

# R2R3-MYB Transcription Factor Family in *Boehmeria nivea* (L.) Gaudich: Potential Associations between Cd<sup>2+</sup> Stress and Flavonoid Metabolism

**Xinkang Feng**

Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205

**Aminu Shehu Abubakar**

Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205

**Kunmei Chen**

Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205

**Chunming Yu**

Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205

**Aiguo Zhu**

Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205

**Jikang Chen**

Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205

**Gang Gao**

Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205

**Xiaofei Wang**

Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205

**Pan Mou**

Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205

**Ping Chen** (✉ [chenping02@caas.cn](mailto:chenping02@caas.cn))

Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205

---

## Research Article

**Keywords:** Ramie, MYB transcription factor, Expressed profiles, Fiber development, Cadmium stress, Protein interaction

**Posted Date:** September 21st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-871051/v1>

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

[Read Full License](#)

# Abstract

**Background:** The characterization of gene families especially MYB being the largest transcription factor (TFs) families in plants is a crucial step for functional studies. The completion of ramie genome sequencing provides a great opportunity to investigate the organization and evolutionary traits of ramie MYB genes at the genome-wide level.

**Results:** A total of 105 BnGR2R3-MYB genes were identified in ramie and divided into 35 distinct subfamilies according to phylogeny divergence and sequences similarity. These genes were unevenly distributed among 14 chromosomes. Collinearity analysis showed that the segmental and tandem duplication events is the dominant form of the gene family expansion, and duplications prominent in distal telomeric regions. 88 BnGR2R3-MYB genes showed syntenic relationship with those in *Apocynum venetum*, followed by *Cannabis sativa* (58), *Arabidopsis* (53), maize (8) and rice (4). 103 of the 105 BnGMYB proteins were predicted nuclear subcellular, the remaining two were in either chloroplast or cytoplasm. The combination of transcriptome data and phylogenetic tree allows us to propose some powerful MYB candidates that might be involved in the regulation of secondary wall-associated cellulose synthases (*BnGMYB14*) secondary cell wall thickening (*BnGMYB66/67*) and flavonoids synthesis (*BnGMYB60*). The MYBs subgroups involve in regulating anthocyanin were different from arabidopsis and tomato pointing that BnGMYB in other groups play a role in regulating anthocyanin synthesis. qPCR results revealed 8 MYB TFs candidate genes for cadmium stress in ramie. There was an increased in synthesis of procyanidins under the cadmium stress, which suggest a new regulatory pathway in response to the stress. The predicted network identifies the interface between flavonoid metabolic pathways and adversity stress, and found evidence for the involvement of flavonoid synthetic pathways in the stress regulation.

**Conclusions:** This work provides a basic understanding of BnGR2R3-MYB gene family characteristics and provides valuable information that may contribute in improving the tolerance of ramie to cadmium stress and fiber quality.

## Background

Transcription factors (TFs) also called trans-acting factors are DNA-binding proteins that can specifically interact with the cis-acting elements of eukaryotic genes and activate or inhibit the transcription of genes. From the perspective of protein structure, these TFs are generally composed of four functional regions: DNA binding domain, transcription regulation domain (including activation domain or inhibition domain), oligomerization site, and nuclear localization signal[1]. Structure determines function, so most transcription factors can be classified into different gene families according to their conserved sequence structure. The MYB TFs, which are distinguished by the highly conserved MYB domain, comprise one of the largest TF families in the plant kingdom[2] and are usually composed of one to four adjacent imperfect tandem repeats which share a highly conserved MYB binding domain at the N-terminus. Each repeat is approximately 50–53 amino acid residues in length, containing three  $\alpha$ -helices that together

form a helix-turn-helix (HTH) secondary structure that interacts with the major DNA at the specific recognition site C/TAACG/TG during transcription[3]. On the contrary, the C-terminal region is highly variable that leads to the wide range of regulatory roles of the MYB gene family[4]. According to the number of the adjacent MYB repeats, MYB genes are distinguished into the following four subfamilies: 1R-MYB, R2R3-MYB, R1R2R3-MYB and 4R-MYB. Generally, the R2R3-MYB members, which have diverse functions, are the most extensively studied[5].

The MYB TF is present in all eukaryotes. *Zea mays* C1, the first plant MYB gene to be cloned and functionally characterized, is responsible for regulation of anthocyanin biosynthesis[6]. Nowadays, with the advance of plant genome-wide association and molecular biology methods, an increasing number of studies have shown that the MYB proteins are involved in the regulation of plant primary and secondary metabolism, the control of plant development, and the partake in response to various biotic and abiotic stresses[7–10]. For example, *CsTSl* regulates the biosynthesis of theanine in different organs and tissues of tea plants[11]. While *CaMYB31* is reported to involve in regulation of biosynthesis of capsaicinoids in *Capsicum annum*[12], *DkMyb4* controls proanthocyanidin biosynthesis in persimmon[13]. A number of MYB TFs regulating anthocyanin biosynthetic pathway have been identified from many species, such as *AtMYB75*, *AtMYB90*, *AtMYB113* and *AtMYB114* in *Arabidopsis*[14–16], *CsRuby1* and *CsMYB3* in citrus[17] and *SIMYB12* in tomato[18]. Many of the reported MYB TFs have been shown to control cellular morphogenesis, secondary cell wall biosynthesis, meristem formation and cell cycle regulation[19–22]. A rice R2R3-MYB gene (*OsCSA*) directly triggers expression of sugar partitioning and metabolic genes during pollen and seed development[23]. *GbMYB5* confers drought tolerance in cotton and transgenic tobacco plants[24]. Another rice MYB (*OsMYB3R-2*) exhibited enhanced cold tolerance by alteration in cell cycle and ectopic expression of stress genes[25].

Ramie is one of the most important fiber crops in the family Urticaceae. It is also widely used as feed and other industrial raw materials[26–28]. The completion of the genome sequencing project at the chromosome level of ramie provides a valuable platform for revealing the genomic fabric of ramie R2R3-MYB gene family and exploring the evolutionary characteristics among different species[26, 29]. R2R3-MYB transcription factor plays an important role in plant specific processes making the MYB TFs a major research topic in many plants. Despite that, very limited reports are present to date on the functional characteristics of R2R3-MYB transcription factor in ramie. Therefore, this study was carried out to elucidate the function, evolution and expression profile of BnGR2R3-MYB transcription factor in ramie based on sequencing data. Comprehensive investigations involving gene structures, gene duplications, chromosomal locations, phylogeny and cis-acting elements of the BnGR2R3-MYB genes were performed. Expression profiles under cadmium stress and different stages of fiber development were used to focus and discuss the regulatory role of BnGR2R3-MYB genes.

## Materials And Methods

# Identification of MYB protein in *Boehmeria nivea* (L.) Gaudich

To identify more comprehensive MYB TFs, 168 MYB family sequences and 97 MYB-related family sequences (*A. thaliana*), 208 MYB family sequences and 85 MYB-related family sequences (*G. arboreum*), 130 MYB family sequences and 106 MYB-related family sequences (*O. sativa*) were retrieved from the PlantTFDB (<http://planttfdb.gao-lab.org/>). Protein sequences of these plant species were used as reference to filter the possible sequence of target specie by Blast Wrapper, and the expectation cut-off value (E-value) of 0.01 set as threshold significance. Subsequently, the MYB sequences obtained were aligned to Swissport Database / Reference Species Whole Protein Sequence Library, and screened by annotation information. The resulting datasets of MYB sequences were confirmed based on the completeness of MYB domains using Pfam, CDD and SMART with an E-value of 1e-2. To correct for deletion of some conserved sites, we used FGENESH-M (<http://www.softberry.com/berry.phtml>) to predict multiple variants potential genes in genomic DNA. All ramie R2R3-MYB proteins were manually inspected to ensure that the putative gene models contained two complete MYB domains and ultimately identified and classified all members of the MYB family in *Boehmeria nivea*. Length of sequences, molecular weights, GRAVY and pI of MYB proteins were analyzed using the ExpASy online tool[80] and subcellular localization using the pLoc-mPlant tools for Batch Prediction.

## Sequence analysis and structural characterization of BnGR2R3-MYB genes

Multiple sequence alignments of the MYB domains sequences were performed using MEGA-X with default parameters. The deduced amino acid sequences in MYB motifs were then adjusted manually with Jalview software. WEBLOGO was used to show up the sequence logos of R2 and R3 MYB domain repeats with default setting. The exon-intron organizations of the BnGR2R3-MYB genes, including intron distribution patterns, phases, intro-exon boundaries and highlighted region of the MYB domains (R2, R3), were graphically displayed by TBtools[81]. The conserved motifs of the ramie MYB proteins were predicted by using the MEME[82] version 5.1.0 online program, with optimized parameters: zero or one per sequence; the maximum number of motifs, 20 respectively, and visualized by using TBtools.

## Analyses of chromosome distribution, gene duplication and synteny for BnGR2R3-MYB genes

The detailed chromosome distribution information of each BnGR2R3-MYB genes was obtained from the ramie genome annotation documents. Circos[81] was used to locate all the BnGR2R3-MYB genes on the ramie chromosome. Tandem, segmental duplication and collinearity within species were obtained by using Multiple Collinearity Scantoolkit (MCscanX). The synteny blocks of the orthologous R2R3-MYB genes between ramie and other species were obtained by using Multiple Collinearity Scan toolkit (MCscanX)[82]. The above results were visualized by TBtools[79] (Multiple Synteny Plot function). Non-

synonymous and synonymous substitution of each duplicated BnGMYB genes were calculated using KaKs\_Calculator 2.0[83].

## **Phylogenetic analysis and classification of ramie BnGR2R3-MYB proteins and WGDI analysis**

Multiple sequence alignments of the R2R3-MYB proteins from ramie and *Arabidopsis thaliana* were performed using MEGA 7.0 with default parameters[84]. MEGA 7.0 was used to construct a neighborhood linkage (NJ) phylogenetic tree with the following parameters: Poisson model; pairwise deletion; and 1000 bootstrap replications. R2R3-MYB proteins from other species (*Arabidopsis*[37], rice[85], maize[86], tomato[32], pineapple[42]) were obtained based on the description in corresponding literatures, and adopted the same method as above.

To detect WGD events, the paralogous genes of ramie were detected using all-vs-all homology searches in BLASTP with an E-value threshold of 1e-5. Syntenic blocks within a genome were identified based on the detected homologous gene pairs using MCscan.

