Send Orders for Reprints to reprints@benthamscience.net

138 *Current Genomics,* 2020, *21,* 138-154

RESEARCH ARTICLE

Phylogenomic Analysis of R2R3 MYB Transcription Factors in Sorghum and their Role in Conditioning Biofuel Syndrome

Vinay Singh[1], Neeraj Kumar[1], Anuj K. Dwivedi[1], Rita Sharma[2] and Manoj K. Sharma[1,*]

[1]*Crop Genetics & Informatics Group, School of Biotechnology, Jawaharlal Nehru University, New Mehrauli Road, New Delhi-110067, India;* [2]*Crop Genetics & Informatics Group, School of Computational & Integrative Sciences, Jawaharlal Nehru University, New Mehrauli Road, New Delhi-110067, India*

ARTICLE HISTORY

Received: December 06, 2019
Revised: March 17, 2020
Accepted: March 19, 2020

DOI:
10.2174/1389202921666200326152119

Abstract: ***Background*:** Large scale cultivation of sorghum for food, feed, and biofuel requires concerted efforts for engineering multipurpose cultivars with optimised agronomic traits. Due to their vital role in regulating the biosynthesis of phenylpropanoid-derived compounds, biomass composition, biotic, and abiotic stress response, R2R3-MYB family transcription factors are ideal targets for improving environmental resilience and economic value of sorghum.

***Methods*:** We used diverse computational biology tools to survey the sorghum genome to identify R2R3-MYB transcription factors followed by their structural and phylogenomic analysis. We used in-house generated as well as publicly available high throughput expression data to analyse the R2R3 expression patterns in various sorghum tissue types.

***Results*:** We have identified a total of 134 R2R3-MYB genes from sorghum and developed a framework to predict gene functions. Collating information from the physical location, duplication, structural analysis, orthologous sequences, phylogeny, and expression patterns revealed the role of duplications in clade-wise expansion of the R2R3-MYB family as well as intra-clade functional diversification. Using publicly available and in-house generated RNA sequencing data, we provide MYB candidates for conditioning biofuel syndrome by engineering phenylpropanoid biosynthesis and sugar signalling pathways in sorghum.

***Conclusion*:** The results presented here are pivotal to prioritize MYB genes for functional validation and optimize agronomic traits in sorghum.

Keywords: Biofuel, phenylpropanoids, R2R3-MYB, transcription factors, sorghum, stress.

1. INTRODUCTION

Sweet sorghum accumulates high levels of directly fermentable sugars within the culm and therefore, has immense potential as a biofuel crop [1]. The lignocellulosic biomass left after grain harvesting and sugar extraction, can also be used as feedstock for biofuels [1, 2]. Due to limited resource requirements and abiotic stress tolerance, sorghum is considered as a future-ready multipurpose crop. However, it has a very brief history of breeding. To adapt to changing climatic conditions and evolving market model, the development of regional ideotypes with enhanced grain yields, brix content, amenability to deconstruction and ability to thrive under adverse environmental conditions is required for large-scale deployment of sorghum as a biofuel crop [1].

Transcription factors (TF), due to their regulatory roles, have been widely used in the past to improve the agronomic performance of crop plants [3-5]. MYB proteins comprise of, one of the largest in number and functionally heterogeneous TF family, and is widely distributed in all the eukaryotes. Since the identification of the first MYB protein in maize, a large number of MYB proteins have been identified from diverse plant species (Fig. **1**). MYB TFs are characterized by the presence of 1-4 highly conserved DNA binding MYB domains, with each of them typically denoted by "R" [6]. Based on the number of conserved domains, MYB TFs have been classified into four groups namely, 1R, 2R, 3R and 4R (four R1/R2 like repeats) [7]. Each MYB repeat is ~52 amino acids long that includes three hydrophobic residues (usually tryptophan) placed at regular intervals. Each MYB domain comprises three alpha-helices, of which 2nd and 3rd helix form a helix-turn-helix (HTH) structure that interacts with the major groove of the DNA [8-10]. The highly conserved hydrophobic tryptophan residues play a significant role in sequence-specific interactions with DNA [11].

The evolutionary presence of MYB domain-containing proteins in lower organisms dates back to the time of divergence of plant and animal lineages [12-14]. However,

*Address correspondence to this author at the Crop Genetics & Informatics Group, School of Biotechnology, Jawaharlal Nehru University, New Mehrauli Road, New Delhi-110067, India; Tel: +91-011-26738895;
E-mail: manojdarolia@gmail.com

1389-2029/20 $65.00+.00

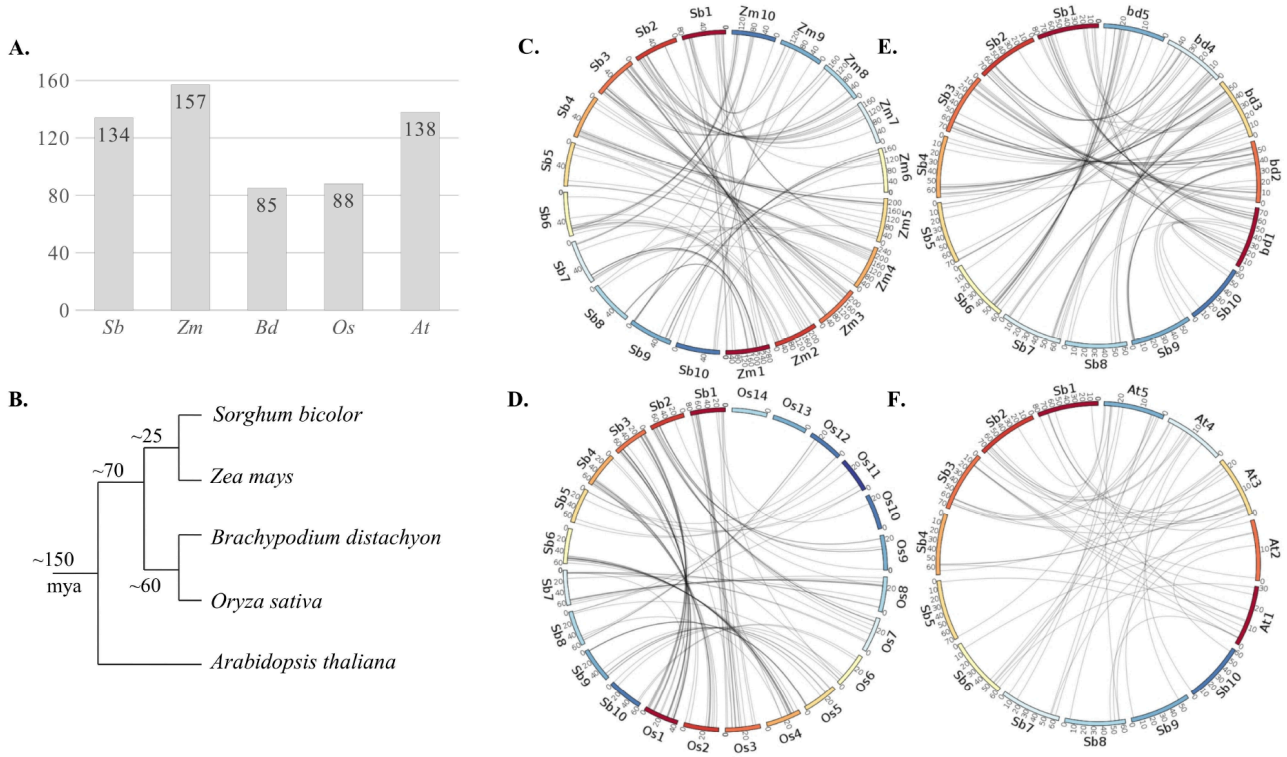

Fig. (1). Comparative mapping of MYB genes in different plant species. [**A**]. The number of MYB genes annotated from different plant species. [**B**]. The estimated divergence time between selected plant species is shown in the phylogenetic context. [**C-F**]. Distribution of R2R3-MYB orthologous gene pairs of sorghum with maize (**C**), rice (**D**), *Brachypodium* (**E**) and *Arabidopsis* (**F**). Sb: *Sorghum bicolor*, Bd: *Brachypodium distachyon*, Zm: *Zea mays*, Os: *Oryza sativa* and At: *Arabidopsis thaliana*. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

lineage-specific expansion of MYB proteins resulted in diversity in MYB proteins in diverse species. 3R-MYB proteins are dominant MYBs in the animal kingdom, while, 2R-MYB proteins are more prevalent in plants. The rapid expansion of 2R-MYB proteins occurred during the evolution of land plants and this expansion in plants is consistent with the whole-genome duplication events in angiosperms.

Based on the evolutionary analysis, MYB proteins have been divided into three groups namely, A-, B- and C-groups. Earliest angiosperms had all the three groups of MYB proteins while later during the evolution, lineage-specific duplication and domain loss resulted in variation in MYB family members in modern higher plants [15, 16]. Further, lineage-specific conservation of small regulatory motifs within the MYB domains and intron gain, supporting the intron late hypothesis, has been reported during MYB protein evolution in plants [12, 15, 17-19].

The 2R-MYB's, also referred as R2R3-MYB TFs, are the highest in number in plants. Several of these proteins have been characterized through genetic approaches and shown to play role in diverse biological processes ranging from morphogenesis, meristem formation, secondary cell wall biosynthesis, hormonal signal transduction, light signalling, male fertility, flowering, seed development, disease resistance, herbivore resistance, abiotic stress tolerance, and secondary metabolite biosynthesis [7, 16, 20, 21]. Many of these traits, such as flowering time, saccharification efficiency, cell wall composition, abiotic stress tolerance and disease resistance (collectively referred to as biofuel syndrome), are directly associated with biofuel production making them ideal candidates for enhancing biofuel-related traits [21, 22]. In fact, the application of R2R3-MYB genes in improving sugar release from plant-based biomass has already been demonstrated in several plant species. For instance, overexpression of *MYB31* and *42* in sugarcane improves glucose release with a concomitant decrease in acid-insoluble lignin [23]. Similarly, *PtoMYB170* positively regulates lignin biosynthesis genes and lignin deposition in secondary walls of xylem cells during wood formation in poplar [24]. Whereas, *ZmMYB31* downregulates several monolignol synthesis genes with its overexpression in transgenic plants resulting in significant reduction in lignin content [25]. Similarly, overexpression of *PvMYB4* led to reduced lignin and increased saccharification in switchgrass [26]. Therefore, systematic identification and characterization of R2R3-MYB genes in sorghum is a key step towards optimising sweet sorghum for enhanced biofuel-related traits.

Leveraging recent improvements in assembly and gene annotations of sorghum [27, 28], we performed genome-wide data mining to identify the complete repertoire of R2R3-MYB family genes in sorghum. Subsequently, we used inferences from localization, duplication, structure, phylogeny, orthology and gene expression to predict R2R3-MYB gene functions in sorghum. We identified a total of 134 R2R3-MYB genes in sorghum and classified them into 14 groups based on the phylogenetic analysis. Mapping of expression data along with information about orthologous gene functions from other plant species on phylogenetic tree

facilitated gene function prediction. RNA-sequencing based expression profiling of R2R3-MYB genes from temporal stages of internodes of sweet sorghum cultivar SSV84 revealed contrasting roles of genes belonging to the same clade in phenylpropanoid biosynthesis and sugar signalling.

2. MATERIALS AND METHODS

2.1. Identification of R2R3 MYB Proteins

The Hidden Markov Model (HMM) profile of MYB DNA-binding domain (PF00249), obtained from Pfam v31.0 [29] (http://pfam.xfam.org/family/PF00249), was queried against the sorghum proteome available at Phytozome v3.1 [30] (https://phytozome.jgi.doe.gov/pz/portal.html). The hits were filtered using E-value cut off of 1.0 and the presence of MYB domain in all the proteins was confirmed using NCBI-Conserved Domain Database [31] (CDD; https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) as well as SMART domain search tool [32] (http://smart.embl-heidel berg.de/). The R2R3 type MYB proteins were segregated based on the number of MYB repeats.

2.2. Chromosomal Localization and Duplication Analysis

The chromosomal coordinates of all R2R3-MYB proteins were extracted from phytozome and mapped onto sorghum chromosomes using Mapchart v2.30 [33] (https://www.wur.nl/en/show/Mapchart-2.30.htm). Tandemly duplicated genes were identified using Plant Tandem Duplicated Genes Database (PTGBase) [34], whereas, information about segmental duplications of sorghum MYB genes was extracted from Plant Genome Duplication Database (PGDD) [35]. The segmental duplications were illustrated using Circos v0.69-5 (http://circos.ca/; [36]). The subcellular localization of all proteins was checked using five different online available tools including Plant-mPLoc [37] (http:// www.csbio.sjtu.edu.cn/bioinf/plant-multi/), MultiLoc2 [38] (https://abi-services.informatik.uni-tuebingen.de/multiloc2/ webloc.cgi), Cello v 2.5 [39] (http://cello.life.nctu.edu.tw/), DeepLoc-1.0 [40] (http://www.cbs.dtu.dk/services/DeepLoc/ cite.php) and WoLF PSORT [41] (https://wolfpsort.hgc.jp/).

2.3. Identification of Orthologs and Synteny with Other Genomes

The orthologous MYB proteins from other genomes were identified using Inparanoid v4.1 [42] with 100% bootstrap confidence (http://inparanoid.sbc.su.se/cgi-bin/index.cgi). For identifying orthologous genes on syntenic regions between maize, rice, Arabidopsis, Brachypodium and sorghum, chromosomal coordinates of all R2R3-MYB proteins were mapped onto PGDD [35] (http://chibba.agtec.uga.edu/duplication/) and visualized using Circos v0.69-5 [36] (http://circos.ca/).

