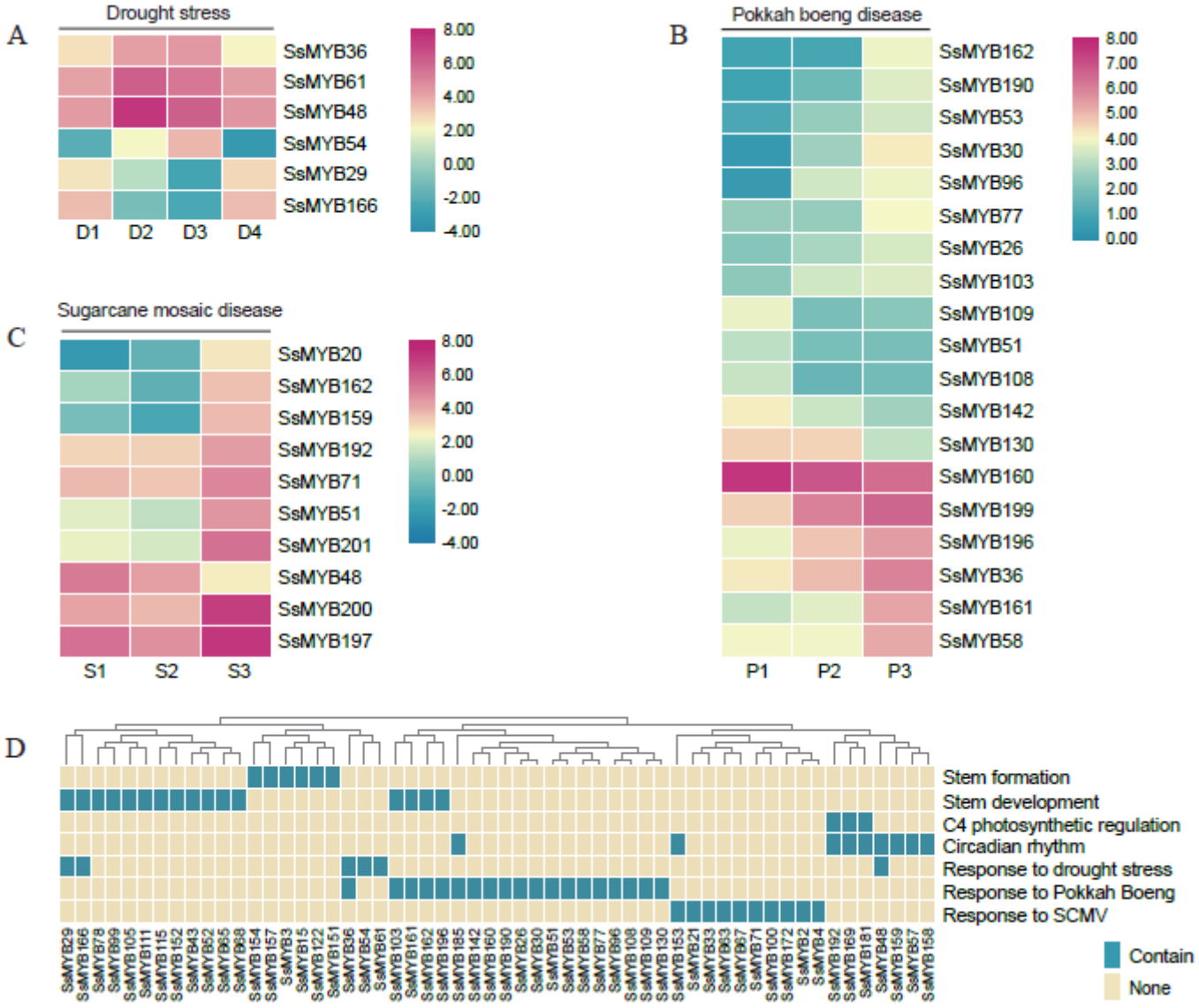


Figure 6

Heatmaps of differentially expressed MYB genes in response to stress conditions and presumed function. The heatmaps of DEGs expression value were shown in (A) (B) (C) based on RNA-seq data from drought stress, Pokkah boeng disease, and sugarcane mosaic disease. DEGs were identified due to their expression having significant variation after suffering stress stimulation (FPKM >2, fold change >2, p-value <0.05). Abbreviation: D1, CK; D2, mild; D3, severe; D4, rehydrate. P1, CK; P2, inchoate; P3, advanced. S1, CK; S2, CK detoxify; S3, post-infection. (D) 56 SsR2R3-MYB genes with specific expression patterns with putative functionalities. Blue boxes represent this function, and yellow with none.