## **Identification of the cis-elements of BnGR2R3-MYB genes and GO annotation**

To investigate the promoter regions of the BnGR2R3-MYB genes, 1000 bp genomic DNA upstream sequences of each of the 105 BnGR2R3-MYB genes were selected and the cis-elements predicted using PlantCare ([http://bioinformatics.psb.ugent.be/web\\_tools/plantcare/html/](http://bioinformatics.psb.ugent.be/web_tools/plantcare/html/)). Response class elements such as light-responsiveness, plant growth, stress-responsiveness and phytohormone-responsiveness were filtered and preserved. The cis-acting elements were visualized in TBtools, and each elements classified and the result displayed using heatmap. Functional annotation of BnGMYB proteins was performed in Blast2GO Tool by first blasting in BLASTP using Swissport database, followed by mapping Gene Ontology (GO) terms, and finally visualization with R ggpolt2.

## **Expression patterns of BnGR2R3-MYB genes in the representative tissues of the ramie**

To explore the expression patterns of the BnGR2R3-MYB genes, 9 samples involving different tissues at different developmental stages were used, RNA-Seq data was obtained from Genome Sequence Archive database and transcript abundance of the BnGR2R3-MYB genes was calculated as fragments per kilobase of exon model per million mapped reads (FPKM).  $\text{Log}_2(\text{FPKM} + 1)$  values were displayed according to the color code. Detailed FPKM values were listed in Table S7.

## **Stress treatment ( $\text{Cd}^{+2}$ ) of ramie under hydroponic conditions**

Two ramie varieties (ZZ\_1; HX\_1) were used for cadmium stress treatment. Shoots from each of the varieties at similar growth stage during the same period were selected for hydroponic cutting according to

Gao et al., [87]. These were monitored for 15 days and seedlings with inconsistent growth were removed. CdCl<sub>2</sub> to a final concentration of 50mg/ L was added and tissues were sampled at 0, 12, 24, and 48 h after the treatments, and three separate biological replicates. All samples were frozen quickly in liquid nitrogen, and stored at -80°C until used.

## **Prediction of BnGMYB protein–protein interaction network**

The orthologous relationships of eight selected cadmium-responsive MYB regulatory genes were determined between *Arabidopsis thaliana* and ramie using OrthoVeen2[88]. Prediction of interactions between BnGMYB proteins and other proteins based on the *Arabidopsis* homologs was obtained using the online program STRING version 11.5 with high confidence > 0.700[89], filtered genes were used to construct the correlation network. The interaction network was visualized by Cytoscape v3.8.2.

## **RNA isolation and expression analysis of BnR2R3-GMYB genes**

Total RNA was extracted using EasySpin Plus plant RNA rapid extraction Kit (Aid-lab Biotechnologies Co., Ltd). The RNA was reverse-transcribed (Thermo Scientific RevertAid First cDNA Synthesis Kit) into cDNA and quantitative RT-PCR (qPCR) analysis conducted using gene specific primers (Table S8). 18s gene was used as internal control. The qPCR was conducted using SYBR Green Premix Pro Taq HS qPCR Kit (Accurate Biotechnology Co., Ltd) on a CFX96 Touch Deep Well Real-Time PCR System (Bio-Rad) according to standard procedure. Relative transcript levels were calculated using the  $2^{-\Delta\Delta C_t}$  formula and the result displayed using histograms drawn with GraphPad Prism 8 software and all the histograms merged using Adobe Photoshop (2020) software. All qPCR analyses were performed with three biological and technical replications.

## **Results And Discussion**

### **Identification and sequence feature of ramie MYB genes**

To identify the MYB family genes in ramie genome, the amino acid sequence of hidden Markov model (HMM) profile of the Pfam MYB domain (PF00249) was used as a query in BLASTP searches. 245 deduced amino acid sequences that contain MYB or MYB-like repeats were examined. The redundant sequences and MYB genes with incomplete ORFs were corrected by FGENESH-M. Subsequently, 211 sequences were further examined to verify the presence of the MYB domains on the basis of the PROSITE, Pfam and NCBI-CDD analyses and a total of 200 non-redundant ramie MYB proteins were identified which including 89 MYB-related proteins (1R-MYB), 105 R2R3-MYB proteins (2R-MYB), 5 R1R2R3-MYB proteins (3R-MYB) and 1 4R-MYB proteins. According to previous reports, R2R3-MYB family members not only represent more than half of the proportion of the total MYB proteins, but also contain most functional genes, and thus constitute the functional group of MYB genes. Therefore, only the R2R3-MYB family members were further analyzed in this study.

Based on their order on the chromosome, the 105 BnGR2R3-MYB genes were renamed *BnGMYB1* to *BnGMYB102* and also *BnGMYB103* – *BnGMYB105* for the three MYB genes (*Maker00022368*, *Maker00031449*, and *Maker00022368*) which could not be conclusively mapped to any chromosome (Table S1). BnGR2R3-MYB genes were further characterized by computing such information as length of the CDS (Coding Sequence), the length of the protein sequence, the protein molecular weight (MW), isoelectric point (pI) and the subcellular localization (Table S1). Among the 105 BnGMYB proteins, *BnGMYB52* was the smallest protein with 135 amino acids (aa), whereas the largest one was *BnGMYB1* (593 aa). The MW of the proteins ranged from 15.6 to 65.0 kDa, and the pI ranged from 4.56 (*BnGMYB23*) to 10.08 (*BnGMYB96*). The predicted subcellular localization results showed that 103 of the 105 BnGMYB proteins were located in the nuclear region, whereas the remaining two were in either chloroplast or cytoplasm.

R2, R3 repeats are significantly conserved sequences within the MYB domain regions. To investigate the homologous domain sequence features and the frequency of amino acids at each position into the ramie R2R3-MYB domains, multiple alignments (MEGA 7.0) and sequence logos were constructed (Fig. 1). In general, R2R3-MYB conservative region contains 108 basic residues (including the linker region) on average. The results revealed that 11 and 6 completely conserved amino acid residues were identical among all BnGMYB proteins detected in the R2 and R3 MYB repeat regions, respectively (Fig. 1). These included the most prominent series of evenly distributed and highly conserved triplet tryptophan (Trp., W) residues in each repeat, and the characteristic W residues was located at positions 9, 29 and 49 of the R2 repeat and 28 and 47 of the R3 repeat in ramie. These characteristic amino acids are known to play an important role in sequence-specific DNA binding, and are considered as landmarks of plant MYB proteins [30].

R2 repeats contain three fully conserved tryptophan residues; for 101 out of the 105 BnGMYB proteins, the R3 repeat sequences contained two Trp residues, whereas, these conserved sequences were not immutable, three R3 repeats (*BnGMYB66*, *BnGMYB56*, *BnGMYB67*) only contained first Trp residues, the last Trp was replaced by Tyrosine (Tyr., Y) or phenylalanine (Phe., F). The results were consistent with those of other species[31, 32]. It has been reported that the insertion of a leucine residue (Leu., L) between the second and third helices of the R2 repeat represents the evolutionary relationship of the R2R3-MYB domain[33]. In this study, 83 (79.05%) BnGR2R3-MYBs were observed to have Leu-38 inserted at the same site (Table S2). As shown in Fig. 1, the highly conserved amino acid residues of the MYB domain were mainly distributed near the third conserved Trp residues indicating that the last HTH domain of the MYBs was the most conserved among the 105 BnGMYB proteins.

## **The classification, gene structure and motif composition of BnGMYBs**

We performed a phylogenetic analysis of R2R3-MYB proteins using neighbor-joining (NJ) and maximum likelihood (ML) methods with 1000 bootstrap replicates. The tree topologies obtained using the two methods were largely similar with a very few variations in the gene classification (Fig. 2a and Figure S1).

Since the ML method can choose the optimal alternative model and optimize the evolutionary tree with certain topological structure and branch length, we adopted it for further characterization. The R2R3-MYB members of ramie were subdivided into 30 subgroups (designated A1-A30 in this study) using Arabidopsis MYB proteins[4] as reference based on the phylogenetic analysis result (Fig. 2a), each subgroup included 2–8 members and most subgroups were supported with high bootstrap value. Most of the large subgroups in our classification were consistent with those of Arabidopsis, and only one small subgroup (*AtMYB24*; S19) not retrieved from the previously constructed phylogenetic trees of Arabidopsis MYB protein. Nine clades only contained ramie R2R3-MYB members, and four did not fit into any subgroup.

To investigate the relationship between gene structural function and evolution, we analyzed the structural diversity of the BnGR2R3-MYB family, which showed genes having the same genetic structure clustering together as obtained in the phylogenetic tree. Most of gene coding sequences were disrupted by introns, except for almost all members from A30 subgroup, ungrouped *BnGMYB27/37* and *BnGMYB72* (Fig. 2c). The number of introns ranges from zero to eleven, but most ramie MYB genes possessed three exons and two introns in line with previous reports[34]. Interestingly, the majority of intron insertions occur in R2, R3 conserved domains indicating the important role of these conserved domains in plants. In conclusion, the highly conserved gene structure and domain influence each other, which provided important evidence for the division of subgroups.

MYB conservative motifs as obtained using MEME online tool with the aim of unraveling diversification of these BnGR2R3-MYB genes (Fig. 2b) revealed the presence of 20 conserved motifs in the ramie MYB proteins (Table S3). Similarly, we also found that the protein architecture was conserved within a specific subgroup. Phylogenetic tree, gene structure, and motif analysis suggested that MYB proteins within the same subgroup were likely to share similar functions. In addition to the conserved motif representing R2R3, other similar motifs were shared by the BnGR2R3-MYB members within the same subgroup with motifs such as 11/17/19 found to be unique in A25, A23 and A1 respectively, while motif 8/9/12/18 were unique in several subgroup, which strongly supporting our subgroup classifications.

## Chromosomal distribution and synteny analysis of BnGMYB genes

As shown in Fig. 3, ramie R2R3-MYB genes were distributed throughout all 14 chromosomes, though it was uneven with Chr10 having the largest number (14) of the R2R3-MYB genes followed by Chr4 (13) and only 2 genes were found on chr12. No significant correlation was obtained between the chromosome length and the number of BnGR2R3-MYB genes, though, it's concentrated on almost every chromosome, and on the top and bottom of chromosome (Fig. 3). This was in line with the theory that the closer the more prone to cross both ends of the chromosomes, ectopic, and other mutations[35].

The segmental and tandem duplication events is the dominant form of the gene family expansion[36]. Based on BLASTP and MCScanX methods, we investigated the gene duplication events. Reference to Holub[37] for a description of tandem duplication, among the BnGR2R3-MYB genes, the blue lines in

Fig. 3 showed four pairs of tandemly duplicated genes (*BnGMYB32/33*, *BnGMYB51/52*, *BnGMYB61/62*, *BnGMYB93/94*) located in chr5, chr7, chr8, chr13, respectively, while the red lines showed 39 segmental (37.1%) duplication pairs between BnGR2R3-MYB genes (Table S4) including two MYB-related genes (*Maker00025472*; *Maker00000772*) and one 3R-MYB gene (*Maker00072694*). Some BnGMYBs located on chr13, chr6 and chr4, such as *BnGMYB91*, *BnGMYB90*, *BnGMYB89*, *BnGMYB21*, *BnGMYB44* and *BnGMYB45*, were involved in multiple duplication events. Interestingly, chr12 (*BnGMYB89*), chr13 (*BnGMYB90/91*) and chr6 (*BnGMYB45*) were interrelated with one another in a segmental duplication manner, all segmentally duplicated genes belonged to A30, and in the same cluster as Arabidopsis *AtMYB45* (Fig. 2a) making us speculate that this specific population of R2R3-MYB genes duplication in ramie might be closely related to photomorphogenesis[38], and might play a vital role in the growth and development of the plant.