2.4. Structural Analysis

The protein sequences of all R2R3-MYB proteins were aligned using MAFFT v7.1 [43] with the iterative refinement method (1000 cycles; https://mafft.cbrc.jp/alignment/software/). Sequence features of the MYB domains were analysed by generating the weblogo of the aligned MYB domain using Weblogo 3 [44] (http://weblogo.threeplusone.com/create.cgi). Molecular weights and theoretical isoelectric point (pI) values of all the proteins were calculated using the ProtParam program at the Expasy bioinformatics resource portal [45] (https://web.expasy.org/protparam/). The information about exon-intron distribution was extracted from Phytozome and plotted using the Gene Structure Display Server [46] (http://gsds.cbi.pku.edu.cn/). The conserved motifs in R2R3-MYB proteins were identified using MEME search tool [47] (http://meme-suite.org/tools/meme) using the following parameters: motif discovery mode: classic; site distribution: zoops (zero or one occurrence per sequence); the number of motifs: 15; motif length: 6-50 amino acids. All 15 MEME motifs were scanned in TOMTOM (http://meme-suite.org/tools/tomtom) using all plant motif matrix profiles of JASPAR (http://jaspar.genereg.net/) with plant PFMs (position frequency matrices) as a target for similarity with any known motifs. Promoter sequences (2000 bp upstream region) of all SbMYB genes were extracted using the BioMart tool of Phytozome [30]. Sequences were submitted to the PlantCARE website (http://bioinformatics.psb.ugent.be/webtools/plant care/html/) to identify the *cis*-regulatory elements. The element matrix was visualized using Gephi, an open graph *vis* platform [48].

2.5. Phylogenetic Analysis of MYB Proteins

The multiple alignment of R2R3-MYB proteins was performed using standalone MAFFT v7.407 [43] (https://mafft.cbrc.jp/alignment/software/). Gap opening and gap extension penalty were set to 2.3 and 0.63, respectively. Alignment results, obtained above, were further checked and manually corrected using Jalview [49]. The corrected alignment was imported in MEGA (version 7.0) [50] to generate phylogenetic trees using neighbour-joining (NJ) and Maximum Likelihood (ML) methods with the Poisson model and pairwise deletion of gaps. The bootstrap analysis was performed with 1,000 replicates. For phylogenomic analysis of sorghum R2R3-MYB proteins with the genetically characterized R2R3-MYB proteins from other plant species, information about characterized MYB proteins was manually extracted from literature and their protein sequences were retrieved from Phytozome or NCBI sequence databases. A combined phylogenetic tree of sorghum R2R3-MYB proteins with previously characterized R2R3-MYB proteins from other plant species was generated using MEGA7. For the visualization of phylogenetic tree, Interactive Tree of Life (iTOL) server was used [51] (https://itol.embl.de/).

2.6. Expression Analysis Using Publicly Available Data

To analyse the expression profiles of MYB genes in different developmental tissues, publicly available RNA-seq data from different stages of vegetative and reproductive development, and four different stages of stem internodes were extracted from previously published studies [52, 53]. To visualize expression patterns in a phylogenetic context, expression data were plotted onto the phylogenetic tree using the Interactive Tree of Life (iTOL) server [51] (https://itol.embl.de/).

2.7. Brix Content

To measure brix content, five biological replicates of middle internodes were collected from field-grown *Sorghum bicolor* cultivar SSV84 plants at three distinct stages of development corresponding to booting, milky to soft dough and physiological maturity. The juice was extracted by squeezing internodes and brix content was determined through a digital hand saccharometer/refractometer (Atagotype) [54]. Three biological readings were subjected to ANOVA (Analysis Of Variance) using ToolPak analysis in Microsoft Excel.

2.8. RNA Sequencing and Data Analysis

To analyse the expression patterns of R2R3-MYB genes during sugar accumulation in sweet sorghum, middle internodes from field-grown sorghum plants (variety SSV84) were harvested at three distinct developmental stages namely, booting, milky to soft dough and physiological maturity. RNA was extracted using TRIzol reagent and assessed using Agilent 2100 Bioanalyzer [55]. High-quality RNA (RIN >8), extracted from two biological replicates of each stage, was used for the preparation of sequencing libraries using TruSeq® Stranded Total RNA kit as per manufacturer's instructions followed by sequencing using Illumina HiSeq 2000 paired-end sequencing platform with an average read length of 100 bp. The low-quality reads (Phred score <30), adaptors and ribosomal RNA were removed using AdapterRemoval version 2.2.0 (https://adapterremoval.readthedocs.io/en/latest/) and Bowtie version 2.2.9, respectively. High-quality reads were aligned to the sorghum genome using the Hisat2 program (https://ccb.jhu.edu/software/hisat2/index.shtml) [56]. The expression values of all genes were estimated using StringTie version 1.3.3b (https://ccb.jhu.edu/software/stringtie/). The expression data of all MYB genes of sorghum was subjected to K-means clustering using MultiExperiment Viewer [57].

3. RESULTS AND DISCUSSION

3.1. Sorghum Genome Encodes 134 R2R3-MYB Genes with Evolutionarily Conserved Structural Features

MYB family transcription factors are characterized by a highly conserved helix-turn-helix DNA binding (MYB) domain, usually at the N-terminus of the proteins. MYB domain in R2R3-MYB proteins consists of two consecutive imperfect repeats of 50-53 amino acids. Each repeat forms three alpha-helices with the first two helices forming helix-turn-helix structure while the third recognition helix is involved in DNA binding [9, 58]. After removing redundancy and confirming the presence of two MYB repeats, we annotated 134 R2R3-MYB proteins from sorghum (Fig. **1A**; Supplementary Table **S1**). Based on the chromosomal coordinates, provided in Phytozome, we named sorghum R2R3-MYB genes from *SbMYB001* to *SbMYB134*. In most of the proteins, R2 and R3 domains were present in tandem, as expected. Conversely, thirteen MYB proteins contained long linker regions between two repeats that ranged from 35 to 345 amino acids (Fig. **2A**).

The average length of these proteins was around 356 aa with *SbMYB070* encoding the smallest protein of 174 aa and the *SbMYB119* coding for the longest protein of 1562 aa (Supplementary Table **S1**). Similarly, the average molecular weight of sorghum MYB proteins was estimated to be 38.7 kDa with an isoelectric point (pI) ranging from 4.57 to 10.68 (Supplementary Table **S1**). Analysis of subcellular localization using five different online prediction tools suggested nuclear localization for all the R2R3-MYB proteins conforming to their role in transcriptional regulation (Supplementary Table **S1**).

The number of R2R3-MYB genes accounts for 0.392% of the total genes annotated in the sorghum genome. This is similar to the proportion of R2R3-MYB genes annotated in rice with 88 genes (0.393%), *Brachypodium* with 85 genes (0.389%) and maize with 157 genes (0.387%) though the proportion of MYB genes in *Arabidopsis* is higher (0.59%) with 138 genes [59, 60] (Fig. **1A**). To compare the orthologous relationships among MYB proteins in these species, we identified orthologs of sorghum R2R3-MYB genes in all four genomes. The results were in agreement with the estimated evolutionary distance among these species [61] (Fig. **1B**). For instance, maize, estimated to have diverged about 25 million years ago from sorghum, had the maximum number (117) of orthologs of sorghum genes (Fig. **1B** and **C**). Whereas, rice and *Brachypodium*, estimated to have diverged about 70 million years ago from sorghum clade, had 107 orthologs each (Fig. **1B**, **D**, and **E**). As expected, dicot plant *Arabidopsis*, with an estimated divergence time of more than 150 million years, had orthologs for only 51 sorghum R2R3-MYB genes (Fig. **1B** and **F**; Supplementary Table **S2**).

Further, to examine the conserved sequence features of R2 and R3 MYB repeats, we performed multiple alignments and generated separate sequence logos for both R2 and R3 domains (Fig. **2B**). As shown in their counterparts from other plant species, landmark tryptophan (W) residues (three in R2 and two in R3 domain), were conserved in both the repeats of sorghum proteins (Fig. **2B**). These residues are involved in sequence-specific binding of DNA [62]. The first tryptophan in the R3 domain was replaced by another hydrophobic residue mostly phenylalanine (Fig. **2B**). A cysteine residue in the third helix of the R2 domain was also conserved (Fig. **2B**). Cysteine at this position in a maize protein is essential for DNA binding [58]. Several other residues such as glutamic acid (E)-11, aspartic acid (D)-12, leucine (L)-15, glycine (G)-23, leucine (L)-36 and leucine (L)-51 were also conserved in R2 repeat. Whereas in R3 repeat, proline (P)-2, glutamic acid (E)-13, isoleucine (I)-31, arginine (R)-8, asparagine (N)-41 and asparagine (N)-45 were also conserved. It would be interesting to explore the functional relevance of these conserved amino acids.

In addition to conserved motifs within the proteins, *cis*-regulatory elements may further strengthen the functional categorization based on their phylogenetic placement. Towards this, MYB gene promoter sequences (2000 bp) were evaluated for the presence of various *cis*-regulatory elements (Supplementary Table **S3**). Promoter elements, identified from regulatory regions of MYB genes, were characterized into four major categories *i.e.* growth and development, biosynthetic regulation, biotic stress and abiotic stress (Supplementary Fig. **1**). More than 50% of the total *cis*-regulatory elements were associated with growth and development,

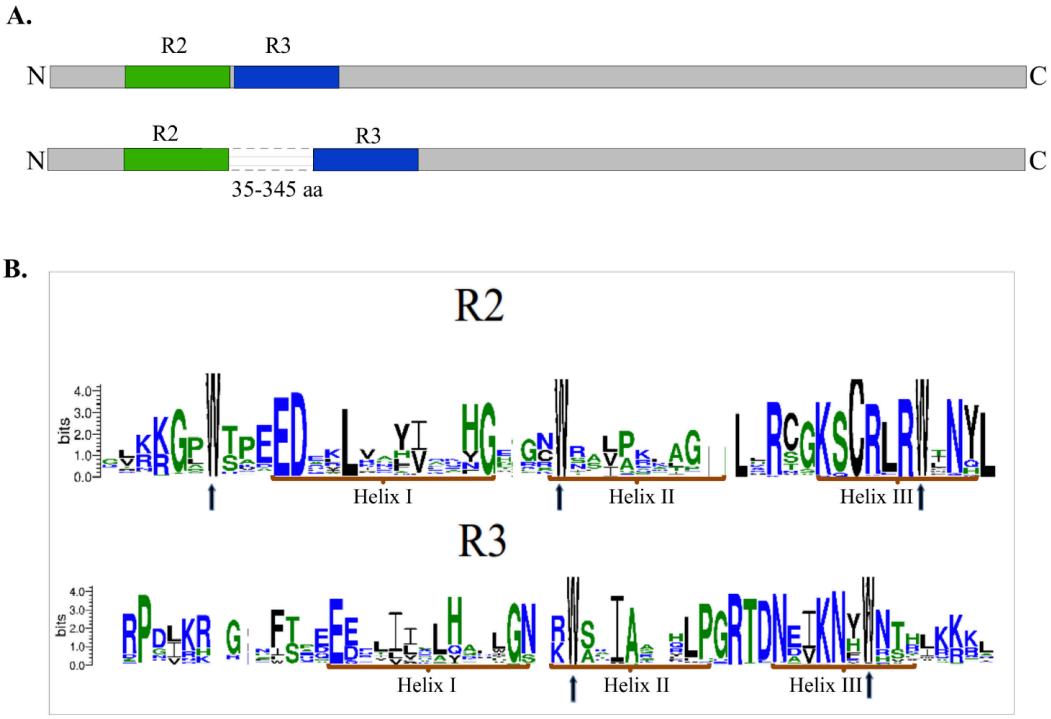

Fig. (2). Illustration of gene structures (**A**) and sequence logos of R2 and R3 MYB repeats (**B**) of sorghum R2R3-MYB proteins. The bits scores on the Y-axis provide information for each position in the sequence. The position of three alpha helices is marked in maroon, while arrows highlight landmark tryptophan residues in the repeat regions. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

which is in consensus with the predicted roles of R2R3 MYB proteins in regulating plant growth and development [20]. Further, among the growth and development-related elements, light-responsive elements were present in almost all the promoters with the highest frequency compared to other categories (Supplementary Table **S3**).

3.2. Analysis of Evolutionary Relationships between Sorghum R2R3-MYB Proteins

To investigate the evolutionary relationships among sorghum MYB proteins, we aligned complete protein sequences of R2R3-MYB proteins and constructed an unrooted phylogenetic tree using the NJ method (Fig. **3**) and ML methods (Supplementary Fig. **2**). Five proteins (*MYB008*, *017*, *079*, *086* and *119*) with long linker regions between MYB repeats did not align and therefore, were not included in the phylogenetic analysis. Based on the sequence similarity and tree topology, R2R3-MYB proteins were divided into 14 groups numbered from I to XIV using Roman numerals (Fig. **3**). Two proteins, *SbMYB115* and *070*, that did not fall in any of the groups, are highlighted by letters a and b in the phylogenetic tree. The number of proteins in each group ranged from 2 to 24. The classification of MYB proteins was also supported by the structural features of these proteins. Although the number of introns varied from 0 to 13, 80% of the R2R3-MYB genes had one or two introns (Supplementary Table **S1**). Further, we also noticed group-specific patterns such as all genes belonging to group IX, except one, lacked introns (Fig. **4**). On the contrary, all genes from group XII and XIV had three or more introns. A total of fifteen genes were predicted to encode multiple transcripts, while for rest 119 genes, only one transcript was annotated. Although the length of introns varied among genes belonging to the same group, group I genes had the smallest introns. The distribution of proteins into different clades mostly remained the same in the tree generated using the ML method (Supplementary Fig. **2**).