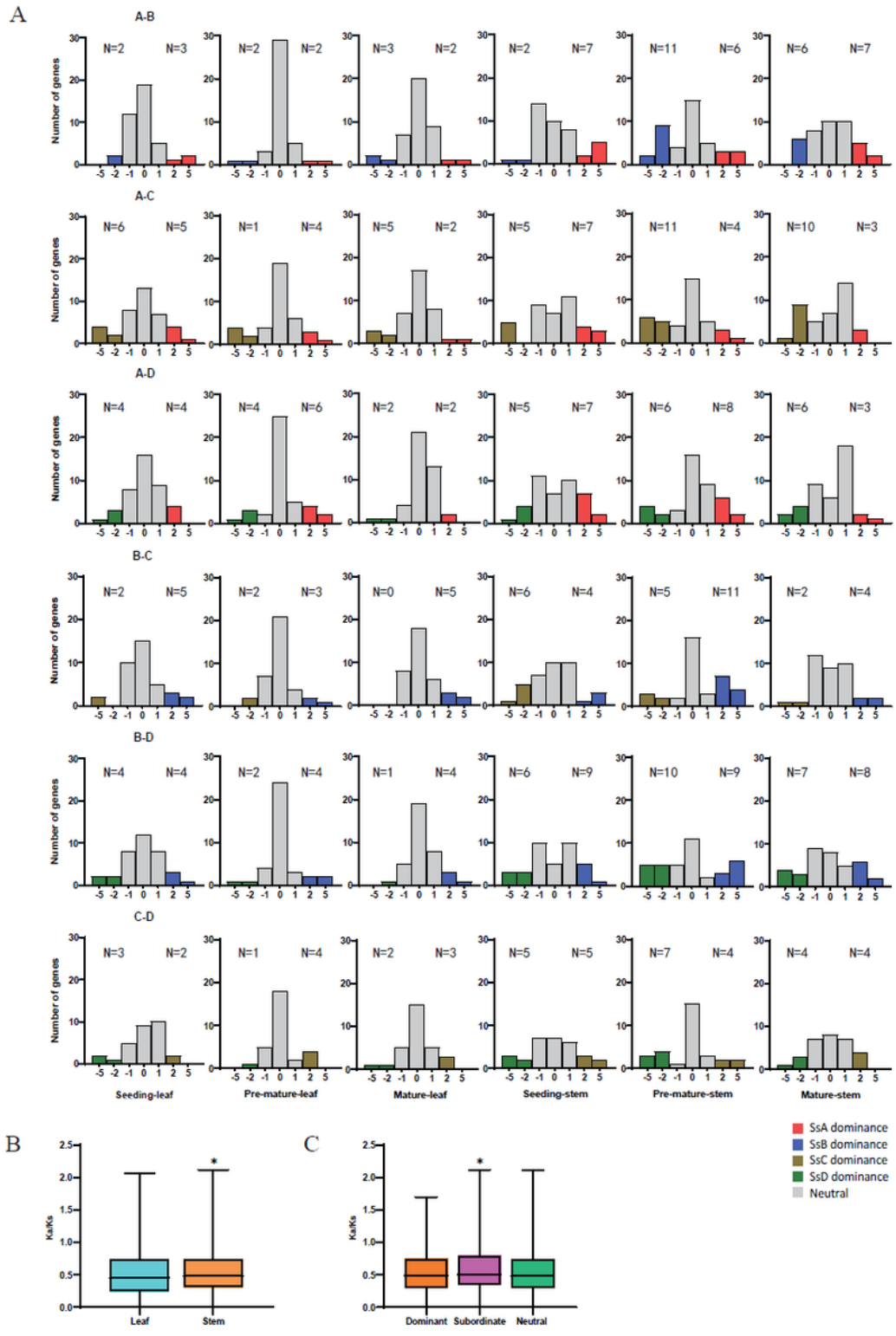


Figure 7

Allelic expression dominance and selective pressure analysis in SsR2R3-MYB family (A) Expression histograms of SsR2R3-MYB allelic genes among the tissue and development stage of S. spontaneum. N values indicate the number of dominant genes in allelic genes identified R2R3-MYB genes. (B) Boxplot of the distribution of Ka/Ks values of expression dominance genes in leaf and stem. Ka/Ks median values of leaf and stem are 0.455 and 0.482, respectively. (C) Boxplot of the distribution of Ka/Ks values among

allelic expression dominance genes as dominant, subordinate, and neutral (non-dominance). Ka/Ks median values of dominant, subordinate, and neutral are 0.491, 0.506, and 0.485, respectively.

Supplementary Files

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

- FigureS1.pdf
- FigureS2.pdf
- FigureS3.pdf
- FigureS4.pdf
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx
- TableS5.xlsx
- TableS6.xlsx
- TableS7.xlsx
- TableS8.xlsx