The positive pressure criteria were selected based on M. Lynch[39]:  $Ka/Ks < 1$  stands for purifying selection,  $Ka/Ks = 1$  means neutral selection, while  $Ka/Ks > 1$  signifies positive selection. We found that there was a high sequence divergence value among the four segmental gene pairs. These sequences diverged greatly and evolution distance was long, owing to the occurrence of a large number of synonymous mutations. The remaining segmental and tandem duplicated BnGMYB gene pairs were  $Ka/Ks < 1$  except *BnGMYB66/67* ( $Ka/Ks = 1.04$ ). These results suggest that the ramie R2R3-BnGMYB gene family had evolved under the effect of purifying selection (Table S4).

To further explore the potential evolutionary mechanisms of BnGR2R3-MYB genes, we constructed five comparative syntenic maps of ramie with five other plants, three from which are dicots (*A. thaliana*, *C. sativa*, *A. venetum*) and two monocots (rice and maize) (Fig. 4). The numbers of orthologous pairs between the ramie and *A. venetum*, *C. sativa*, *A. thaliana*, *O. sativa* and *Z. mays* were 88, 58, 53, 8 and 4 respectively. Collinearity analysis showed that ramie and *A. venetum* and *C. sativa* had the most collinearity segments, which was consistent with the phylogenetic divergence time estimation which supported the placement of ramie within the same clade as *C. sativa*[29]. Some of the BnGMYBs were found to be associated with multiple syntenic gene pairs, particularly *BnGMYB89* and four *A. thaliana* genes, suggestion that these genes might have played an important role in MYB gene family during evolution. Remarkably, the number of collinear gene pair between ramie and *A. venetum*, *C. sativa*, and *A. thaliana* (dicotyledonous) was far greater than between ramie and *O. sativa*/*Z. mays* (monocotyledonous)[29, 40, 41]. 19 BnGMYBs identified in dicotyledonous plant were absent in monocotyledonous plant, suggestion that these orthologous pairs (19 BnGMYBs) formed after the divergence of dicotyledonous and monocotyledonous plants. Additionally, one collinear pair (*BnGMYB64*) was identified in all the six detected plants, indicating that these orthologous pairs might have already existed before the ancestral divergence (Table S5).

## Comparative phylogenetic analysis of BnGMYB and R2R3-MYB family from six different plant species

A Neighbor-joining (NJ) phylogenetic tree was constructed using the 105 full-length R2R3-MYB protein sequences of ramie and those from *Arabidopsis* (126), rice (81), maize (156), tomato (121) and pineapple (93). Due to the large number of proteins (682), only condensation tree and statistics were provided (Fig. 5). The complete version was shown in Figure S2. The phylogenetic analysis generated 35 clades (C1-C35) among six species and the numbers of the MYB members in each species were listed. The fact that this tree was in good agreement with the classification results of *Arabidopsis*[4] and other plants[31, 32, 42] demonstrated the reliability of the data. Moreover, the comprehensive phylogenetic analysis of the R2R3-MYB proteins in these six species might provide more evolutionary historical clues. For example, monocotyledon proteins in the same group were often found clustered together, while MYB proteins from dicotyledonous plants were clustered side by side in lateral branches. These data were also in the order Poales. Meanwhile, several clades were found only in some particular species. For example, C5, C11, and C16 not existed in rice, maize, and pineapple. C31 was only found in monocotyledons, but not in dicotyledons. This indicated a great evolutionary distance between monocotyledons and dicotyledons, and these MYB genes might not be present prior to monocotyledonous genomic differentiation.

Most of the clades contained members from all six plants, suggesting that these genes were present in the common ancestor before the species diverged, while each clade had a different number of representative MYB proteins suggesting that MYB genes expansion may be more active in some plant species, and this expansion inequality may be related to the well conserved chromosome karyotype of ramie[43].

Interestingly, A subgroup (C24) containing all members of other species but ramie may have disappeared during the evolutionary course of ramie, and their function need further exploration. C12 contains only *Arabidopsis* members, further demonstrating that the *Arabidopsis* corresponding S12 subgroup was a specific clade of *Arabidopsis* that is involved in the regulation of Glucosinolates and Camalexin biosynthesis which is consistent with the results of the *Prunus salicina* MYB study[31]. *Maker00077618* (*BnGMYB33*), *Maker00077519* (*BnGMYB32*) in ramie were closely clustered with *AtMYB103* (from *Arabidopsis*), *SIMYB52* (from tomato) which negatively regulate endoreduplication during trichome initiation and expansion, suggestion that the two genes in ramie might also involve in trichomes and tapetal cells functions. In addition, the annotation in the reference model plant showed the unique functions of each clade and provided a direction for the functional identification of ramie MYB proteins[4, 38, 44].

Polyploidy and subsequent genes loss exist in most species is an important driving force of species evolution. Through comparative genomic analysis, it could be inferred (Figure S3) that ramie has undergone at least one round of genome-wide duplication events. Phylogenetic analysis of MYB gene family showed that the number of BnGMYB members in some clades was consisted with rice and maize. Ramie also undergone the genome-wide duplication event common to rice[45] and maize[46].

## **Analysis of the cis-acting elements of the BnGR2R3-MYB genes and Gene ontology annotation**

The upstream promoter regions (1000bp) of all BnGR2R3-MYB genes were predicted to better understand the functions of the R2R3-MYB genes in the ramie plants. 41 responsive cis-acting elements obtained were categorized into 4 groups (Fig. 6 and Table S6), which made up of 22 light responsive elements, 8 phytohormone responses, 6 plant growth development-related responses and 5 abiotic stresses (adversity). Light response elements being the most abundant are present in all genes. Box-4, G-box (light responsiveness), ARE (anaerobic induction) and ABRE (abscisic acid responsiveness) were the most prominent elements, and more than 100 were found in the 105 BnGR2R3-MYB genes, suggesting that these elements might play an important role in regulating gene expression.

A total of 65 (61.9%) genes were found to contain ABRE. 118 CGTCA-motif and TGACG-motif were involved in the response to MeJA and as well 77 TGA-element and TCA-element involved in the response to auxin. These results indicate the potential function of BnGR2R3-MYB genes in response to hormonal stresses.

Almost all BnGR2R3-MYB genes as obtained in this study contained several cis-acting elements associated with abiotic stresses such as drought-inducibility, anaerobic induction, low-temperature-responsive, defense and stress-responsive, and wound-responsive which were similar to what was reported[47]. The abiotic stress response elements were mostly clustered, which might be related to the function of gene groups. For example: A29 and A30 groups of phylogenetic trees in the red line in Fig. 6 (corresponding to the grouping C32 and C33 constructed by NJ method in Fig. 5) constructed by ML method showed concentrated distribution of abiotic stress response elements, which were consistent with their functional clustering. This further illustrates the accuracy of the results.

Plant growth elements were dispersed (Fig. 6), and CAT-box and O2-site were the main responses to plant growth and development. Summarily, the overall results indicated that BnGR2R3-MYB genes play an irreplaceable role in plant growth and development, stress and phytohormone responses.

Gene ontology (GO) annotations of 105 BnGMYB proteins were obtained using protein blast in Blast2GO tool. Based on the identified GO terms, the involvement of the identified BnGMYB proteins in biological processes (BP), cellular component (CC) and molecular function (MF) were shown in Fig. 7[48]. The GO term “binding” (GO: 0005488) best described the greatest number of genes (96, 92.38%), and “single-organism developmental process” (GO: 0044767), “developmental process” (GO:0032502) were significantly enriched in top 20 of biological process. These GO annotations of BnGMYB proteins were in agreement with the experimental findings in other species[49–51].

## **Expression profiling of BnGR2R3-MYB genes with RNA-seq in different tissues**

To explore the temporal and spatial expression patterns of BnGMYBs, we analyzed their transcript levels using transcriptome data of 9 different tissues (phloem, root, leaf) or developmental stages of ramie (Fig. 8 and Table S7a). The results showed most BnGMYBs have different expression patterns, consistent with the results of other plants[52]. Three BnGMYBs can't detected in all nine samples, suggesting that

they were pseudogenes, or only expressed in other tissues or special temporal (Fig. 8). We performed hierarchical cluster analysis using expression data of 102 other BnGMYBs from 9 different samples. BnGMYBs were categorized into three main groups, high expression, preferential expression, relatively lower expression. We found that the phylogenetic grouping was inconsistent with the expression grouping in most cases, indicating that the expression patterns (period, location) of genes with similar functions were significantly different. However, some closely related MYB genes showed highly similar transcript levels, for example, three members of A30 (*BnGMYB89/90/91*) were all grouped in high expression group. There were 57 genes expressed in all tissues, and 16 of them showed constitutive expression (FPKM > 2). In order to better understand the preferential expression of genes, we filtered genes based on meeting two set rules: FPKM > 2 (at least in one tissue); tissues with the highest expression levels should have twice as much expression in at least one of the other tissues. Based on this, 11 genes in phloem\_third period, 19 genes in leaf of variety ZZ\_1, 22 genes in terrestrial root were found to have exhibited preferential expression over the others (Table S7c).