The analysis of MEME motifs revealed similar motif distribution among members of the same group. A total of 15 motifs were identified from R2R3-MYB proteins of sorghum [63, 64] (Supplementary Table **S4**). Eight of these motif (1-6, 9 & 10) represent the smaller conserved regions within the R2R3 MYB domains. Motif 1-5 were conserved in most of the MYB proteins except group XIII. Motif-6 was conserved in the MYB domain of proteins belonging to group I-VII and XIII, while, group IX and X members had motif-10 in place of motif-6 (Fig. **4**). In rest of the proteins, this region seems to have diverged (Fig. **4**). Members of group XIII and XIV showed rearrangement of the MEME motifs corresponding to the MYB domain. In some of the group XIII and XIV proteins, motif-9 corresponded to the part of R2 MYB domain while motif-11 corresponded to the R3 MYB domain of group XIII members. Closer investigation revealed that the linker region between R2 and R3 repeats of group XIII members is unusually long. Conservation of phylogenetic clade-specific MEME motifs in the regions of R2R3 MYB domains and variation in the linker region joining R2 and R3 domain suggest divergence of MYB domain architecture in sorghum MYB proteins and subsequent functions (Fig. **4**). Some motifs present in the highly variable C-terminal region of R2R3-MYB proteins may also contribute to the functional divergence of MYB genes by acting as transcriptional

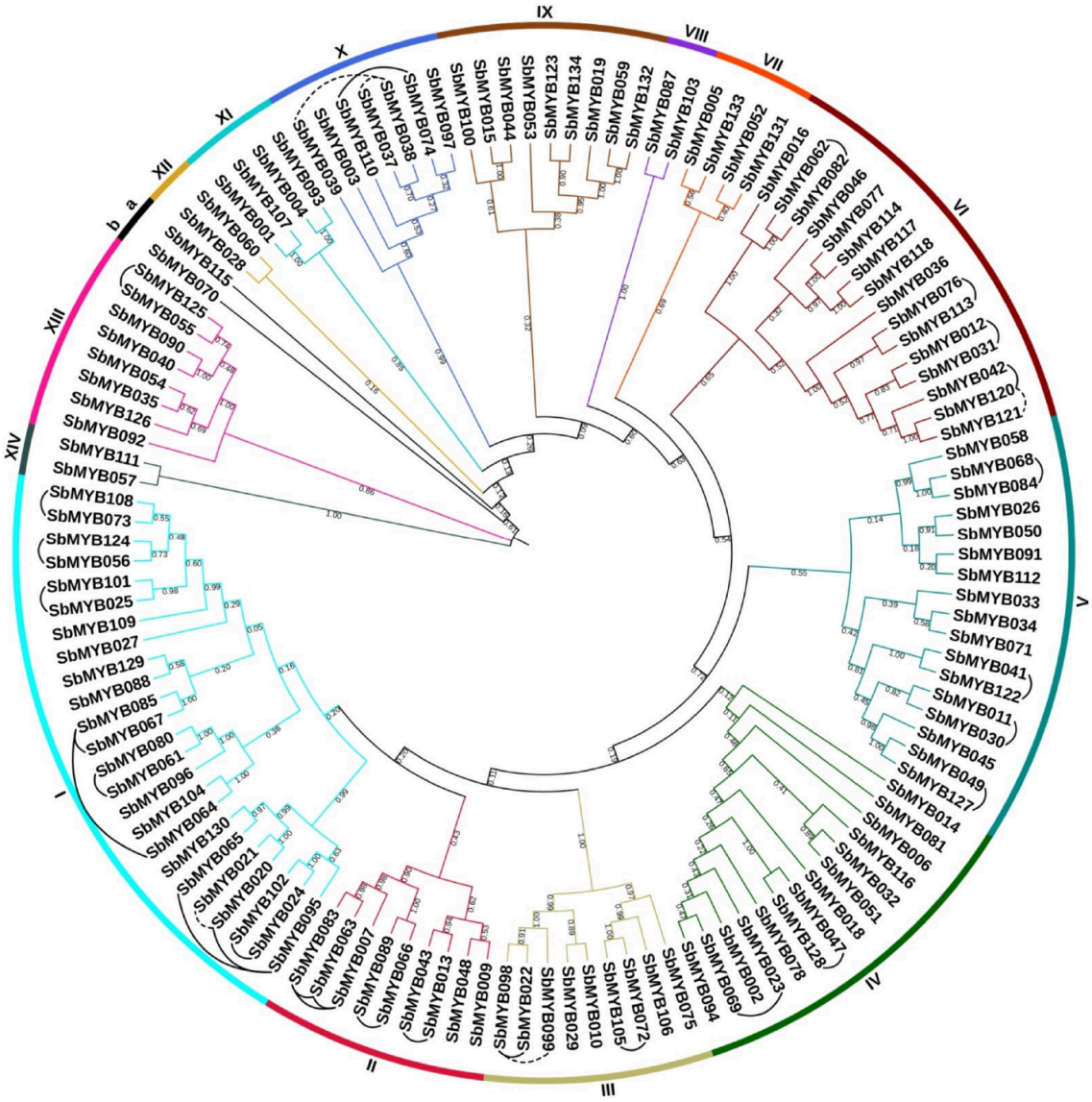

Fig. (3). Phylogenetic analysis of sorghum R2R3-MYB proteins. A total of 129 proteins were used to construct a neighbour-joining (NJ) tree. Bootstrap values are given for all the branches. The 14 groups are marked by distinct colors and, named using Roman numerals (I to XIV). The solid lines highlight segmentally duplicated gene pairs while those connected through dashed lines represent tandem duplicates. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

activators or repressors (Supplementary Table **S4**). Further, some of them likely provide sites for post-translational modifications and physical interactions [65].

Motif-7 is only present in group VI members and is found close to the R3-MYB domain. It has conserved histidine, leucine, aspartic acid, serine, methionine, and arginine residues and has been detected in genes that specifically regulate JA-dependent transcriptional responses [66]. Further, MYB proteins having this motif align with characterized MYB genes implicated in abiotic stress (Fig. **5**). Motif-8 is specifically rich in glutamine (Q) residues and is present in members of group 1-6 (Supplementary Table **S4**). Poly (Q) motifs have been found to stabilize the protein-protein interactions and are associated with protein aggregation in plants as well as animals [67, 68]. In humans, the presence of poly(Q) motifs has been associated with molecular pathogenesis responsible for several diseases [69]. Motif-12 is only present in group III members and has ExWLL/FDD motif (Supplementary Table **S4**). It is located at the C-terminal end of the proteins. NCBI search of this motif showed that it is specifically associated with MYB or MYB-related proteins. When searched in the plant motif matrix profile of JASPAR for similarity with known motifs, results suggested that it may recognize GA-rich DNA motifs and regulate anthocyanin or flavanol content in plants (Supplementary Table **S4**). Interestingly, some of the characterized MYB proteins that align with group III members are known to regulate anthocyanin pathway and stress tolerance. Motif-13 has a highly conserved "Phe-Leu-Gly-Val-Gly", which has been shown to bind to DNA [70]. It is exclusively present in group X members and is present at the C-terminal end of the MYB proteins. Motifs-14 and 15 are single amino acid repeats of histidine and proline, respectively. Histidine rich motifs can interact with metal ions and have been shown

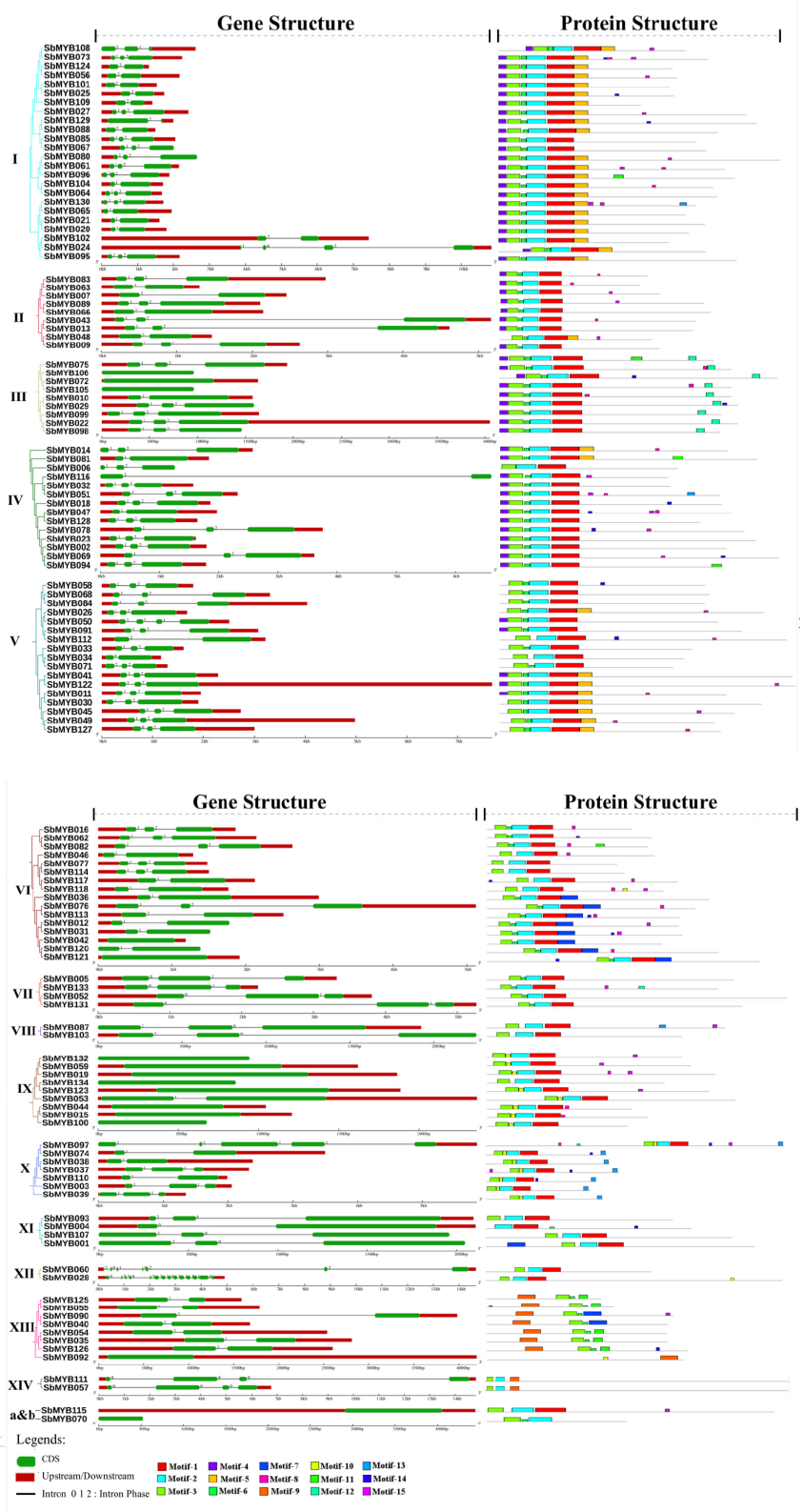

Fig. (4). Structural analysis of sorghum genes and proteins in their phylogenetic context. The genes are arranged separately for all 14 groups as per their order in the phylogenetic tree. The intron-exon distribution of each gene has been illustrated with exons represented by filled green bars and introns shown by black lines. The red bars stand for 5' and 3'UTRs. The length of genes is scaled to the bar (Intron phase is marked 0-phase if it does not disrupt a codon, 1-intron phase disrupts a codon between first and second bases, and in phase 2 introns codon is disrupted in the second and third bases). The second panel shows the distribution of conserved motifs in corresponding protein sequences as identified by MEME motif search tool. The color legend and sequence for MEME motifs are given at the base. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