As a bast fiber crop, ramie fiber underwent successively co-growth with the stem. The expression level of *Maker00025443* (*BnGMYB14*) increased gradually during the five developmental stages of phloem (Fig. 8). The homologous gene *AtMYB46/83* are gatekeeper of secondary wall biosynthesis in *Arabidopsis*, which directly regulates the gene expression of secondary wall-associated cellulose synthases[53, 54]. Therefore, we speculated that *BnGMYB14* was related to the synthesis of secondary cell wall in ramie. The bark thickness and phloem fiber length were important yield traits of ramie and increased with the growth of ramie[29], which is consistent with the expression patterns of some MYB TFs. For example, *Maker09720* (*BnGMYB67*), *Maker00041853* (*BnGMYB66*) both clustered into C30 subgroup and also have higher homologies to *AtMYB69*, which was linked to regulating biosynthesis of lignin, xylan and cellulose, participating in secondary cell wall thickening[55]. The expression levels of these two genes in the fifth cell wall thickening stage were significantly higher than those in the first four stages. According to its expression pattern and phylogenetic relationship, the two genes could be involved in cell wall regulation during ramie fiber development. In previous studies, we explored the mechanism of red leaf formation using transcriptomic analysis, based on which, we identified MYB TFs that regulate leaf color. It is also reported that, the highly conserved MYB–bHLH–WD repeat (MBW) transcriptional complex model plays an important role in regulating flavonoid pigment pathway[56]. In this study, we found some MYB TFs related to flavonoid syntheses which are expressed in HX\_1 leaves more than in ZZ\_1. C1 cluster subgroup have several experimentally verified anthocyanin regulators, such as *AtMYB75*[57], *AtMYB90*[15], *AtMYB113/114*[58] from *Arabidopsis*, *SIMYB75*[14] from tomato and was the largest anthocyanin subgroup. However, no BnGMYB was classified into this subgroup, which must mean that MYB in other groups such as group C4[59] and C6[60] also play a role in regulating anthocyanin synthesis, and further pointed to the fact that the mechanisms of anthocyanin synthesis regulation are differed with species. Furthermore, adequate flavonoids and other components in ramie may be the reason for its mildew-proof and bacteriostatic effects[61, 62]. *Maker00087487* (*BnGMYB60*, C4) was closely clustered with *AtMYB12*, which acted as an activator and enhanced the biosynthesis of flavonoids in *Arabidopsis*[59, 63], and its expression was higher in leaves of ZZ\_1 than that of HX\_1. Thus,

*BnGMYB60* might be involved in the regulation of flavonoids synthesis (Table S7b). The above two substances represented the flavanols synthesis pathway and proanthocyanin synthesis pathway[64, 65], respectively, and there may be a substrate competition relationship (antagonistic relationship) between them. Therefore, it was not difficult to understand that the expression of proanthocyanin biosynthesis related gene (C1) was high in the red variety, while the expression of flavanols biosynthesis related gene (*BnGMYB60*) was high in the green variety. Further study should be conducted to verify these results.

## Expression profiles of BnGMYBs under Cd<sup>2+</sup> stress

Increasing studies revealed that R2R3-MYB family genes were involved in various stress response. However, not much information is available about the MYB genes involvement into the cadmium tolerance response. It is well known that ramie has strong cadmium tolerance and accumulation ability, and is a promising plant for remediation of land polluted by heavy metal cadmium[66]. With aid of RNA-seq expression profile data combined with phylogenetic analysis, we preliminarily screened 11 MYB genes related to stress and proanthocyanidin synthesis in ramie and investigated the expression pattern of these MYB genes.

According to the results, BnGMYB gene showed different levels of expression under different cadmium treatment times, suggesting that they would be candidate genes for cadmium tolerance in ramie (Fig. 9). Among the 11 genes investigated, no two genes showed exactly the same expression trend. Most of the genes were up-regulated, while some genes were down-regulated. *BnGMYB41/104* in the stem and *BnGMYB89* in the root remained unchanged. The expression trend of *BnGMYB78* showed a two-stage pattern: its expression level in stem and leaf first increased before decreased after the cadmium stress treatment, while the trend was reversed in the root. In previous studies, Liu et al[67] selected *BnGMYB78* as a corresponding factor to cadmium stress. *BnGMYB78* and *BnGMYB28* belong to the same group, though their patterns of expression were inconsistent indicating the complexity of the response pathway of cadmium stress in ramie. Previous studies have reported that under cadmium stress, expression of *AtMYB4* was induced in *Arabidopsis* and resulted in cadmium tolerance regulation via enhanced protection against oxidative damage and increases expression of *PCS1* and *MT1C*[60]. Interestingly, *BnGMYB41* and *AtMYB4* are orthologous genes, and the expression of this gene in roots and leaves increased with the extension of cadmium treatment time reaching a significant level. This implied that *BnGMYB41* might also activate the protective mechanism of oxidative damage in ramie, contributing to ramie cd-tolerance.

### Coi

ncidentally, the expression patterns of the 11 genes in the three tissues were consistent in the two ramie varieties, which indicated similarity in the response mechanism of different ramie varieties to cadmium stress. However, the response of many other genes such as *BnGMYB28* in root; *BnGMYB9* in stem; *BnGMYB78* in leaf varies greatly between the two, with the degree of responses more intense in of HX\_1 than in ZZ\_1. This might be the reason for the differences in cadmium tolerance between the two.

Some genes like *BnGMYB9/11/41/89* showed a bottom-up (root-stem-leaf) time sequential response pattern, which might be related to the direction of cadmium transport and signal transduction. All genes were up-regulated in leaves after cadmium treatment. *BnGMYB10/11/12* were all up-regulated in all the three tissues, and their expression trends were very similar. *BnGMYB10* has been identified as a cadmium stress response factor and belongs to the same C1 subgroup with the other two (*BnGMYB11/12*) and by extension, these three were found very close to the S5 subfamily of *Arabidopsis thaliana*, whose function was related to proanthocyanidin synthesis[4]. Therefore, we inferred those plants initiate proanthocyanidin synthesis under cadmium stress, and proanthocyanidin functions in alleviating the damage caused due to the stress via producing an antioxidant active substance[68] or produced a large number of reactive oxygens to alleviate.

According to report, *AtMYB49* regulates cadmium accumulation[69], *AtMYB59* regulates calcium signaling during plant growth and stress response[70]. In ramie, we found their respective homologous genes, *BnGMYB9* and *BnGMYB 104* and their expression patterns were similar to those of *AtMYB49/59* under cadmium treatment[69, 70], so we deduced that they also had similar functions. *BnGMYB45/89/90* which fall to the C33 group were according to phylogenetic analysis valuable candidate for abiotic stress[71–73]. Taken together, these above findings might be valuable for improving the cadmium resistance of ramie via the manipulation of the BnGMYBs.

## **BnGR2R3-MYB protein-protein interaction network prediction**

To perform a functional analysis of BnGR2R3-MYB proteins, interaction network was generated using protein families based on their orthologues AtR2R3-MYB in *Arabidopsis*. Several interactions were predicted and we could preliminarily infer the possible function and potential contact of some BnGR2R3-MYB genes (Fig. 10). Previous studies have shown *AtMYB4* (homolog of *BnGMYB41*), *AtMYB12* (homolog of *BnGMYB10*) and *AtMYB123* (homolog of *BnGMYB12*) involvement in the regulation of multiple flavonoid synthesis pathway genes like *F3H*, *FLS*, *C4H*, *ANR*. These regulatory genes, in addition to their role flavonoid biosynthesis also regulate some stress resistance-related proteins such as *LPP2* (Lipid phosphate phosphatase 2). And *AtMYB123*[74] which involved in the control of flavonoid late metabolism in developing siliques also plays a key role in determining tissue-specific activation of leucoanthocyanidin reductase, implying that the mechanism by which *LPP2* responded to stress via attenuating the signal generated by phosphodiesterase was also present in ramie[75]. We found that there is a strong bidirectional interaction between *SAD2* and *AtMYB4*. *SAD2* functions in regulating abscisic acid (ABA)-mediated pathways in response to cold or salt stress in the form of an autonomous nuclear transport receptor or a junction-like protein[76, 77]. Furthermore, multiple genes in the flavonoid metabolic pathway interact with *JOX4* (Jasmonate-induced oxygenase 4) involved in the oxidation of jasmonate (JA) and stress-induced synthesis of phytohormones[78, 79], indicating an intrinsic link between flavonoid metabolic pathways and stress.

It is worth mentioning that *AtMYB30* (homolog of *BnGMYB28*), *AtMYB49* (homolog of *BnGMYB9*) and *AtMYB96* (homolog of *BnGMYB78*), which are associated with biotic and abiotic stresses, interact with an extremely rich set of resistance proteins including *SIZ1* (participating in abiotic stress-induced sumoylation), *RD22* (moisture stress), *YAK1* (ABA-mediated, drought response), *ABI5* (response to abscisic acid, chitin, gibberellin, salt stress, water deprivation), and finally connects the two possible coercive coping mechanisms at the *TGG1* node (Fig. 10). Also, this verifies our later conjecture. In conclusion, the predicted network identifies the interface between flavonoid metabolic pathways and adversity stress, and found evidence for the involvement of flavonoid synthetic pathways in the stress regulation.

## Conclusions

A total of 105 BnGR2R3-MYBs, phylogenetically divided into 35 distinct subfamilies, were identified in ramie and unevenly distributed among 14 chromosomes. The segmental duplication events especially prominent in distal telomeric regions, played a crucial role in the expansion of BnGR2R3-MYB gene family. The numbers of orthologous pairs between ramie and other plants had the highest value (58) with *Apocynum venetum*. We also established the existence of functional divergence during ramie evolution. Expression analysis led to the identification of tissue preferential and fiber development and cadmium stress responsive expression patterns of the BnGR2R3-MYB genes. Additionally, putative functions of ramie MYB genes were assigned based on the phylogenomic results and gene expression data. The integration of our results allowed us to propose some powerful MYB candidates that might be involved in the regulation of phenylpropanoid, flavonoids, lignin, anthocyanidin and proanthocyanidin biosynthesis, providing a new perspective into the role of MYB transcription factors in secondary metabolism. In addition, 11 BnGMYB genes, including MYB related to procyanidin synthesis, were identified, suggesting a new regulatory pathway for ramie's resistance to cadmium stress, which will provide valuable information for the study of the mechanism of cadmium stress in ramie and the breeding of cadmium-tolerant varieties. Further functional characterization of BnGR2R3-MYB genes is needed for a better understanding of the role and regulatory mechanisms of the BnGR2R3-MYB gene family.

## Abbreviations

HTH: helix–turn–helix; MEME: Multiple Em for Motif Elicitation; NJ: Neighbor–Joining; ORF: open reading frame; GSDS: Gene structure display server; PlantTFDB: Plant Transcription Factor Database; RNA-seq: RNA-sequencing; SMART: Simple Modular Architecture Research Tool; TAIR: The Arabidopsis Information Resource; TFs: Transcription Factors; qPCR: Quantification real-time PCR; MWs: Molecular weights; MEME: Multiple Expectation maximization for Motif Elicitation; pIs: Isoelectric points; GRAVY: Grand average of hydropathicity; HX\_1: Hongxuan No.1 variety; ZZ\_1: Zhongzhu No.1 variety

## Declarations

## Acknowledgements

We would like to thank the Institute of bast fiber crops, Chinese Academy of Agricultural Sciences for kindly supplying the experimental material.

### **Author Contributions**

PC, XKF conceived and designed the experiments; XKF, AGZ, JKC, XFW, PM performed the experiments; XKF, CMY, KMC, GG, analyzed the data; XKF, ASA, PC wrote and reviewed the paper. All authors have read and approved the manuscript.