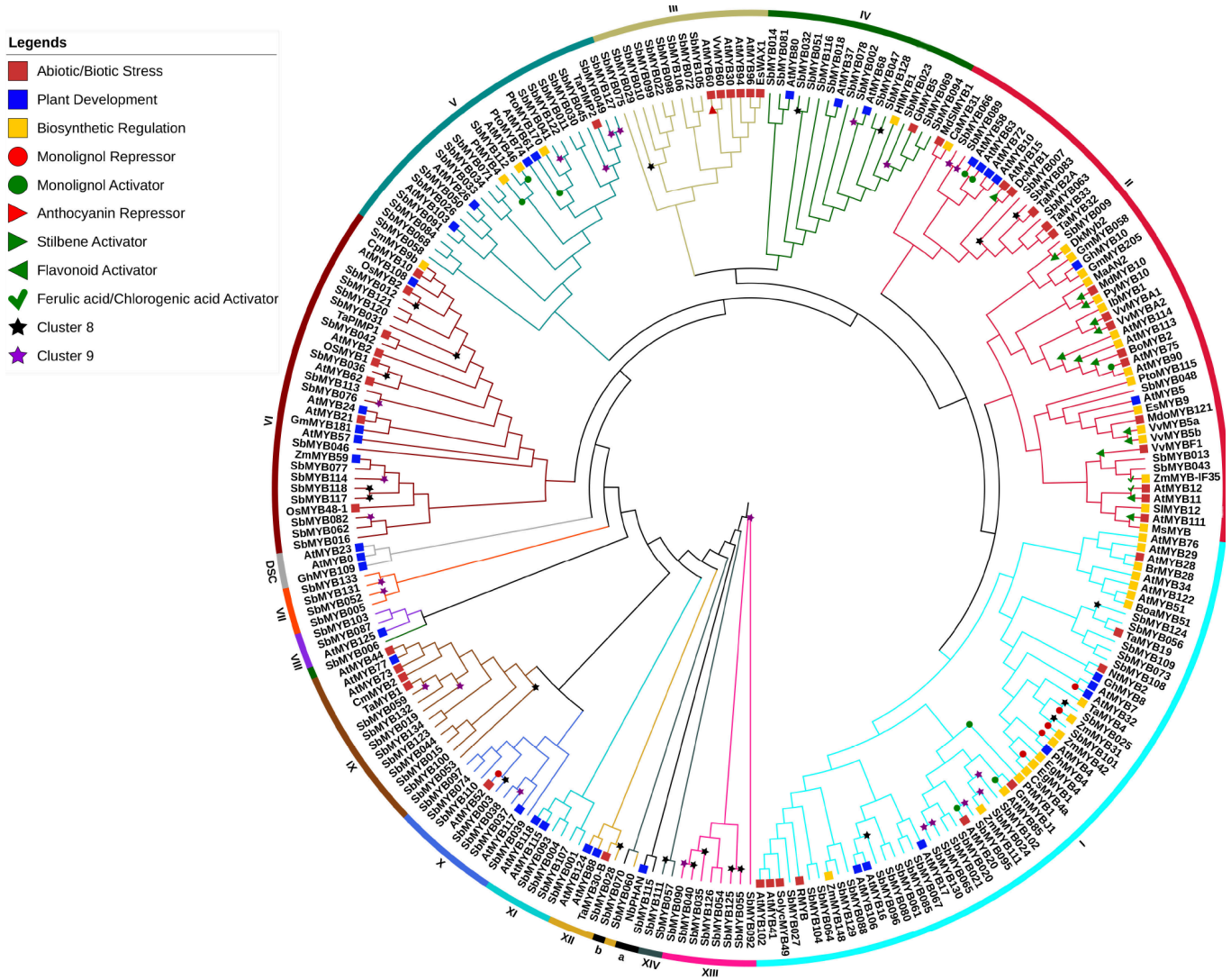

Fig. (5). Comparative phylogenetic tree showing evolutionary relationships between sorghum R2R3-MYB proteins and experimentally characterized R2R3-MYB proteins from other plant species. The groups are designated by roman numerals. The DSC stands for dicot specific clade as no gene from sorghum or other monocot species falls in this clade. Functions of the characterized genes are highlighted using symbols of distinct colors as per legend given on the top. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

to play diverse roles in DNA-protein interactions, protein conformation, nuclear targeting and transcriptional regulation [71-73]. Whereas, consecutive proline residues in the coding sequence have been shown to cause translational slow down through ribosome stalling. Recently, Gall and coworkers [74] described Mg^{2+}-dependent control of translational speed as a metabolite sensor mechanism, which utilizes relatively slow translation of proline codons. Therefore, the presence of motifs that regulate the translation speed may be important for translational regulation of MYB proteins [75, 76].

3.3. Chromosomal Localization and Duplication Analysis Suggest a Key Role of Gene Duplications in the Expansion of R2R3-MYB Family in Sorghum

Based on the chromosomal coordinates, we localized sorghum R2R3-MYB genes on respective chromosomes (Fig. **6A**). All 133 genes mapped onto sorghum chromosomes except one that could be traced on unanchored contigs. The R2R3-MYB genes showed an uneven distribution with the maximum number of genes (26) localized on chromosome 3, while, only five genes were located on chromosome 10 (Fig. **6A**). Interestingly, most of the genes mapped on the terminal regions of chromosomes (Fig. **6A**).

Analysis of tandem arrays of identical sequences in close genomic proximity revealed six clusters of tandemly duplicated genes. Except for a cluster of three genes on chromosome 3, rest five clusters contained only two genes each (Fig. **6A**). The percent identity between tandem duplicated genes varied from 33 to 67% (Supplementary Table **S5**). Conversely, 28 pairs of duplicated genes that mapped to duplicated segments of the sorghum genome, exhibited 39 to 70% identity among duplicated gene pairs (Supplementary Table **S5**). Except for chromosome 10, all chromosomes contained segmentally duplicated genes with the maximum number of genes mapped on syntenic regions between chromosomes 4 and 6 as well as chromosomes 3 and 9 (Fig. **6B**).

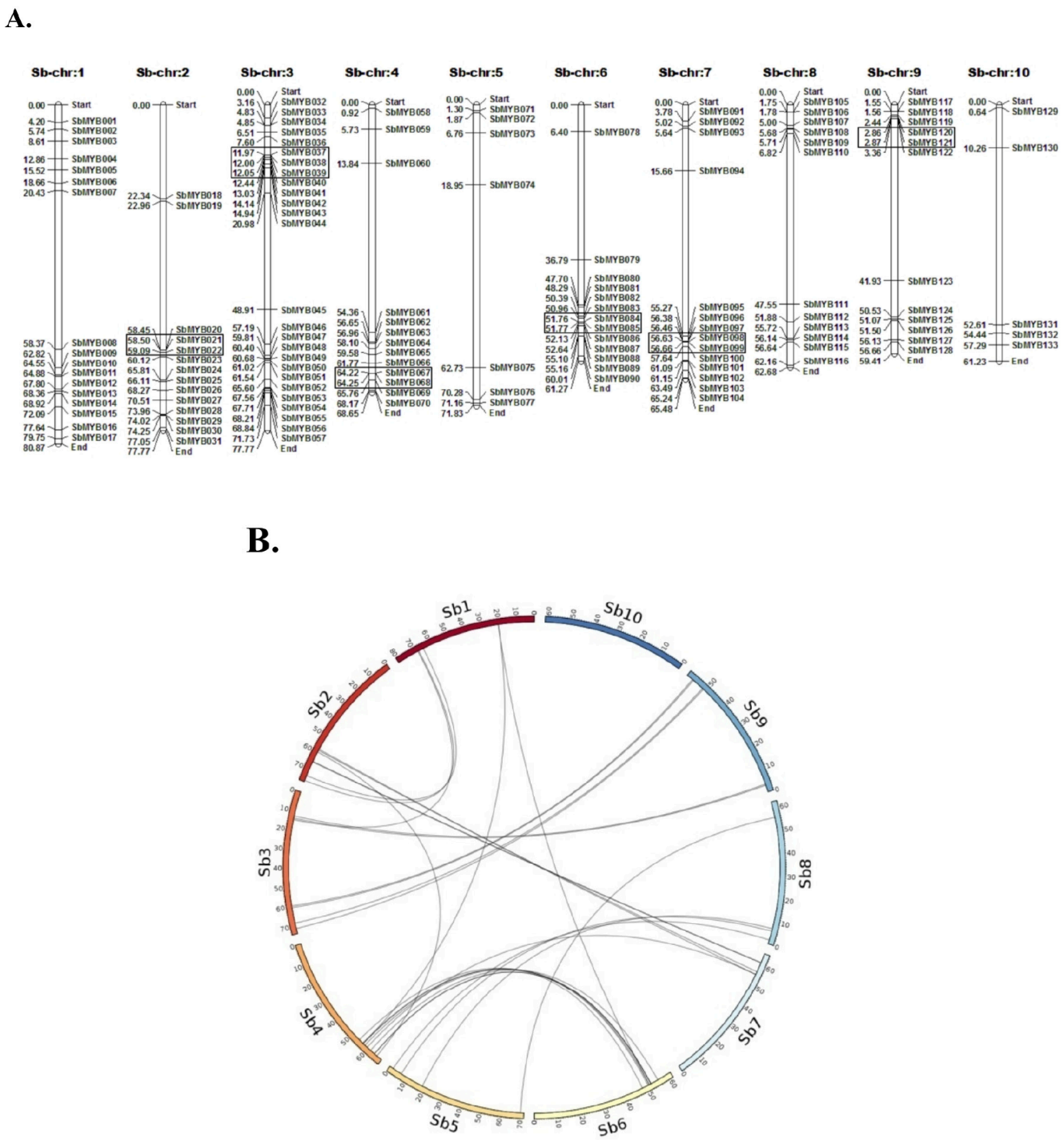

Fig. (6). Chromosomal localization and duplication analysis of sorghum R2R3-MYB genes. **A**. The position of all SbMYB genes on sorghum chromosomes is indicated as per their coordinates. Tandemly duplicated gene clusters are marked with a rectangular box. The numbers on the left of chromosomes indicate chromosomal positions of R2R3-MYB genes of sorghum in Mb. **B**. Inter-chromosomal relationships of sorghum R2R3-MYB genes. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

Mapping of duplicated genes on the phylogenetic tree revealed clade-specific expansion of R2R3-MYB genes. Group I with eight pairs of duplicated genes has the highest number of paralogous genes. Similarly, group VI with nine paralogous genes also seems to have undergone expansion due to segmental duplications. The non-synonymous/ synonymous substitution (Ka/Ks) ratio of all duplicated gene pairs was less than 1 suggesting that R2R3-MYB proteins have evolved under strong purifying selection (Supplementary Table **S5**). The lower number of tandem duplicates compared to gene pairs on segmentally duplicated regions in R2R3-MYB proteins may also be attributed to positive selection driving higher retention of segmental duplicates [77]. The divergence time of duplicated gene pairs was estimated to vary between 30 to 100 million years ago.

3.4. Combined Phylogenetic Analysis with Experimentally Characterized Orthologs from Other Species Reveals Intra-clade Functional Diversity

Several R2R3-MYB genes have been experimentally characterized by different plant species. We thoroughly mined literature for experimentally characterized R2R3-MYB proteins and shortlisted 123 genes from different plant species (Supplementary Table **S6**). Based on the major phenotype in overexpression, mutant or silencing lines, these were categorized into three major functional classes including plant development, biotic/abiotic stress response and biosynthetic regulation. A combined phylogenetic tree using 129 sorghum R2R3-MYB proteins and 123 characterized R2R3-MYB proteins from various plant species was generated (Fig. **5**). The topology of the tree was the same as that of the sorghum R2R3-MYB tree presented in Fig. (**3**) except for minor rearrangements. A group IV protein, *SbMYB006* aligned with group VIII clade, whereas, *SbMYB060* that aligned with *SbMYB028* of group XII earlier, aligned with *SbMYB070* (designated with b) in the combined tree.

Except for groups VII, XIII and XIV, experimentally characterized R2R3-MYB proteins were dispersed throughout different phylogenetic groups (Fig. **5**). Group I had 34 characterized genes with seven of them implicated in plant growth and development, nine genes regulating biotic and abiotic stress response and 18 genes involved in biosynthetic processes. Similarly, 39 characterized genes clubbed with group II with about an equal number of genes implicated in stress response (17) and biosynthetic processes (16) and, six genes associated with plant development (Supplementary Table **S6**). Group III contained six characterized genes, all of which are characterized in dicot plants and form a separate subclade within the group III. These determine abiotic stress response by regulating cuticular wax biosynthesis. Group IV had five characterized genes, three of which regulate plant development (Supplementary Table **S6**). Similarly, group V contained eight characterized genes, four of which are involved in regulating plant development, while the remaining four are associated with biosynthetic processes associated with secondary cell wall biosynthesis and lignification (Supplementary Table **S6**). Group VI had 14 characterized genes with eight of them associated with stress response while the rest five genes regulate male reproductive organ (anther) development. Group VIII had only one characterized gene involved in the specification of sperm cells. Group IX, on the other hand, had five characterized genes with four of them making a separate clade implicated in abiotic stress response. Group X had only two characterized genes with one of them involved in ovule development and the other gene associated with multiple stress responses. Furthermore, group XI contained two characterized genes involved in vegetative to embryonic transition during embryogenesis, whereas, three characterized genes of group XII determine drought tolerance by regulating stomatal development (Supplementary Table **S6**). The representation of all three functions in most of the clades highlights intra-clade functional diversity in R2R3-MYB genes.

3.5. Role of R2R3-MYB Genes in Phenylpropanoid Biosynthesis

Due to high biomass content, lignocellulosic biomass of sorghum holds immense potential as biofuel feedstock [2]. However, recalcitrance due to the presence of lignin is a major constraint to enzymatic degradation of lignocellulosic biomass. Therefore, reducing the amount of lignin using biotechnological approaches is a major area of investigation in engineering lignocellulosic feedstock [78]. R2R3-MYB transcription factors have been shown to play a significant role in lignin biosynthesis and its deposition on plant secondary cell walls. While some of them have been shown to activate lignin biosynthesis, others act as negative regulators by directly binding to AC elements in the promoters of lignin biosynthetic genes [53, 79-83]. The suggested functions of R2R3-MYB genes in secondary cell wall development are conserved in both monocot and dicot lineages [84]. Such as, ectopic expression of *ZmMYB31* and *42* resulted in reduced lignin content and improved saccharification efficiency in *Arabidopsis* [85].

To investigate the role of R2R3-MYB genes in secondary cell wall lignification in sorghum stems, we leveraged the RNA sequencing data available from four top internodes of bioenergy sorghum, R.07020, that represent sequential phases of development involving cell division, elongation and differentiation [53]. The first internode, being the youngest, exhibits active cell division while the fourth internode demonstrates cessation of growth and accumulation of secondary cell walls. Out of 123 R2R3-MYB genes, for which expression data in internodes is available, 31 exhibited the highest expression in the fourth internode thereby, suggesting a role in secondary cell wall biosynthesis (Supplementary Table **S7**). However, eleven genes *viz*., *SbMYB009*, *037*, *060, 086, 090, 092, 111, 115, 126, 131* and *132* exhibited higher expression in younger internodes that gradually declined in mature internodes suggesting their involvement in cell division. Interestingly most of the 31 genes with high expression in 4th internode, belong to groups I-VI. Whereas, genes showing higher expression in younger internodes are distributed across all phylogenetic groups with most of them having diversified R2 or R3 MYB domain (Fig. **7**). Three genes *SbMYB019*, *059* and *117* exhibited a similar level of expression in all four internodes, while the rest of the genes mostly exhibited variable or very low-level expression.