### **Funding**

This work was supported by the Agricultural Science and Technology Innovation Project of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-2020), the China Agriculture Research System of MOF and MARA (CARS-16-E03).

### **Availability of data and materials**

Raw reads, assembly genome sequences and corresponding annotation used in this study source cngb (China National GeneBank) database. Project: CNP0001467, review link: <https://db.cngb.org/cnsa/project/CNP0001467/reviewlink>. Bast fiber tissue transcriptomics in 5 developmental stages source cngb under project: CNP0001453, review link: <https://db.cngb.org/cnsa/project/CNP0001453/reviewlink>. We have received administrative permission to access and use these. The transcriptome of root and leaf used in this study are available through the NCBI SRA (PRJNA717370, GSE98903).

### **Ethics approval and consent to participate**

The plant materials are owned by the Institute (the Institute of bast fiber crops, Chinese Academy of Agricultural Sciences) and we have the right to use them. Sampling of plant materials were performed in compliance with institutional, national, and international guidelines. The materials were publicly available for non-commercial purposes. No specific permits were required.

### **Consent for publication**

Not applicable.

### **Competing interests**

The authors declare no conflict of interest.

### **Author details**

Institute of Bast Fiber Crops, Chinese Academy of Agricultural Sciences, Changsha 410205, China.

## **References**

1. Liu L, White MJ, MacRae TH: **Transcription factors and their genes in higher plants functional domains, evolution and regulation.** *European journal of biochemistry* 1999, **262**(2):247–257.
2. Riechmann JL, Ratcliffe OJ: **A genomic perspective on plant transcription factors.** *Current opinion in plant biology* 2000, **3**(5):423–434.
3. 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.** *Nature structural biology* 1996, **3**(2):178–187.
4. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L: **MYB transcription factors in Arabidopsis.** *Trends in plant science* 2010, **15**(10):573–581.
5. Jiang CK, Rao GY: **Insights into the Diversification and Evolution of R2R3-MYB Transcription Factors in Plants.** *Plant physiology* 2020, **183**(2):637–655.
6. Paz-Ares J, Ghosal D, Wienand U, Peterson PA, Saedler H: **The regulatory c1 locus of Zea mays encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators.** *The EMBO journal* 1987, **6**(12):3553–3558.
7. Ng DW, Abeysinghe JK, Kamali M: **Regulating the Regulators: The Control of Transcription Factors in Plant Defense Signaling.** *International journal of molecular sciences* 2018, **19**(12).
8. Baldoni E, Genga A, Cominelli E: **Plant MYB Transcription Factors: Their Role in Drought Response Mechanisms.** *International journal of molecular sciences* 2015, **16**(7):15811–15851.
9. Li J, Han G, Sun C, Sui N: **Research advances of MYB transcription factors in plant stress resistance and breeding.** *Plant signaling & behavior* 2019, **14**(8):1613131.
10. Feller A, Machemer K, Braun EL, Grotewold E: **Evolutionary and comparative analysis of MYB and bHLH plant transcription factors.** *The Plant journal: for cell and molecular biology* 2011, **66**(1):94–116.
11. Zhang S, Chen Y, He X, Du J, Zhang R, Ma Y, Hu X, Zhang Z, Chen Q, Wan X: **Identification of MYB Transcription Factors Regulating Theanine Biosynthesis in Tea Plant Using Omics-Based Gene Coexpression Analysis.** *Journal of agricultural and food chemistry* 2020, **68**(3):918–926.
12. Arce-Rodríguez ML, Ochoa-Alejo N: **An R2R3-MYB Transcription Factor Regulates Capsaicinoid Biosynthesis.** *Plant physiology* 2017, **174**(3):1359–1370.
13. Akagi T, Ikegami A, Tsujimoto T, Kobayashi S, Sato A, Kono A, Yonemori K: **DkMyb4 is a Myb transcription factor involved in proanthocyanidin biosynthesis in persimmon fruit.** *Plant physiology* 2009, **151**(4):2028–2045.
14. Jian W, Cao H, Yuan S, Liu Y, Lu J, Lu W, Li N, Wang J, Zou J, Tang N *et al.*: **SIMYB75, an MYB-type transcription factor, promotes anthocyanin accumulation and enhances volatile aroma production in tomato fruits.** *Horticulture research* 2019, **6**:22.
15. Bac-Molenaar JA, Fradin EF, Rienstra JA, Vreugdenhil D, Keurentjes JJ: **GWA Mapping of Anthocyanin Accumulation Reveals Balancing Selection of MYB90 in Arabidopsis thaliana.** *PloS one* 2015, **10**(11):e0143212.

16. Muñoz-Gómez S, Suárez-Baron H, Alzate JF, González F, Pabón-Mora N: **Evolution of the Subgroup 6 R2R3-MYB Genes and Their Contribution to Floral Color in the Perianth-Bearing Piperales.** *Frontiers in plant science* 2021, **12**:633227.
17. Huang D, Tang Z, Fu J, Yuan Y, Deng X, Xu Q: **CsMYB3 and CsRuby1 form an 'Activator-and-Repressor' Loop for the Regulation of Anthocyanin Biosynthesis in Citrus.** *Plant & cell physiology* 2020, **61**(2):318–330.
18. Ballester AR, Molthoff J, de Vos R, Hekkert B, Orzaez D, Fernández-Moreno JP, Tripodi P, Grandillo S, Martin C, Heldens J *et al.*: **Biochemical and molecular analysis of pink tomatoes: deregulated expression of the gene encoding transcription factor SIMYB12 leads to pink tomato fruit color.** *Plant physiology* 2010, **152**(1):71–84.
19. Ito M, Araki S, Matsunaga S, Itoh T, Nishihama R, Machida Y, Doonan JH, Watanabe A: **G2/M-phase-specific transcription during the plant cell cycle is mediated by c-Myb-like transcription factors.** *The Plant cell* 2001, **13**(8):1891–1905.
20. Araki S, Ito M, Soyano T, Nishihama R, Machida Y: **Mitotic cyclins stimulate the activity of c-Myb-like factors for transactivation of G2/M phase-specific genes in tobacco.** *The Journal of biological chemistry* 2004, **279**(31):32979–32988.
21. Sun X, Gong SY, Nie XY, Li Y, Li W, Huang GQ, Li XB: **A R2R3-MYB transcription factor that is specifically expressed in cotton (*Gossypium hirsutum*) fibers affects secondary cell wall biosynthesis and deposition in transgenic *Arabidopsis*.** *Physiologia plantarum* 2015, **154**(3):420–432.
22. Kobayashi K, Suzuki T, Iwata E, Magyar Z, Bögre L, Ito M: **MYB3Rs, plant homologs of Myb oncoproteins, control cell cycle-regulated transcription and form DREAM-like complexes.** *Transcription* 2015, **6**(5):106–111.
23. Zhu X, Liang W, Cui X, Chen M, Yin C, Luo Z, Zhu J, Lucas WJ, Wang Z, Zhang D: **Brassinosteroids promote development of rice pollen grains and seeds by triggering expression of Carbon Starved Anther, a MYB domain protein.** *The Plant journal: for cell and molecular biology* 2015, **82**(4):570–581.
24. Chen T, Li W, Hu X, Guo J, Liu A, Zhang B: **A Cotton MYB Transcription Factor, GbMYB5, is Positively Involved in Plant Adaptive Response to Drought Stress.** *Plant & cell physiology* 2015, **56**(5):917–929.
25. Ma Q, Dai X, Xu Y, Guo J, Liu Y, Chen N, Xiao J, Zhang D, Xu Z, Zhang X *et al.*: **Enhanced tolerance to chilling stress in OsMYB3R-2 transgenic rice is mediated by alteration in cell cycle and ectopic expression of stress genes.** *Plant physiology* 2009, **150**(1):244–256.
26. Wang Y, Li F, He Q, Bao Z, Zeng Z, An D, Zhang T, Yan L, Wang H, Zhu S *et al.*: **Genomic analyses provide comprehensive insights into the domestication of bast fiber crop ramie (*Boehmeria nivea*).** *The Plant journal: for cell and molecular biology* 2021.
27. Mu L, Cai M, Wang Z, Liu J, Liu T, Wanapat M, Huang B: **Assessment of ramie leaf (*Boehmeria nivea* L. gaud) as an animal feed supplement in P.R. China.** *Tropical animal health and production* 2020, **52**(1):115–121.

28. Kiprioti, M., Alexopoulou, E., Vafeiadakis, T., Kipriotis, Heping, Crops XJI, Products: **Ramie and kenaf as feed crops**. 2015.
29. Luan MB, Jian JB, Chen P, Chen JH, Chen JH, Gao Q, Gao G, Zhou JH, Chen KM, Guang XM *et al*: **Draft genome sequence of ramie, *Boehmeria nivea* (L.) Gaudich.** *Molecular ecology resources* 2018, **18**(3):639–645.
30. Jiang C, Gu X, Peterson T: **Identification of conserved gene structures and carboxy-terminal motifs in the Myb gene family of *Arabidopsis* and *Oryza sativa* L. ssp. indica.** *Genome biology* 2004, **5**(7):R46.
31. Liu C, Hao J, Qiu M, Pan J, He Y: **Genome-wide identification and expression analysis of the MYB transcription factor in Japanese plum (*Prunus salicina*).** *Genomics* 2020, **112**(6):4875–4886.
32. Li Z, Peng R, Tian Y, Han H, Xu J, Yao Q: **Genome-Wide Identification and Analysis of the MYB Transcription Factor Superfamily in *Solanum lycopersicum*.** *Plant & cell physiology* 2016, **57**(8):1657–1677.
33. Williams CE, Grotewold E: **Differences between plant and animal Myb domains are fundamental for DNA binding activity, and chimeric Myb domains have novel DNA binding specificities.** *The Journal of biological chemistry* 1997, **272**(1):563–571.
34. Du H, Yang SS, Liang Z, Feng BR, Liu L, Huang YB, Tang YX: **Genome-wide analysis of the MYB transcription factor superfamily in soybean.** *BMC plant biology* 2012, **12**:106.
35. Schilling S, Kennedy A, Pan S, Jermiin LS, Melzer R: **Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization.** *The New phytologist* 2020, **225**(1):511–529.
36. Cannon SB, Mitra A, Baumgarten A, Young ND, May G: **The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*.** *BMC plant biology* 2004, **4**:10.
37. Holub EB: **The arms race is ancient history in *Arabidopsis*, the wildflower.** *Nature reviews Genetics* 2001, **2**(7):516–527.
38. Kranz HD, Denekamp M, Greco R, Jin H, Leyva A, Meissner RC, Petroni K, Urzainqui A, Bevan M, Martin C *et al*: **Towards functional characterisation of the members of the R2R3-MYB gene family from *Arabidopsis thaliana*.** *The Plant journal: for cell and molecular biology* 1998, **16**(2):263–276.
39. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes.** *Science (New York, NY)* 2000, **290**(5494):1151–1155.
40. Liu C, Zeng L, Zhu S, Wu L, Wang Y, Tang S, Wang H, Zheng X, Zhao J, Chen X *et al*: **Draft genome analysis provides insights into the fiber yield, crude protein biosynthesis, and vegetative growth of domesticated ramie (*Boehmeria nivea* L. Gaud).** *DNA research: an international journal for rapid publication of reports on genes and genomes* 2018, **25**(2):173–181.
41. Huang KY, Zhu AG, Chen XR, Shi YL, Tang Q, Wang XF, Sun ZM, Luan MB, Chen JH: **Comparative transcriptomics reveals the selection patterns of domesticated ramie.** *Ecology and evolution* 2019, **9**(12):7057–7068.