By collating information of the characterized orthologs from other plant species, phylogenetic placement and expression data in internodes, we have predicted key targets for secondary cell wall biosynthesis in sorghum. Notable *SbMYB025* and *SbMYB101* of group I are syntelogs of *ZmMYB31* and *42,* which have already been shown to negatively regulate lignin biosynthesis in maize [85]. Similarly, three genes from one of the clades in group I, *AtMYB20*, *AtMYB85,* and *PtMYB1* have earlier been shown to act as activators of monolignol biosynthesis [79, 86]. Also, sorghum genes in this group have a greater number of *cis*-regulatory elements associated with abiotic stress and biosynthetic regulation. Therefore, based on the expression patterns and sequence homology with these genes, we predict *SbMYB024, 065*, *095*, *102* and *130* of group I as important candidates for engineering lignin biosynthesis.

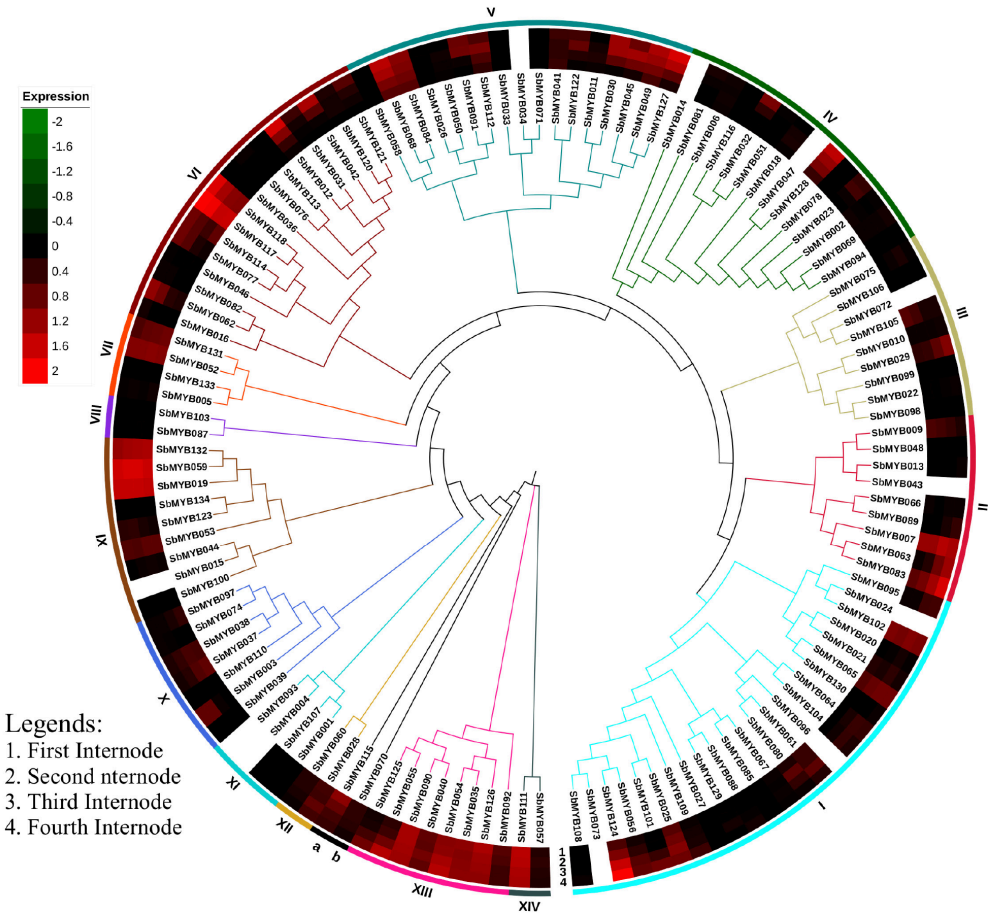

Fig. (7). Expression analysis of MYB genes in stages of internode development from bioenergy sorghum (genotype R.07020). The expression profiles of sorghum R2R3 MYB proteins were analysed in four sequential stages of internodes with the first internode being youngest and fourth internode corresponding to the oldest internode. Log_2 FPKM values were plotted as per color code given on the top left with green being low expression and red representing high-level expression. The phylogenetic groups are represented by Roman numerals. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

Further, *SbMYB066* and *089* of group II are closely related to *MYB63* and *58* of *Arabidopsis,* which have been reported to act as monolignol activators [79]. *SbMYB066* in our study corresponds to a previously characterized gene, *SbMYB60*, that has been shown to affect both primary and secondary metabolism by regulating UDP-sugar and cellulose biosynthesis-related genes [87]. Also, biosynthetic regulation- and stress-associated *cis*-elements were detected in their promoters.

Another gene, *SbMYB112* of group V forms a separate clade with both *PtMYB4* and *AtMYB46* acting as monolignol activators [86, 88]. Overexpression of *AtMYB46* resulted in the ectopic deposition of secondary walls around parenchymatous cells in *Arabidopsis*. In fact, Secondary wall-associated NAC Domain protein 1 (SND1), a regulator of the developmental program of secondary wall biosynthesis, binds to the MYB promoter and activates the biosynthetic pathways of secondary cell wall thickenings, suggesting that *AtMYB46* is a master regulatory component of secondary wall biosynthesis in *Arabidopsis*. Hence, it would be interesting to study if *SbMYB112* performs a similar function in sorghum. Similarly, *SbMYB003* is orthologous to *AtMYB52* which acts as a monolignol repressor [89]. These genes provide important candidates for engineering lignocellulosic biomass composition and thereby, saccharification efficiency in sorghum.

In addition to lignin, R2R3-MYB genes also regulate the biosynthesis of a diverse array of flavonoids (anthocyanins, proanthocyanidins, flavonols, isoflavonoids, and phlobaphenes), phenolic acids and stilbenes [21, 90]. These are of significant interest from both commercial and human health perspectives due to their anti-inflammatory and chemoprotective properties [90]. A total of eighteen genes of group II, from diverse plant species, have been experimentally shown to regulate flavonoid biosynthesis (Fig. **5**). Interestingly, all of them act as activators. Therefore, sorghum genes belonging to group II including *SbMYB007, 009, 043*, *048*, *063*, *083* and *103* are putative candidates for engineering flavonoid biosynthesis in sorghum. Two motifs, TGACG and CGTCA, associated with biosynthetic regulation were also identified in their promoters. Among these, *SbMYB103* shows high homology with *Yellow seed 1* (*Y1*), which is essential for the synthesis of 3-deoxy flavonoid pigments in sorghum [91]. The loss of function of *Y1* leads to yellow color due to the lack of visible phlobaphene in seed pericarp.

3.6. Role of Sorghum R2R3-MYB Genes in Vegetative and Reproductive Organ Development

Several R2R3-MYB genes have been shown to play a major role in regulating different aspects of plant growth and development ranging from female reproductive development

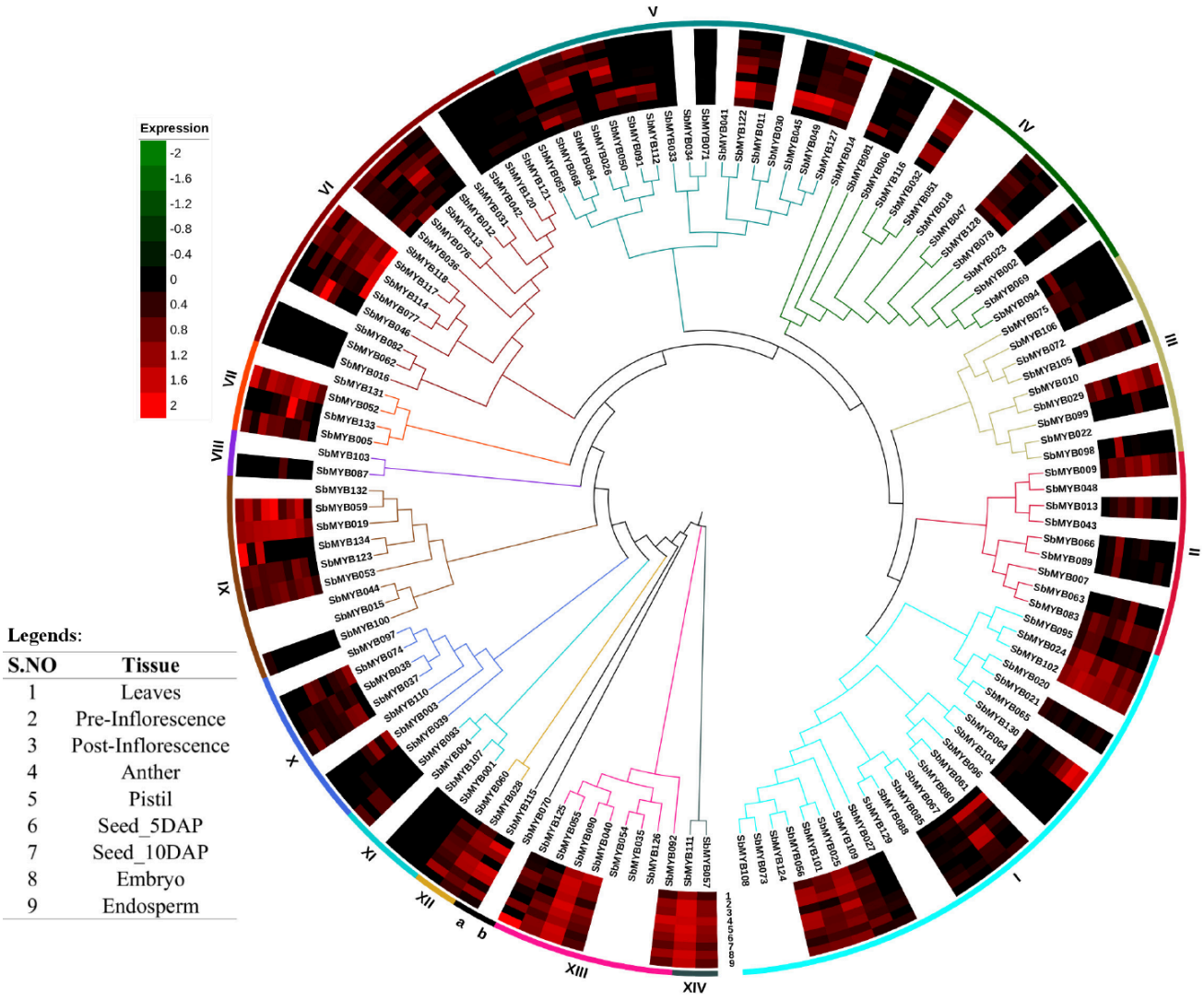

Fig. (8). Expression analysis of MYB genes in stages of vegetative and reproductive development. The expression profiles of sorghum R2R3 MYB proteins were analysed in nine stages of vegetative and reproductive development. Log_2 FPKM values were plotted as per color code given on the top left with green being low expression and red representing high-level expression. The phylogenetic groups are represented by Roman numerals. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

[92], morphogenesis [93], branch number [94], sperm cell differentiation [95] and seed size [96]. To investigate the spatial and temporal expression patterns of sorghum R2R3-MYB genes, we used publicly available data from nine developmental tissues *viz*., leaves, pre-emergence inflorescence, post-emergence inflorescence, anthers, pistils, seeds collected 5 and 10 days after pollination, embryo and endosperm (Supplementary Table **S8**). The expression data was mapped on the phylogenetic tree so that we could determine group-specific conservation in expression patterns (Fig. **8**). Although all the genes exhibited diverse expression patterns, a few clade-specific inferences could be made. For instance, genes belonging to groups X and XI exhibited low-level expression, whereas, most of the genes belonging to groups XII, XIII and XIV exhibited high-level ubiquitous expression. Genes belonging to groups IV and V exhibited preferential accumulation in inflorescence tissues.

Some tissue-specific expression patterns were also distinct, such as *SbMYB130* showed embryo and endosperm specific accumulation; *SbMYB113* and *SbMYB046* of group VI exhibited pistil-preferential expression, whereas, *SbMYB133* of group VII exhibited the highest expression in anthers. Seed-specific *cis*-regulatory elements (RY-element, O2-site, GCN4_motif, HD-Zip 1, AACA_motif) were detected in the promoter region of *SbMYB133*. *SbMYB125* and *SbMY134* expressions was confined to endosperm tissue, while *SbMYB060* exhibited maximum accumulation in embryos. Similarly, *SbMYB066*, *089* and *098* expression was specific to post-emergence inflorescence, while *SbMYB061*, *078*, *081*, *094* and *128* exhibited preferential expression in the pre-emergence inflorescence. Many of the tissue-specific genes have also been implicated in the biosynthesis of phenylpropanoids or related compounds indicating that they may be regulating plant development by fine-tuning secondary metabolites.

3.7. Contrasting Roles of Sorghum MYB Proteins in Phenylpropanoid and/or Sugar Signalling

Although R1/2-type MYB family gene of rice, *OsMYBS1* and its ortholog in *Arabidopsis* have been demonstrated to

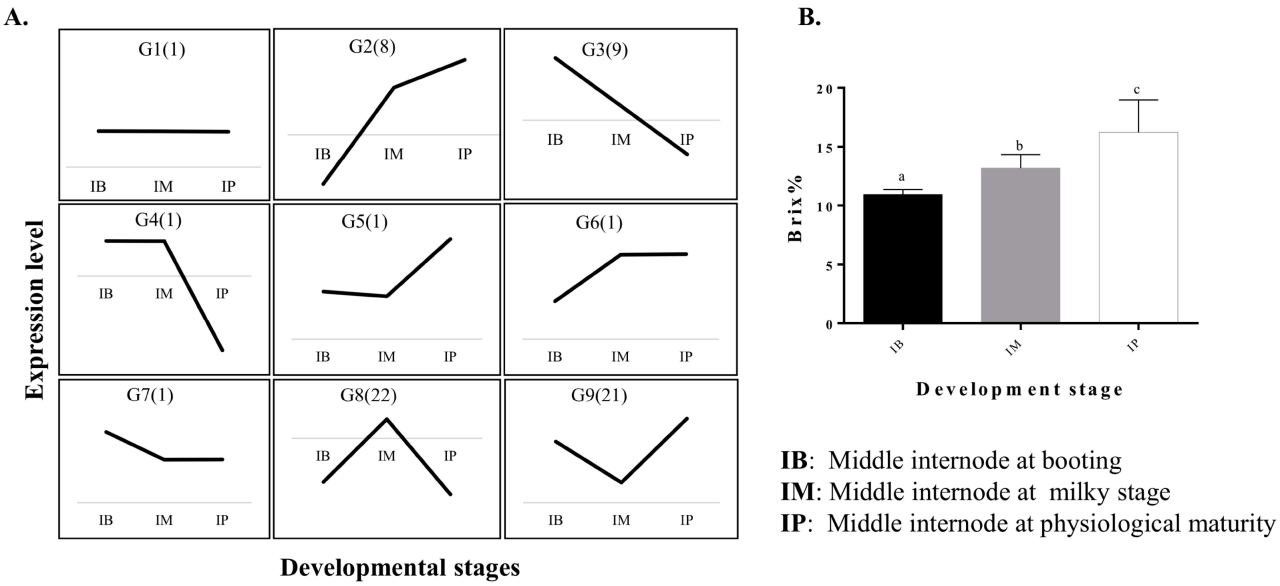

Fig. (9). Analysis of brix content and expression patterns exhibited by R2R3-MYB genes in three temporal stages of internodes of sweet sorghum cultivar SSV84. **A**. Clustering of expression patterns exhibited by MYB genes in three temporal stages of the internode. Cluster numbers (G) are given with each pattern and the numbers in brackets represent the number of R2R3-MYB genes exhibiting that pattern (IB internode at booting, IM internode at milky to soft dough and IP internode at physiological maturity). **B**. The bar graph shows relative brix content in middle internodes collected at booting, milky to soft dough and physiological maturity stages. The error bars represent the significance level analysed using ANOVA (Analysis Of Variance). (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

play role in sugar signalling, the role of R2R3-MYB genes in sugar signalling remains unexplored [97]. Some MYB genes have been shown to regulate phenylpropanoid biosynthesis through sugar signalling. For example, loss of function of *PAP1*/*MYB71* gene function in *Arabidopsis* impairs sugar-dependent inducibility of dihydroflavonol reductase, a key enzyme in anthocyanin biosynthesis [98, 99].

To examine the role of R2R3-MYB genes in sugar accumulation and/or signalling, we performed transcriptome profiling in internodes collected from sweet sorghum cultivar, SSV84. The brix assay with internodes revealed that the brix content increases from booting to the milky stage with a peak at physiological maturity [100] (Fig. **9A** and **B**). Two biological replicates of middle internodes were collected at booting, milky to a soft dough and physiological maturity stages and used for RNA sequencing. More than 500 million reads were generated out of which about 95% mapped to the sorghum genome (Supplementary Table **S9**). Out of 134 R2R3-MYB genes, annotated in this study, 65 exhibited FPKM >1 in sorghum internodes (Supplementary Table **S10**). Based on their expression patterns, they could be divided into nine distinct clusters (Fig. **9A**). Cluster 1 comprised genes exhibiting minimal variability in expression among all three stages. Cluster 2 contained nine genes whose expression gradually increased from booting to physiological maturity. Two of these genes fall in group V of the phylogenetic tree. Cluster 3 contained nine genes with the highest expression at booting stage that gradually declined in subsequent stages. Cluster 4, 5, 6 and 7 comprised one gene each with each of them exhibiting distinct expression patterns, while cluster 6 exhibited the highest expression in second and third internode stages. Both clusters 5 and 6 genes exhibited the highest expression in the last stage of the internode, whereas, clusters 4 and 7 genes exhibited highest expression in the booting stage (Fig. **9A**). Clusters 8 and 9 contained the maximum number of genes with 22 genes exhibiting cluster 8 pattern and 21 genes exhibiting cluster 9 pattern (Fig. **9A**). Cluster 8 genes showed the highest expression in internodes collected from milky to soft dough stage, whereas, cluster 9 genes exhibited contrasting pattern with peak expression in internodes collected from booting and physiological maturity stages.

Phylogenetic placement of the genes belonging to clusters 8 and 9 revealed a remarkably interesting pattern. Although it was common for different genes from the same clade to exhibit cluster 8 and cluster 9 pattern, most of the sorghum genes exhibiting cluster 8 pattern grouped with experimentally characterized genes known to act as activators of phenylpropanoid biosynthesis, whereas, those exhibiting cluster 9 pattern clubbed with known repressors. These results suggest that the genes belonging to clusters 8 and 9 play contrasting roles in primary or/and secondary metabolic pathways. Such opposite regulatory effects of R2R3-MYB genes have earlier been demonstrated in white spruce and *Arabidopsis*, proposed as a fine-tuning mechanism for regulation of plant metabolites [101].

CONCLUSION

R2R3-MYB transcription factors (TFs) form one of the largest in number and functionally heterogeneous transcription factor family in plants. They have been implicated in secondary cell wall development by regulating the expres-

sion of genes involved in the synthesis of cell wall components and phenylpropanoids [102]. The specialized secondary compounds regulated by R2R3-MYB genes are important to confer abiotic stress tolerance and resistance to pathogens and herbivores [103].

Using bioinformatic analyses coupled with in-depth expression profiling, we have generated a framework for hypothesizing R2R3-MYB gene functions in sorghum. Our results revealed selective expansion of specific clades attributed to gene duplications. Genes belonging to the same clade inside the major clades showed diverse functionalities indicating intra-clade functional diversification. While tissue-specific enrichment of R2R3-MYB sorghum genes provide candidates for engineering plant development, contrasting patterns in internodes of bioenergy and sweet sorghum provide targets for fine-tuning phenylpropanoid biosynthesis and, conditioning biofuel syndrome in sorghum cultivars.

AUTHOR'S CONTRIBUTIONS

MKS conceived and designed the study. VS and NK performed experiments, data analysis, prepared all the figures and participated in the manuscript writing. AD helped in data analysis. RS and MKS drafted and edited the final manuscript. All authors read and approved the manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

No Animals/humans were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data supporting the findings of the article is available in the NCBI-GEO at [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE148558], reference number [GSE148558].

FUNDING

This study was carried out with the research grants from Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Government of India (ECR/2016/001581) and (ECR/2015/000495) to RS and MKS. R.S. and M.K.S. also acknowledge the Department of Biotechnology (DBT), Government of India for Ramalingaswami fellowships and research grants. V.S. acknowledges ICMR for fellowship, N.K. thanks UGC for fellowship.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

REFERENCES

[1] Mathur, S.; Umakanth, A.V.; Tonapi, V.A.; Sharma, R.; Sharma, M.K. Sweet sorghum as biofuel feedstock: recent advances and available resources. *Biotechnol. Biofuels,* **2017**, *10*, 146. http://dx.doi.org/10.1186/s13068-017-0834-9 PMID: 28603553

[2] Somerville, C.; Youngs, H.; Taylor, C.; Davis, S.C.; Long, S.P. Feedstocks for lignocellulosic biofuels. *Science,* **2010**, *329*(5993), 790-792. http://dx.doi.org/10.1126/science.1189268 PMID: 20705851

[3] Rabara, R.C.; Tripathi, P.; Rushton, P.J. The potential of transcription factor-based genetic engineering in improving crop tolerance to drought. *OMICS,* **2014**, *18*(10), 601-614. http://dx.doi.org/10.1089/omi.2013.0177 PMID: 25118806

[4] Wang, H.; Wang, H.; Shao, H.; Tang, X. Recent advances in utilizing transcription factors to improve plant abiotic stress tolerance by transgenic technology. *Front. Plant Sci.,* **2016**, *7*, 67. http://dx.doi.org/10.3389/fpls.2016.00067 PMID: 26904044

[5] Gaponenko, A.K.; Shulga, O.A.; Mishutkina, Y.B.; Tsarkova, E.A.; Timoshenko, A.A.; Spechenkova, N.A. Perspectives of use of transcription factors for improving resistance of wheat productive varieties to abiotic stresses by transgenic technologies. *Russ. J. Genet.,* **2018**, *54*, 27-35. http://dx.doi.org/10.1134/S1022795418010039

[6] Du, H.; Wang, Y.B.; Xie, Y.; Liang, Z.; Jiang, S.J.; Zhang, S.S.; Huang, Y.B.; Tang, Y.X. Genome-wide identification and evolutionary and expression analyses of MYB-related genes in land plants. *DNA Res.,* **2013**, *20*(5), 437-448. http://dx.doi.org/10.1093/dnares/dst021

[7] Dubos, C.; Stracke, R.; Grotewold, E.; Weisshaar, B.; Martin, C.; Lepiniec, L. MYB transcription factors in Arabidopsis. *Trends Plant Sci.,* **2010**, *15*(10), 573-581. http://dx.doi.org/10.1016/j.tplants.2010.06.005 PMID: 20674465

[8] Ogata, K.; Morikawa, S.; Nakamura, H.; Sekikawa, A.; Inoue, T.; Kanai, H.; Sarai, A.; Ishii, S.; Nishimura, Y. Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell,* **1994**, *79*(4), 639-648. http://dx.doi.org/10.1016/0092-8674(94)90549-5 PMID: 7954830

[9] Ogata, K.; Kanei-Ishii, C.; Sasaki, M.; Hatanaka, H.; Nagadoi, A.; Enari, M.; Nakamura, H.; Nishimura, Y.; Ishii, S.; Sarai, A. The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nat. Struct. Biol.,* **1996**, *3*(2), 178-187. http://dx.doi.org/10.1038/nsb0296-178 PMID: 8564545

[10] Jia, L.; Clegg, M.T.; Jiang, T. Evolutionary dynamics of the DNA-binding domains in putative R2R3-MYB genes identified from rice subspecies indica and japonica genomes. *Plant Physiol.,* **2004**, *134*(2), 575-585. http://dx.doi.org/10.1104/pp.103.027201 PMID: 14966247

[11] Ogata, K.; Morikawa, S.; Nakamura, H.; Hojo, H.; Yoshimura, S.; Zhang, R.; Aimoto, S.; Ametani, Y.; Hirata, Z.; Sarai, A. Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-Myb. *Nat. Struct. Biol.,* **1995**, *2*(4), 309-320. http://dx.doi.org/10.1038/nsb0495-309 PMID: 7796266

[12] Braun, E.L.; Grotewold, E. Newly discovered plant c-myb-like genes rewrite the evolution of the plant myb gene family. *Plant Physiol.,* **1999**, *121*(1), 21-24. http://dx.doi.org/10.1104/pp.121.1.21 PMID: 10482656

[13] Dias, A.P.; Braun, E.L.; McMullen, M.D.; Grotewold, E. Recently duplicated maize R2R3 Myb genes provide evidence for distinct mechanisms of evolutionary divergence after duplication. *Plant Physiol.,* **2003**, *131*(2), 610-620. http://dx.doi.org/10.1104/pp.012047 PMID: 12586885

[14] Ito, M. Conservation and diversification of three-repeat Myb transcription factors in plants. *J. Plant Res.,* **2005**, *118*(1), 61-69. http://dx.doi.org/10.1007/s10265-005-0192-8 PMID: 15703854

[15] Feng, G.; Burleigh, J.G.; Braun, E.L.; Mei, W.; Barbazuk, W.B. Evolution of the 3R-MYB gene family in plants. *Genome Biol. Evol.,* **2017**, *9*(4), 1013-1029. http://dx.doi.org/10.1093/gbe/evx056 PMID: 28444194

[16] Du, H.; Liang, Z.; Zhao, S.; Nan, M.G.; Tran, L.S.P.; Lu, K.; Huang, Y.B.; Li, J.N. The evolutionary history of R2R3-MYB proteins across 50 eukaryotes: new insights into subfamily classification and expansion. *Sci. Rep.,* **2015**, *5*, 11037. http://dx.doi.org/10.1038/srep11037 PMID: 26047035

[17] Jiang, C.; Gu, J.; Chopra, S.; Gu, X.; Peterson, T. Ordered origin of the typical two- and three-repeat Myb genes. *Gene,* **2004**, *326*, 13-22. http://dx.doi.org/10.1016/j.gene.2003.09.049 PMID: 14729259

[18] Meneses, E.; Cárdenas, H.; Zárate, S.; Brieba, L.G.; Orozco, E.; López-Camarillo, C.; Azuara-Liceaga, E. The R2R3 Myb protein family in *Entamoeba histolytica*. *Gene,* **2010**, *455*(1-2), 32-42. http://dx.doi.org/10.1016/j.gene.2010.02.004 PMID: 20156532

[19] Yang, T.; Perasso, R.; Baroin-Tourancheau, A. Myb genes in ciliates: a common origin with the myb protooncogene? *Protist,* **2003**, *154*(2), 229-238. http://dx.doi.org/10.1078/143446103322166527 PMID: 13677450

[20] Ambawat, S.; Sharma, P.; Yadav, N.R.; Yadav, R.C. MYB transcription factor genes as regulators for plant responses: an overview. *Physiol. Mol. Biol. Plants,* **2013**, *19*, 307-321. http://dx.doi.org/10.1007/s12298-013-0179-1.