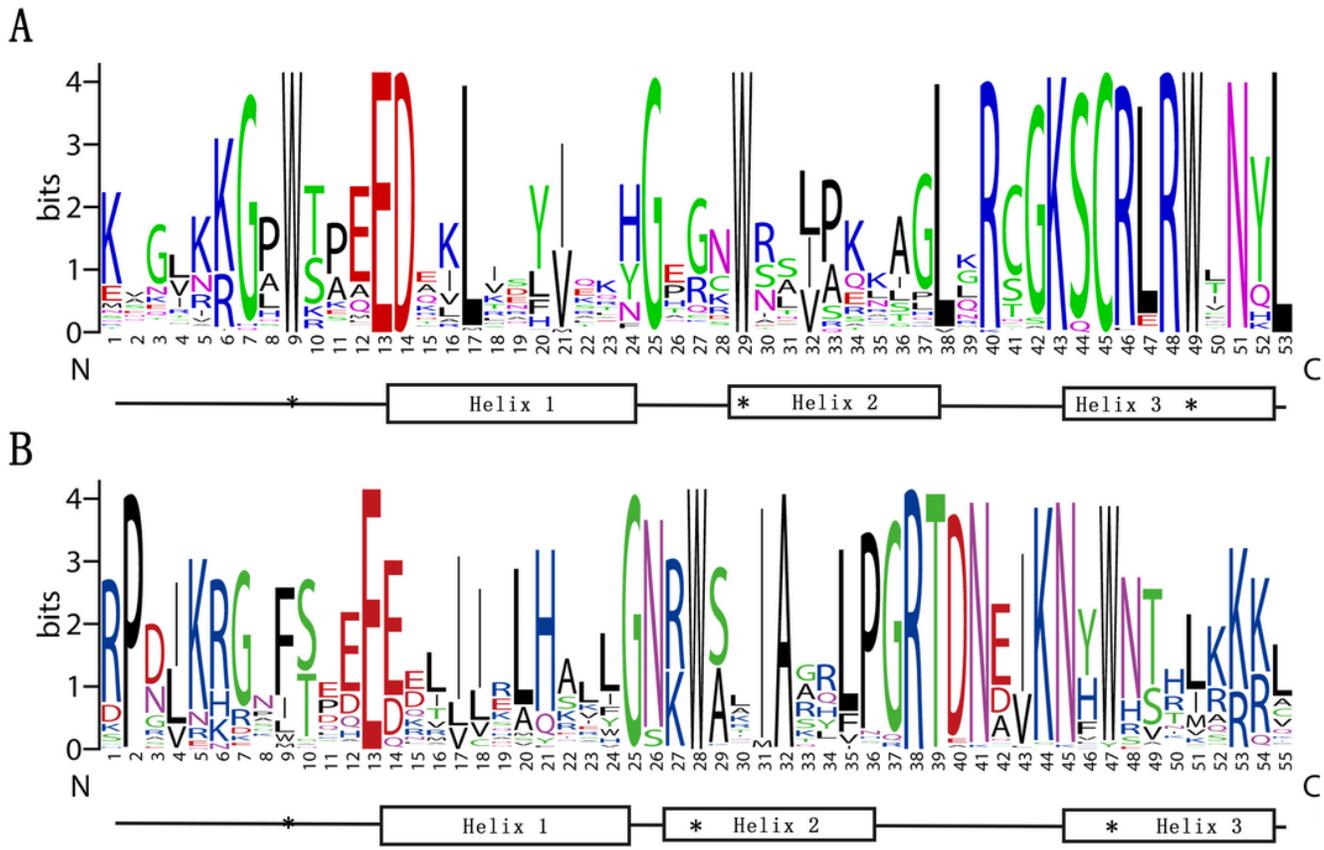
42. Liu C, Xie T, Chen C, Luan A, Long J, Li C, Ding Y, He Y: **Genome-wide organization and expression profiling of the R2R3-MYB transcription factor family in pineapple (*Ananas comosus*)**. *BMC genomics* 2017, **18**(1):503.
43. Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang ML, Chen J, Biggers E *et al*: **The pineapple genome and the evolution of CAM photosynthesis**. *Nature genetics* 2015, **47**(12):1435–1442.
44. Stracke R, Werber M, Weisshaar B: **The R2R3-MYB gene family in *Arabidopsis thaliana***. *Current opinion in plant biology* 2001, **4**(5):447–456.
45. Thiel T, Graner A, Waugh R, Grosse I, Close TJ, Stein N: **Evidence and evolutionary analysis of ancient whole-genome duplication in barley predating the divergence from rice**. *BMC evolutionary biology* 2009, **9**:209.
46. Si W, Hang T, Guo M, Chen Z, Liang Q, Gu L, Ding T: **Whole-Genome and Transposed Duplication Contributes to the Expansion and Diversification of TLC Genes in Maize**. *International journal of molecular sciences* 2019, **20**(21).
47. Yamaguchi-Shinozaki K, Shinozaki K: **Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses**. *Annual review of plant biology* 2006, **57**:781–803.
48. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the unification of biology**. The Gene Ontology Consortium. *Nature genetics* 2000, **25**(1):25–29.
49. Chen S, Niu X, Guan Y, Li H: **Genome-Wide Analysis and Expression Profiles of the MYB Genes in *Brachypodium distachyon***. *Plant & cell physiology* 2017, **58**(10):1777–1788.
50. Smita S, Katiyar A, Chinnusamy V, Pandey DM, Bansal KC: **Transcriptional Regulatory Network Analysis of MYB Transcription Factor Family Genes in Rice**. *Frontiers in plant science* 2015, **6**:1157.
51. 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.
52. Chen X, Wang P, Gu M, Lin X, Hou B, Zheng Y, Sun Y, Jin S, Ye N: **R2R3-MYB transcription factor family in tea plant (*Camellia sinensis*): Genome-wide characterization, phylogeny, chromosome location, structure and expression patterns**. *Genomics* 2021, **113**(3):1565–1578.
53. Kim WC, Ko JH, Kim JY, Kim J, Bae HJ, Han KH: **MYB46 directly regulates the gene expression of secondary wall-associated cellulose synthases in *Arabidopsis***. *The Plant journal: for cell and molecular biology* 2013, **73**(1):26–36.
54. Ko JH, Jeon HW, Kim WC, Kim JY, Han KH: **The MYB46/MYB83-mediated transcriptional regulatory programme is a gatekeeper of secondary wall biosynthesis**. *Annals of botany* 2014, **114**(6):1099–1107.
55. Zhong R, Lee C, Zhou J, McCarthy RL, Ye ZH: **A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis***. *The Plant cell* 2008, **20**(10):2763–2782.

56. Lloyd A, Brockman A, Aguirre L, Campbell A, Bean A, Cantero A, Gonzalez A: **Advances in the MYB-bHLH-WD Repeat (MBW) Pigment Regulatory Model: Addition of a WRKY Factor and Co-option of an Anthocyanin MYB for Betalain Regulation.** *Plant & cell physiology* 2017, **58**(9):1431–1441.
57. Zheng T, Tan W, Yang H, Zhang L, Li T, Liu B, Zhang D, Lin H: **Regulation of anthocyanin accumulation via MYB75/HAT1/TPL-mediated transcriptional repression.** *PLoS genetics* 2019, **15**(3):e1007993.
58. Coego A, Brizuela E, Castillejo P, Ruíz S, Koncz C, del Pozo JC, Piñeiro M, Jarillo JA, Paz-Ares J, León J: **The TRANSPLANTA collection of Arabidopsis lines: a resource for functional analysis of transcription factors based on their conditional overexpression.** *The Plant journal: for cell and molecular biology* 2014, **77**(6):944–953.
59. Mehrtens F, Kranz H, Bednarek P, Weisshaar B: **The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis.** *Plant physiology* 2005, **138**(2):1083–1096.
60. Agarwal P, Mitra M, Banerjee S, Roy S: **MYB4 transcription factor, a member of R2R3-subfamily of MYB domain protein, regulates cadmium tolerance via enhanced protection against oxidative damage and increases expression of PCS1 and MT1C in Arabidopsis.** *Plant science: an international journal of experimental plant biology* 2020, **297**:110501.
61. Kandimalla R, Kalita S, Choudhury B, Devi D, Kalita D, Kalita K, Dash S, Kotoky J: **Fiber from ramie plant (*Boehmeria nivea*): A novel suture biomaterial.** *Materials science & engineering C, Materials for biological applications* 2016, **62**:816–822.
62. Chen Y, Wang G, Wang H, Cheng C, Zang G, Guo X, Liu RH: **Phytochemical profiles and antioxidant activities in six species of ramie leaves.** *PloS one* 2014, **9**(9):e108140.
63. Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B: **Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the Arabidopsis thaliana seedling.** *The Plant journal: for cell and molecular biology* 2007, **50**(4):660–677.
64. Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V: **Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway.** *Journal of experimental botany* 2011, **62**(8):2465–2483.
65. Vogt T: **Phenylpropanoid biosynthesis.** *Molecular plant* 2010, **3**(1):2–20.
66. She W, Zhu S, Jie Y, Xing H, Cui G: **Expression profiling of cadmium response genes in ramie (*Boehmeria nivea* L.) root.** *Bulletin of environmental contamination and toxicology* 2015, **94**(4):453–459.
67. Zheng X, Zhu S, Tang S, Liu T: **Identification of drought, cadmium and root-lesion nematode infection stress-responsive transcription factors in ramie %J Open Life Sciences.** 2016, **11**(1):191–199.
68. Wang N, Qu C, Jiang S, Chen Z, Xu H, Fang H, Su M, Zhang J, Wang Y, Liu W *et al*: **The proanthocyanidin-specific transcription factor MdMYBPA1 initiates anthocyanin synthesis under low-**