[21] Liu, J.; Osbourn, A.; Ma, P. MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol. Plant,* **2015**, *8*(5), 689-708. http://dx.doi.org/10.1016/j.molp.2015.03.012 PMID: 25840349

[22] Roy, S. Function of MYB domain transcription factors in abiotic stress and epigenetic control of stress response in plant genome. *Plant Signal. Behav.,* **2016**, *11*(1), e1117723. http://dx.doi.org/10.1080/15592324.2015.1117723 PMID: 26636625

[23] Poovaiah, C.R.; Bewg, W.P.; Lan, W.; Ralph, J.; Coleman, H.D. Sugarcane transgenics expressing MYB transcription factors show improved glucose release. *Biotechnol. Biofuels,* **2016**, *9*, 143. http://dx.doi.org/10.1186/s13068-016-0559-1 PMID: 27429646

[24] Xu, C.; Fu, X.; Liu, R.; Guo, L.; Ran, L.; Li, C.; Tian, Q.; Jiao, B.; Wang, B.; Luo, K. PtoMYB170 positively regulates lignin deposition during wood formation in poplar and confers drought tolerance in transgenic Arabidopsis. *Tree Physiol.,* **2017**, *37*(12), 1713-1726. http://dx.doi.org/10.1093/treephys/tpx093 PMID: 28985414

[25] Fornale, S.; Shi, X.; Chai, C.; Encina, A.; Irar, S.; Capellades, M.; Fuguet, E.; Torres, J.L.; Rovira, P.; Puigdomenech, P.; Rigau, J.; Grotewold, E.; Gray, J.; Caparros-Ruiz, D. ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux. *Plant J.,* **2010**, *64*(4), 633-644.

[26] Shen, H.; He, X.; Poovaiah, C.R.; Wuddineh, W.A.; Ma, J.; Mann, D.G.J.; Wang, H.; Jackson, L.; Tang, Y.; Stewart, C.N., Jr; Chen, F.; Dixon, R.A. Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. *New Phytol.,* **2012**, *193*(1), 121-136. http://dx.doi.org/10.1111/j.1469-8137.2011.03922.x PMID: 21988539

[27] Paterson, A.H.; Bowers, J.E.; Bruggmann, R.; Dubchak, I.; Grimwood, J.; Gundlach, H.; Haberer, G.; Hellsten, U.; Mitros, T.; Poliakov, A.; Schmutz, J.; Spannagl, M.; Tang, H.; Wang, X.; Wicker, T.; Bharti, A.K.; Chapman, J.; Feltus, F.A.; Gowik, U.; Grigoriev, I.V.; Lyons, E.; Maher, C.A.; Martis, M.; Narechania, A.; Otillar, R.P.; Penning, B.W.; Salamov, A.A.; Wang, Y.; Zhang, L.; Carpita, N.C.; Freeling, M.; Gingle, A.R.; Hash, C.T.; Keller, B.; Klein, P.; Kresovich, S.; McCann, M.C.; Ming, R.; Peterson, D.G.; Mehboob-ur-Rahman, ; Ware, D.; Westhoff, P.; Mayer, K.F.X.; Messing, J.; Rokhsar, D.S. The *Sorghum bicolor* genome and the diversification of grasses. *Nature,* **2009**, *457*(7229), 551-556. http://dx.doi.org/10.1038/nature07723 PMID: 19189423

[28] McCormick, R.F.; Truong, S.K.; Sreedasyam, A.; Jenkins, J.; Shu, S.; Sims, D.; Kennedy, M.; Amirebrahimi, M.; Weers, B.D.; McKinley, B.; Mattison, A.; Morishige, D.T.; Grimwood, J.; Schmutz, J.; Mullet, J.E. The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.,* **2018**, *93*(2), 338-354. http://dx.doi.org/10.1111/tpj.13781 PMID: 29161754

[29] Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; Sonnhammer, E.L.L.; Tate, J.; Punta, M. Pfam: the protein families database. *Nucleic Acids Res.,* **2014**, *42*(Database issue), D222-D230. http://dx.doi.org/10.1093/nar/gkt1223 PMID: 24288371

[30] Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; Rokhsar, D.S. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.,* **2012**, *40*(Database issue), D1178-D1186. http://dx.doi.org/10.1093/nar/gkr944 PMID: 22110026

[31] Gonzales, N.R.; Chitsaz, F.; Derbyshire, M.K.; Geer, L.; Gwadz, M.; Han, L.; He, J.; Hurwitz, D.I.; Lanczycki, C.J.; Lu, F.; Marchler, G.H.; Song, J.S.; Thanki, N.; Wang, Z.; Yamashita, R.A.; Zheng, C.; Bryant, S.H.; Marchler-Bauer, A. Manual curation in the conserved domain database. *Protein Sci.,* **2016**, *25*, 20.

[32] Letunic, I.; Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.,* **2018**, *46*(D1), D493-D496. http://dx.doi.org/10.1093/nar/gkx922 PMID: 29040681

[33] Voorrips, R.E. MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.,* **2002**, *93*(1), 77-78. http://dx.doi.org/10.1093/jhered/93.1.77 PMID: 12011185

[34] Yu, J.; Ke, T.; Tehrim, S.; Sun, F.; Liao, B.; Hua, W. PTGBase: an integrated database to study tandem duplicated genes in plants. *Database (Oxford),* **2015**, *2015,* pii: bav017. http://dx.doi.org/10.1093/database/bav017

[35] Lee, T.H.; Tang, H.; Wang, X.; Paterson, A.H. PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res.,* **2013**, *41*(Database issue), D1152-D1158. PMID: 23180799

[36] Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: an information aesthetic for comparative genomics. *Genome Res.,* **2009**, *19*(9), 1639-1645. http://dx.doi.org/10.1101/gr.092759.109 PMID: 19541911

[37] Chou, K.C.; Shen, H.B. Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS One,* **2010**, *5*(6), e11335. http://dx.doi.org/10.1371/journal.pone.0011335 PMID: 20596258

[38] Blum, T.; Briesemeister, S.; Kohlbacher, O. MultiLoc2: integrating phylogeny and Gene Ontology terms improves subcellular protein localization prediction. *BMC Bioinformatics,* **2009**, *10*, 274. http://dx.doi.org/10.1186/1471-2105-10-274 PMID: 19723330

[39] Yu, C.S.; Chen, Y.C.; Lu, C.H.; Hwang, J.K. Prediction of protein subcellular localization. *Proteins,* **2006**, *64*(3), 643-651. http://dx.doi.org/10.1002/prot.21018 PMID: 16752418

[40] Almagro Armenteros, J.J.; Sønderby, C.K.; Sønderby, S.K.; Nielsen, H.; Winther, O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics,* **2017**, *33*(21), 3387-3395. http://dx.doi.org/10.1093/bioinformatics/btx431 PMID: 29036616

[41] Horton, P.; Park, K.J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C.J.; Nakai, K. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.,* **2007**, *35*(Web Server issue), W585-W587. PMID: 17517783

[42] O'Brien, K.P.; Remm, M.; Sonnhammer, E.L.L. Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res.,* **2005**, *33*(Database issue), D476-D480. http://dx.doi.org/10.1093/nar/gki107 PMID: 15608241

[43] Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.,* **2013**, *30*(4), 772-780. http://dx.doi.org/10.1093/molbev/mst010 PMID: 23329690

[44] Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res.,* **2004**, *14*(6), 1188-1190. http://dx.doi.org/10.1101/gr.849004 PMID: 15173120

[45] Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. In: *The Proteomics Protocols Handbook*; Humana Press: Totowa, NJ, **2005**, pp. 571-607. http://dx.doi.org/10.1385/1-59259-890-0:571

[46] 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*(8), 1296-1297. http://dx.doi.org/10.1093/bioinformatics/btu817 PMID: 25504850

[47] Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.,* **2009**, *37*(Web Server issue), W202-W208. http://dx.doi.org/10.1093/nar/gkp335 PMID: 19458158

[48] Bastian, M.; Heymann, S.; Jacomy, M. *Third international AAAI conference on weblogs and social media,* **2009**.

[49] Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics,* **2009**, *25*(9), 1189-1191. http://dx.doi.org/10.1093/bioinformatics/btp033 PMID: 19151095

[50] Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.,* **2016**, *33*(7), 1870-1874. http://dx.doi.org/10.1093/molbev/msw054 PMID: 27004904

[51] Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.,* **2019**, *47*(W1), W256-W259. http://dx.doi.org/10.1093/nar/gkz239 PMID: 30931475

[52] Davidson, R.M.; Gowda, M.; Moghe, G.; Lin, H.; Vaillancourt, B.; Shiu, S.H.; Jiang, N.; Robin Buell, C. Comparative transcriptomics of three poaceae species reveals patterns of gene expression evolution. *Plant J.,* **2012**, *71*(3), 492-502. http://dx.doi.org/10.1111/j.1365-313X.2012.05005.x.

[53] Kebrom, T.H.; McKinley, B.; Mullet, J.E. Dynamics of gene expression during development and expansion of vegetative stem internodes of bioenergy sorghum. *Biotechnol. Biofuels,* **2017**, *10*, 159. http://dx.doi.org/10.1186/s13068-017-0848-3 PMID: 28649278

[54] Rao, S.S.; Patil, J.V.; Prasad, P.V.V.; Reddy, D.C.S.; Mishra, J.S.; Umakanth, A.V.; Reddy, B.V.S.; Kumar, A.A. Sweet sorghum planting effects on stalk yield and sugar quality in semi-arid tropical environment. *Agron. J.,* **2013**, *105*, 1458. http://dx.doi.org/10.2134/agronj2013.0156

[55] Zhang, J.; Jiang, F.; Shen, Y.; Zhan, Q.; Bai, B.; Chen, W.; Chi, Y. Transcriptome analysis reveals candidate genes related to phosphorus starvation tolerance in sorghum. *BMC Plant Biol.,* **2019**, *19*(1), 306. http://dx.doi.org/10.1186/s12870-019-1914-8 PMID: 31296169

[56] Pertea, M.; Kim, D.; Pertea, G.M.; Leek, J.T.; Salzberg, S.L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.,* **2016**, *11*(9), 1650-1667. http://dx.doi.org/10.1038/nprot.2016.095 PMID: 27560171

[57] Stelpflug, S.C.; Sekhon, R.S.; Vaillancourt, B.; Hirsch, C.N.; Buell, C.R. An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *Plant Genome,* **2016**, *9*(1). http://dx.doi.org/10.3835/plantgenome2015.04.0025

[58] Heine, G.F.; Hernandez, J.M.; Grotewold, E. Two cysteines in plant R2R3 MYB domains participate in REDOX-dependent DNA binding. *J. Biol. Chem.,* **2004**, *279*(36), 37878-37885. http://dx.doi.org/10.1074/jbc.M405166200 PMID: 15237103

[59] 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*(6), e37463. http://dx.doi.org/10.1371/journal.pone.0037463 PMID: 22719841

[60] 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*(10), 1777-1788. http://dx.doi.org/10.1093/pcp/pcx115 PMID: 29016897

[61] Wang, Z.; Tang, J.; Hu, R.; Wu, P.; Hou, X.L.; Song, X.M.; Xiong, A.S. Genome-wide analysis of the R2R3-MYB transcription factor genes in Chinese cabbage (*Brassica rapa ssp. pekinensis*) reveals their stress and hormone responsive patterns. *BMC Genomics,* **2015**, *16*, 17. http://dx.doi.org/10.1186/s12864-015-1216-y PMID: 25613160

[62] Zargarian, L.; Le Tilly, V.; Jamin, N.; Chaffotte, A.; Gabrielsen, O.S.; Toma, F.; Alpert, B. Myb-DNA recognition: role of tryptophan residues and structural changes of the minimal DNA binding domain of c-Myb. *Biochemistry,* **1999**, *38*(6), 1921-1929. http://dx.doi.org/10.1021/bi981199j PMID: 10026273

[63] Jin, H.; Martin, C. Multifunctionality and diversity within the plant MYB-gene family. *Plant Mol. Biol.,* **1999**, *41*(5), 577-585. http://dx.doi.org/10.1023/A:1006319732410 PMID: 10645718

[64] Kranz, H.; Scholz, K.; Weisshaar, B. c-MYB oncogene-like genes encoding three MYB repeats occur in all major plant lineages. *Plant J.,* **2000**, *21*, 231-235.

[65] Millard, P.S.; Kragelund, B.B.; Burow, M. R2R3 MYB transcription factors - functions outside the DNA-binding domain. *Trends Plant Sci.,* **2019**, 24(10), 934-946, http://dx.doi.org/10.1016/j.tplants.2019.07.003

[66] Fernández-Calvo, P.; Chini, A.; Fernández-Barbero, G.; Chico, J.M.; Gimenez-Ibanez, S.; Geerinck, J.; Eeckhout, D.; Schweizer, F.; Godoy, M.; Franco-Zorrilla, J.M.; Pauwels, L.; Witters, E.; Puga, M.I.; Paz-Ares, J.; Goossens, A.; Reymond, P.; De Jaeger, G.; Solano, R. The Arabidopsis bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. *Plant Cell,* **2011**, *23*(2), 701-715. http://dx.doi.org/10.1105/tpc.110.080788 PMID: 21335373

[67] Liu, Y.; Chakrabortee, S.; Li, R.; Zheng, Y.; Tunnacliffe, A. Both plant and animal LEA proteins act as kinetic stabilisers of polyglutamine-dependent protein aggregation. *FEBS Lett.,* **2011**, *585*(4), 630-634. http://dx.doi.org/10.1016/j.febslet.2011.01.020 PMID: 21251910

[68] Schaefer, M.H.; Wanker, E.E.; Andrade-Navarro, M.A. Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks. *Nucleic Acids Res.,* **2012**, *40*(10), 4273-4287. http://dx.doi.org/10.1093/nar/gks011 PMID: 22287626

[69] Pelassa, I.; Corà, D.; Cesano, F.; Monje, F.J.; Montarolo, P.G.; Fiumara, F. Association of polyalanine and polyglutamine coiled coils mediates expansion disease-related protein aggregation and dysfunction. *Hum. Mol. Genet.,* **2014**, *23*(13), 3402-3420. http://dx.doi.org/10.1093/hmg/ddu049 PMID: 24497578

[70] Briggs, G.S.; Mahdi, A.A.; Wen, Q.; Lloyd, R.G. DNA binding by the substrate specificity (wedge) domain of RecG helicase suggests a role in processivity. *J. Biol. Chem.,* **2005**, *280*(14), 13921-13927. http://dx.doi.org/10.1074/jbc.M412054200 PMID: 15695524

[71] Alvarez, M.; Estivill, X.; de la Luna, S. DYRK1A accumulates in splicing speckles through a novel targeting signal and induces speckle disassembly. *J. Cell Sci.,* **2003**, *116*(Pt 15), 3099-3107. http://dx.doi.org/10.1242/jcs.00618 PMID: 12799418

[72] Hoque, M.; Young, T.M.; Lee, C.G.; Serrero, G.; Mathews, M.B.; Pe'ery, T. The growth factor granulin interacts with cyclin T1 and modulates P-TEFb-dependent transcription. *Mol. Cell. Biol.,* **2003**, *23*(5), 1688-1702. http://dx.doi.org/10.1128/MCB.23.5.1688-1702.2003 PMID: 12588988

[73] Gamsjaeger, R.; Liew, C.K.; Loughlin, F.E.; Crossley, M.; Mackay, J.P. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem. Sci.,* **2007**, *32*(2), 63-70. http://dx.doi.org/10.1016/j.tibs.2006.12.007 PMID: 17210253

[74] Gall, A.R.; Datsenko, K.A.; Figueroa-Bossi, N.; Bossi, L.; Masuda, I.; Hou, Y.M.; Csonka, L.N. Mg^{2+} regulates transcription of mgtA in *Salmonella Typhimurium via* translation of proline codons during synthesis of the MgtL peptide. *Proc. Natl. Acad. Sci. USA,* **2016**, *113*(52), 15096-15101. http://dx.doi.org/10.1073/pnas.1612268113 PMID: 27849575

[75] Peil, L.; Starosta, A.L.; Lassak, J.; Atkinson, G.C.; Virumäe, K.; Spitzer, M.; Tenson, T.; Jung, K.; Remme, J.; Wilson, D.N. Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. *Proc. Natl. Acad. Sci. USA,* **2013**, *110*(38), 15265-15270. http://dx.doi.org/10.1073/pnas.1310642110 PMID: 24003132

[76] Qi, F.; Motz, M.; Jung, K.; Lassak, J.; Frishman, D. Evolutionary analysis of polyproline motifs in *Escherichia coli* reveals their regulatory role in translation. *PLOS Comput. Biol.,* **2018**, *14*(2), e1005987. http://dx.doi.org/10.1371/journal.pcbi.1005987 PMID: 29389943

[77] Liu, C.; Xie, T.; Chen, C.; Luan, A.; Long, J.; Li, C.; Ding, Y.; He, Y. Genome-wide organization and expression profiling of the R2R3-MYB transcription factor family in pineapple (*Ananas comosus*). *BMC Genomics,* **2017**, *18*(1), 503. http://dx.doi.org/10.1186/s12864-017-3896-y PMID: 28668094

[78] Weng, J.K.; Li, X.; Bonawitz, N.D.; Chapple, C. Emerging strategies of lignin engineering and degradation for cellulosic biofuel production. *Curr. Opin. Biotechnol.,* **2008**, *19*(2), 166-172. http://dx.doi.org/10.1016/j.copbio.2008.02.014 PMID: 18403196

[79] Zhong, R.; Lee, C.; Zhou, J.; McCarthy, R.L.; Ye, Z.H. A battery of transcription factors involved in the regulation of secondary cell

wall biosynthesis in Arabidopsis. *Plant Cell,* **2008**, *20*(10), 2763-2782.
http://dx.doi.org/10.1105/tpc.108.061325 PMID: 18952777

[80] Zhong, R.; Ye, Z.H. Transcriptional regulation of lignin biosynthesis. *Plant Signal. Behav.,* **2009**, *4*(11), 1028-1034.
http://dx.doi.org/10.4161/psb.4.11.9875 PMID: 19838072

[81] 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*(1), 248-266.
http://dx.doi.org/10.1105/tpc.108.063321 PMID: 19122102

[82] Wang, H.Z.; Dixon, R.A. On-off switches for secondary cell wall biosynthesis. *Mol. Plant,* **2012**, *5*(2), 297-303.
http://dx.doi.org/10.1093/mp/ssr098 PMID: 22138968

[83] Yang, K.; Li, Y.; Wang, S.; Xu, X.; Sun, H.; Zhao, H.; Li, X.; Gao, Z. Genome-wide identification and expression analysis of the MYB transcription factor in moso bamboo (*Phyllostachys edulis*). *PeerJ,* **2019**, *6,* e6242.
http://dx.doi.org/10.7717/peerj.6242 PMID: 30648007

[84] Zhao, K.; Bartley, L.E. Comparative genomic analysis of the R2R3 MYB secondary cell wall regulators of Arabidopsis, poplar, rice, maize, and switchgrass. *BMC Plant Biol.,* **2014**, *14*, 135.
http://dx.doi.org/10.1186/1471-2229-14-135 PMID: 24885077

[85] Fornalé, S.; Sonbol, F.M.; Maes, T.; Capellades, M.; Puigdomènech, P.; Rigau, J.; Caparrós-Ruiz, D. Down-regulation of the maize and *Arabidopsis thaliana* caffeic acid O-methyl-transferase genes by two new maize R2R3-MYB transcription factors. *Plant Mol. Biol.,* **2006**, *62*(6), 809-823.
http://dx.doi.org/10.1007/s11103-006-9058-2 PMID: 16941210

[86] Patzlaff, A.; Newman, L.J.; Dubos, C.; Whetten, R.W.; Smith, C.; McInnis, S.; Bevan, M.W.; Sederoff, R.R.; Campbell, M.M. Characterisation of Pt MYB1, an R2R3-MYB from pine xylem. *Plant Mol. Biol.,* **2003**, *53*(4), 597-608.
http://dx.doi.org/10.1023/B:PLAN.0000019066.07933.d6 PMID: 15010621

[87] Scully, E.D.; Gries, T.; Palmer, N.A.; Sarath, G.; Funnell-Harris, D.L.; Baird, L.; Twigg, P.; Seravalli, J.; Clemente, T.E.; Sattler, S.E. Overexpression of SbMyb60 in *Sorghum bicolor* impacts both primary and secondary metabolism. *New Phytol.,* **2018**, *217*(1), 82-104.
http://dx.doi.org/10.1111/nph.14815 PMID: 28944535

[88] Zhong, R.; Richardson, E.A.; Ye, Z.H. The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in Arabidopsis. *Plant Cell,* **2007**, *19*(9), 2776-2792.
http://dx.doi.org/10.1105/tpc.107.053678 PMID: 17890373

[89] Park, M.Y.; Kang, J.Y.; Kim, S.Y. Overexpression of AtMYB52 confers ABA hypersensitivity and drought tolerance. *Mol. Cells,* **2011**, *31*(5), 447-454.
http://dx.doi.org/10.1007/s10059-011-0300-7 PMID: 21399993

[90] Awika, J.M. Sorghum flavonoids: unusual compounds with promising implications for health. ACS Publications; Washigton: DC, **2011**, 171-200.
http://dx.doi.org/10.1021/bk-2011-1089.ch009

[91] Boddu, J.; Svabek, C.; Ibraheem, F.; Jones, A.D.; Chopra, S. Characterization of a deletion allele of a sorghum Myb gene yellow seed1 showing loss of 3-deoxyflavonoids. *Plant Sci.,* **2005**, *169*, 542-552.
http://dx.doi.org/10.1016/j.plantsci.2005.05.007

[92] Makkena, S.; Lee, E.; Sack, F.D.; Lamb, R.S. The R2R3 MYB transcription factors FOUR LIPS and MYB88 regulate female reproductive development. *J. Exp. Bot.,* **2012**, *63*(15), 5545-5558.
http://dx.doi.org/10.1093/jxb/ers209 PMID: 22915737

[93] Baumann, K.; Perez-Rodriguez, M.; Bradley, D.; Venail, J.; Bailey, P.; Jin, H.; Koes, R.; Roberts, K.; Martin, C. Control of cell and petal morphogenesis by R2R3 MYB transcription factors. *Development,* **2007**, *134*(9), 1691-1701.
http://dx.doi.org/10.1242/dev.02836 PMID: 17376813

[94] Yang, H.; Xue, Q.; Zhang, Z.; Du, J.; Yu, D.; Huang, F. GmMYB181, a soybean R2R3-MYB protein, increases branch number in transgenic arabidopsis. *Front. Plant Sci.,* **2018**, *9*, 1027.
http://dx.doi.org/10.3389/fpls.2018.01027 PMID: 30065741

[95] Borg, M.; Brownfield, L.; Khatab, H.; Sidorova, A.; Lingaya, M.; Twell, D. The R2R3 MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in Arabidopsis. *Plant Cell,* **2011**, *23*(2), 534-549.
http://dx.doi.org/10.1105/tpc.110.081059 PMID: 21285328

[96] Zhang, Y.; Liang, W.; Shi, J.; Xu, J.; Zhang, D. MYB56 encoding a R2R3 MYB transcription factor regulates seed size in *Arabidopsis thaliana*. *J. Integr. Plant Biol.,* **2013**, *55*(11), 1166-1178.
http://dx.doi.org/10.1111/jipb.12094 PMID: 23911125

[97] Chen, Y.S.; Chao, Y.C.; Tseng, T.W.; Huang, C.K.; Lo, P.C.; Lu, C.A. Two MYB-related transcription factors play opposite roles in sugar signaling in Arabidopsis. *Plant Mol. Biol.,* **2017**, *93*(3), 299-311.
http://dx.doi.org/10.1007/s11103-016-0562-8 PMID: 27866313

[98] Kranz, H.; Denekamp, M.; Greco, R.; Jin, H.; Leyva, A.; Meissner, R.; Petroni, K.; Urzainqui, A.; Bevan, M.; Martin, C.; Smeekens, S.; Tonelli, C.; Paz-Ares, J.; Weisshaar, B. Towards functional characterisation of the members of the R2R3-MYB gene family from *Arabidopsis thaliana*. *Plant J.,* **1998**, *16*(2), 263-276.
http://dx.doi.org/10.1046/j.1365-313x.1998.00278.x

[99] Borevitz, J.O.; Xia, Y.; Blount, J.; Dixon, R.A.; Lamb, C. Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell,* **2000**, *12*(12), 2383-2394.
http://dx.doi.org/10.1105/tpc.12.12.2383 PMID: 11148285

[100] Shukla, S.; Felderhoff, T.J.; Saballos, A.; Vermerris, W. The relationship between plant height and sugar accumulation in the stems of sweet sorghum (*Sorghum bicolor (L.) Moench*). *Field Crops Res.,* **2017**, *203*, 181-191.
http://dx.doi.org/10.1016/j.fcr.2016.12.004

[101] Bomal, C.; Duval, I.; Giguère, I.; Fortin, É.; Caron, S.; Stewart, D.; Boyle, B.; Séguin, A.; MacKay, J.J. Opposite action of R2R3-MYBs from different subgroups on key genes of the shikimate and monolignol pathways in spruce. *J. Exp. Bot.,* **2014**, *65*(2), 495-508.
http://dx.doi.org/10.1093/jxb/ert398 PMID: 24336492

[102] Qian, M.; Kalbina, I.; Rosenqvist, E.; Jansen, M.A.K.; Teng, Y.; Strid, Å. UV regulates the expression of phenylpropanoid biosynthesis genes in cucumber (*Cucumis sativus L.*) in an organ and spectrum dependent manner. *Photochem. Photobiol. Sci.,* **2019**, *18*(2), 424-433.
http://dx.doi.org/10.1039/C8PP00480C PMID: 30628617

[103] Ibraheem, F.; Gaffoor, I.; Tan, Q.; Shyu, C.R.; Chopra, S.; Ibraheem, F.; Gaffoor, I.; Tan, Q.; Shyu, C.R.; Chopra, S. A sorghum MYB transcription factor induces 3-deoxyanthocyanidins and enhances resistance against leaf blights in maize. *Molecules,* **2015**, *20*(2), 2388-2404.
http://dx.doi.org/10.3390/molecules20022388 PMID: 25647576