- temperature conditions in red-fleshed apples. *The Plant journal: for cell and molecular biology* 2018, **96**(1):39–55.
69. Zhang P, Wang R, Ju Q, Li W, Tran LP, Xu J: **The R2R3-MYB Transcription Factor MYB49 Regulates Cadmium Accumulation.** *Plant physiology* 2019, **180**(1):529–542.
70. Fasani E, DalCorso G, Costa A, Zenoni S, Furini A: **The Arabidopsis thaliana transcription factor MYB59 regulates calcium signalling during plant growth and stress response.** *Plant molecular biology* 2019, **99**(6):517–534.
71. Maitra Majee S, Sharma E, Singh B, Khurana JP: **Drought-induced protein (Di19-3) plays a role in auxin signaling by interacting with IAA14 in Arabidopsis.** *Plant direct* 2020, **4**(6):e00234.
72. Nguyen NH, Cheong JJ: **H2A.Z-containing nucleosomes are evicted to activate AtMYB44 transcription in response to salt stress.** *Biochemical and biophysical research communications* 2018, **499**(4):1039–1043.
73. Jaradat MR, Feurtado JA, Huang D, Lu Y, Cutler AJ: **Multiple roles of the transcription factor AtMYBR1/AtMYB44 in ABA signaling, stress responses, and leaf senescence.** *BMC plant biology* 2013, **13**:192.
74. Zimmermann IM, Heim MA, Weisshaar B, Uhrig JF: **Comprehensive identification of Arabidopsis thaliana MYB transcription factors interacting with R/B-like BHLH proteins.** *The Plant journal: for cell and molecular biology* 2004, **40**(1):22–34.
75. Pierrugues O, Brutesco C, Oshiro J, Gouy M, Deveaux Y, Carman GM, Thuriaux P, Kazmaier M: **Lipid phosphate phosphatases in Arabidopsis. Regulation of the AtLPP1 gene in response to stress.** *The Journal of biological chemistry* 2001, **276**(23):20300–20308.
76. Zhao J, Zhang W, Zhao Y, Gong X, Guo L, Zhu G, Wang X, Gong Z, Schumaker KS, Guo Y: **SAD2, an importin -like protein, is required for UV-B response in Arabidopsis by mediating MYB4 nuclear trafficking.** *The Plant cell* 2007, **19**(11):3805–3818.
77. Verslues PE, Guo Y, Dong CH, Ma W, Zhu JK: **Mutation of SAD2, an importin beta-domain protein in Arabidopsis, alters abscisic acid sensitivity.** *The Plant journal: for cell and molecular biology* 2006, **47**(5):776–787.
78. Caarls L, Elberse J, Awwanah M, Ludwig NR, de Vries M, Zeilmaker T, Van Wees SCM, Schuurink RC, Van den Ackerveken G: **Arabidopsis JASMONATE-INDUCED OXYGENASES down-regulate plant immunity by hydroxylation and inactivation of the hormone jasmonic acid.** *Proceedings of the National Academy of Sciences of the United States of America* 2017, **114**(24):6388–6393.
79. Smirnova E, Marquis V, Poirier L, Aubert Y, Zumsteg J, Ménard R, Miesch L, Heitz T: **Jasmonic Acid Oxidase 2 Hydroxylates Jasmonic Acid and Represses Basal Defense and Resistance Responses against Botrytis cinerea Infection.** *Molecular plant* 2017, **10**(9):1159–1173.
80. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF: **Protein identification and analysis tools in the ExpASY server.** *Methods in molecular biology (Clifton, NJ)* 1999, **112**:531–552.

81. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R: **TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data.** *Molecular plant* 2020, **13**(8):1194–1202.
82. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS: **MEME SUITE: tools for motif discovery and searching.** *Nucleic acids research* 2009, **37**(Web Server issue):W202-208.
83. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome research* 2009, **19**(9):1639–1645.
84. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H *et al.*: **MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity.** *Nucleic acids research* 2012, **40**(7):e49.
85. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J: **KaKs\_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies.** *Genomics, proteomics & bioinformatics* 2010, **8**(1):77–80.
86. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.** *Molecular biology and evolution* 2016, **33**(7):1870–1874.
87. Du H, Feng BR, Yang SS, Huang YB, Tang YX: **The R2R3-MYB transcription factor gene family in maize.** *PloS one* 2012, **7**(6):e37463.
88. Gao G, Xiong H, Chen J, Chen K, Chen P, Yu C, Zhu A: **Hydroponic method for ramie and removal of nitrogen and phosphorus from livestock wastewater.** *International journal of phytoremediation* 2018, **20**(6):545–551.
89. Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Zhang G, Gu YQ, Coleman-Derr D, Xia Q *et al.*: **OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species.** *Nucleic acids research* 2019, **47**(W1):W52-w58.
90. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P *et al.*: **The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible.** *Nucleic acids research* 2017, **45**(D1):D362-d368.

## Figures



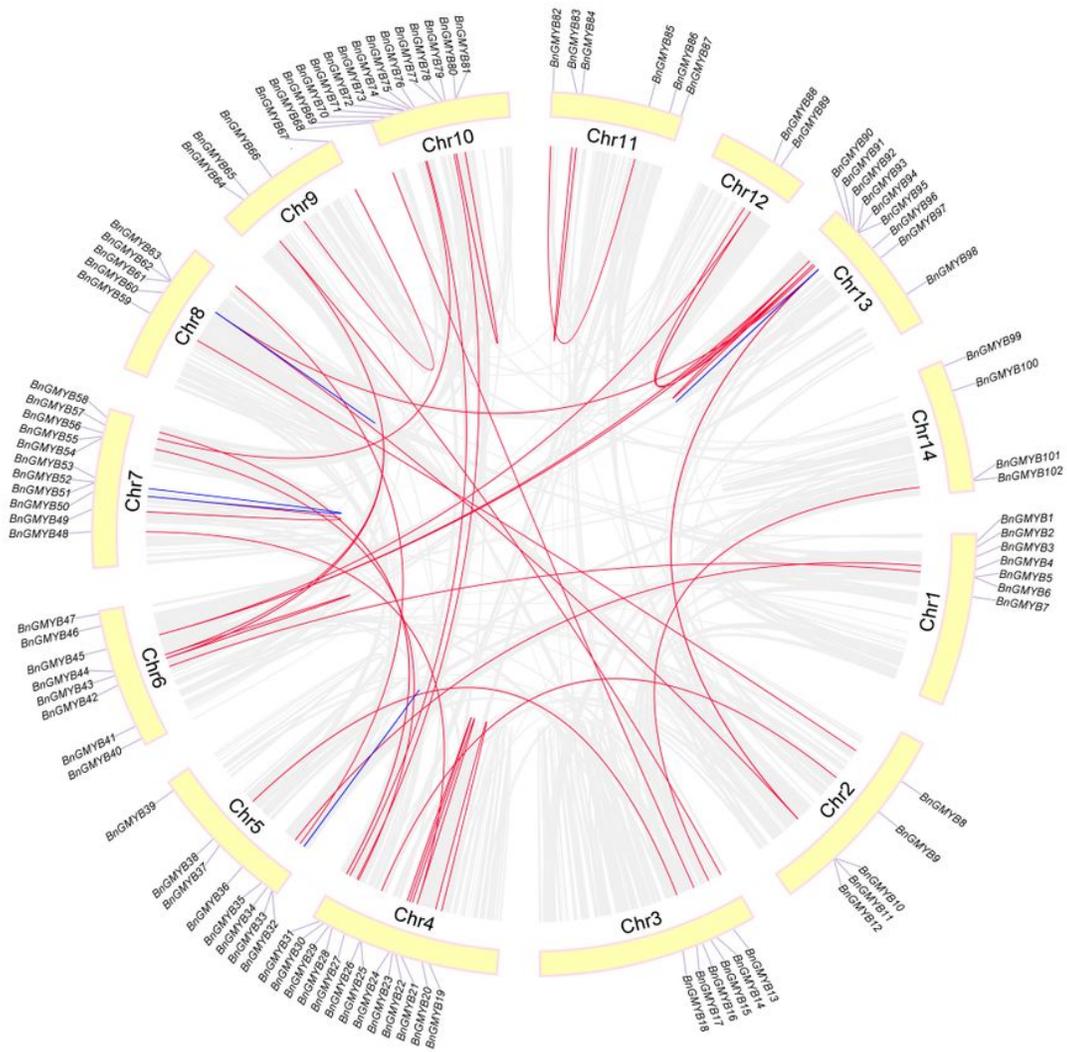
**Figure 1**

Consensus sequence and the level of conservation of R2R3-MYB domains from ramie. The sequence logos of the R2 and R3 MYB repeats are based on full-length alignments of all ramie R2R3-MYB domains. The bit score exhibits the information content for each position in the sequence. The position of the three  $\alpha$ -helices were marked (Helix 1 to 3). The conserved tryptophan residues (Trp, W) in the MYB domain are marked with black asterisks.



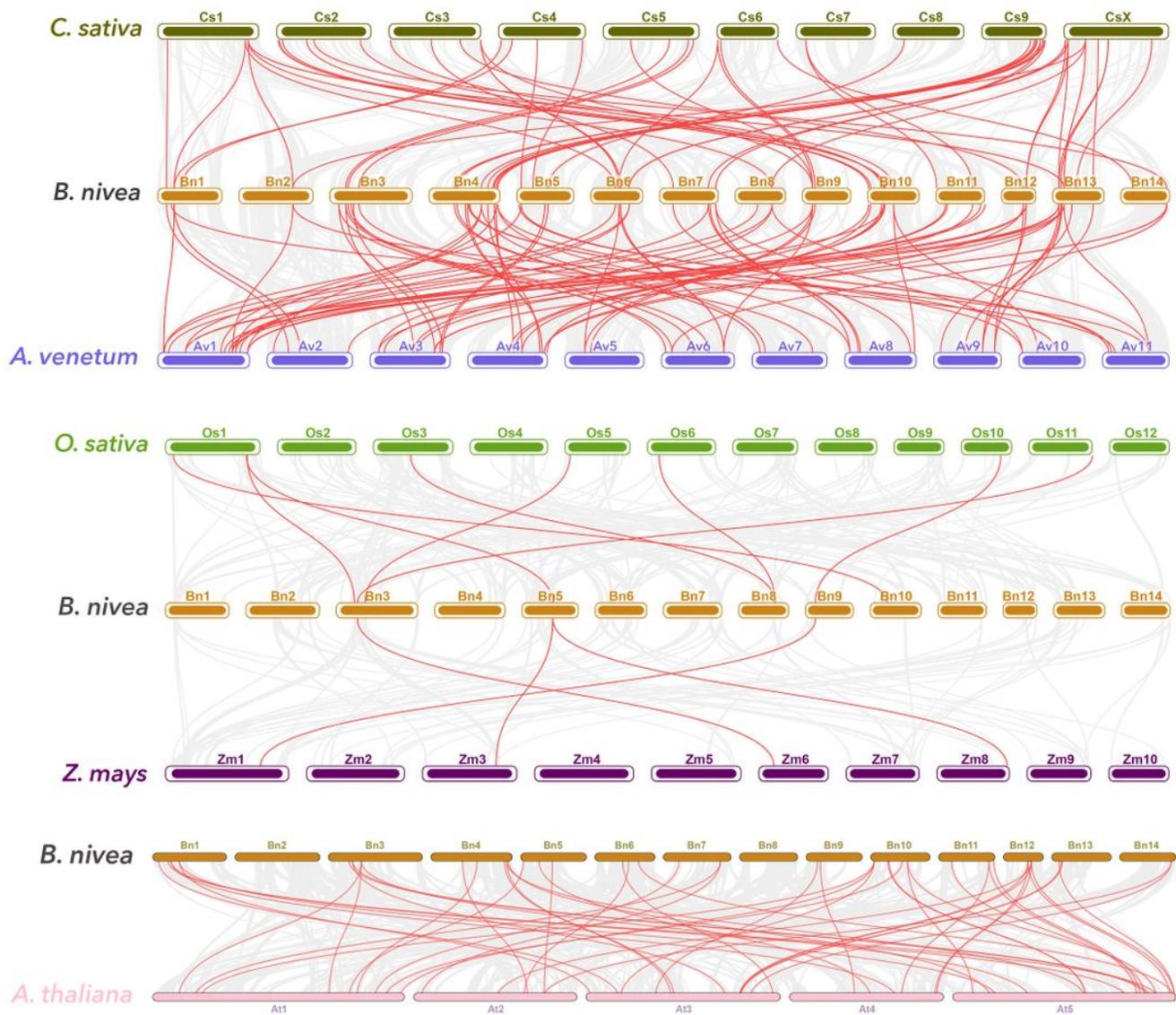
**Figure 2**

Phylogenetic relationships, gene structure and architecture of conserved protein motifs in BnGR2R3-MYB genes. A. The maximum likelihood phylogenetic includes 105 R2R3-MYB proteins from ramie and 23 representatives from Arabidopsis. B. The motif architecture of ramie MYB proteins. 20 different motifs are displayed in different colored boxes. C. Ex-on-intron structure of ramie MYB genes. Green boxes indicate exons; blue boxes indicate untranslated 5' and 3' regions; black lines indicate introns. R2, R3 MYB domain are highlighted by boxes of different shape and color.



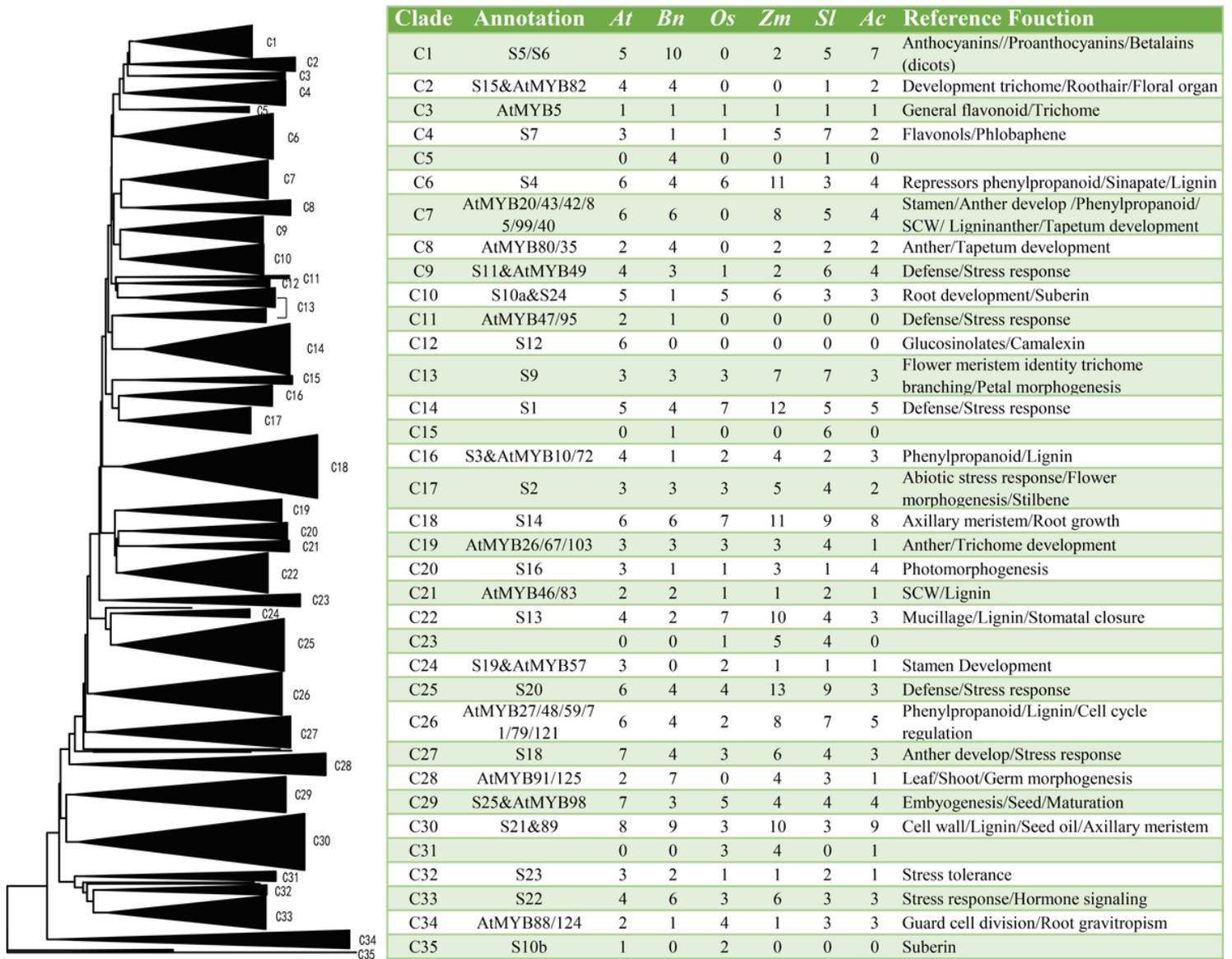
**Figure 3**

Schematic representations for the chromosomal distribution and inter-chromosomal relationships of ramie R2R3-MYB genes. Gray lines indicate all syntenic blocks in the ramie genome, and the red lines indicate duplicated MYB gene pairs. Gene and chromosome names are displayed on the tip and inside of the chromosome, respectively.



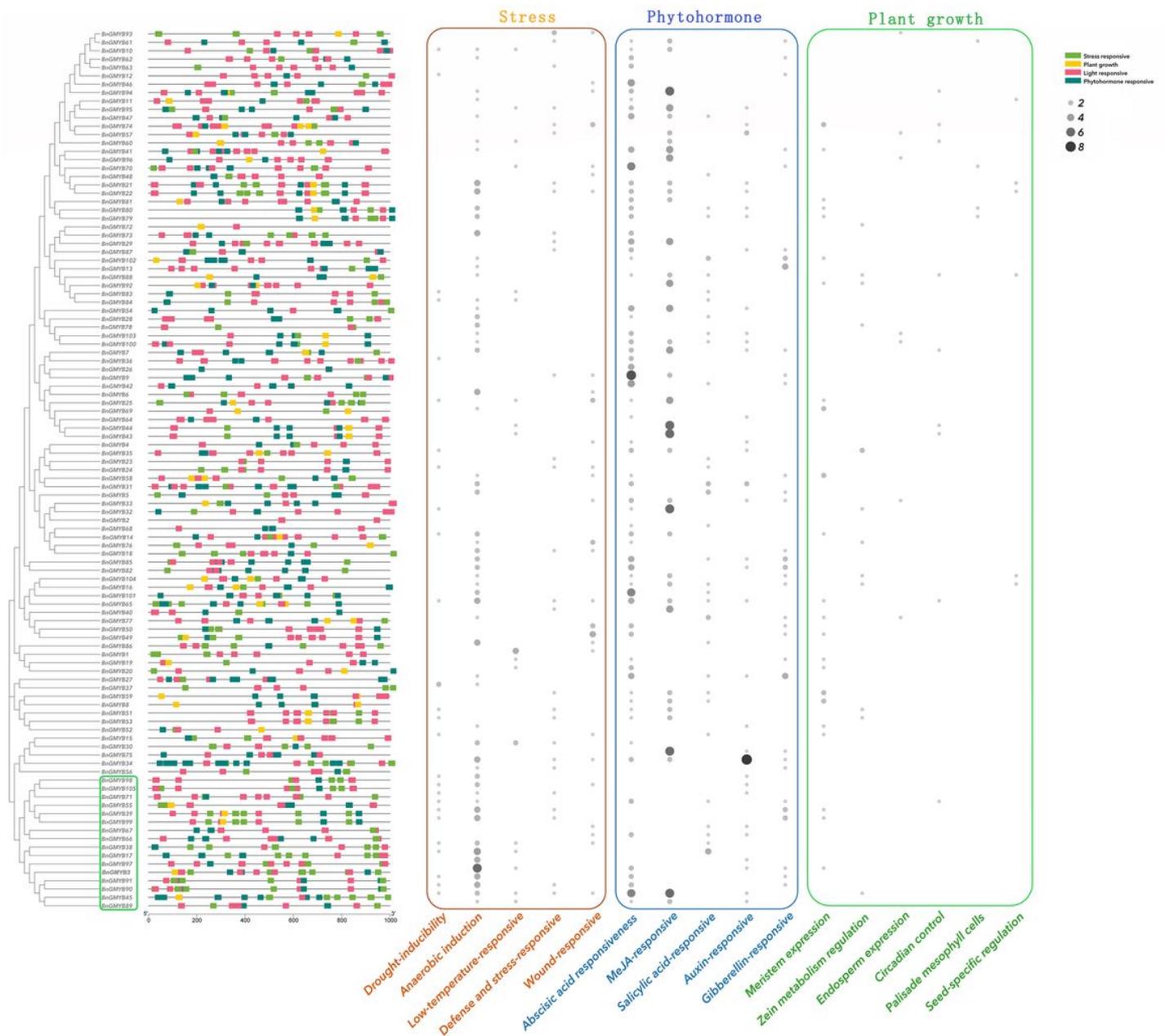
**Figure 4**

Synteny analyses of MYB genes between ramie and five representative plants: *B. nivea*, *C. sativa*, *A. venetum*, *O. sativa*, *Z. mays* and *A. thaliana*. Gray lines in the background indicate the collinear blocks within ramie and other plant genomes, while the red lines highlight the syntenic MYB gene pairs.



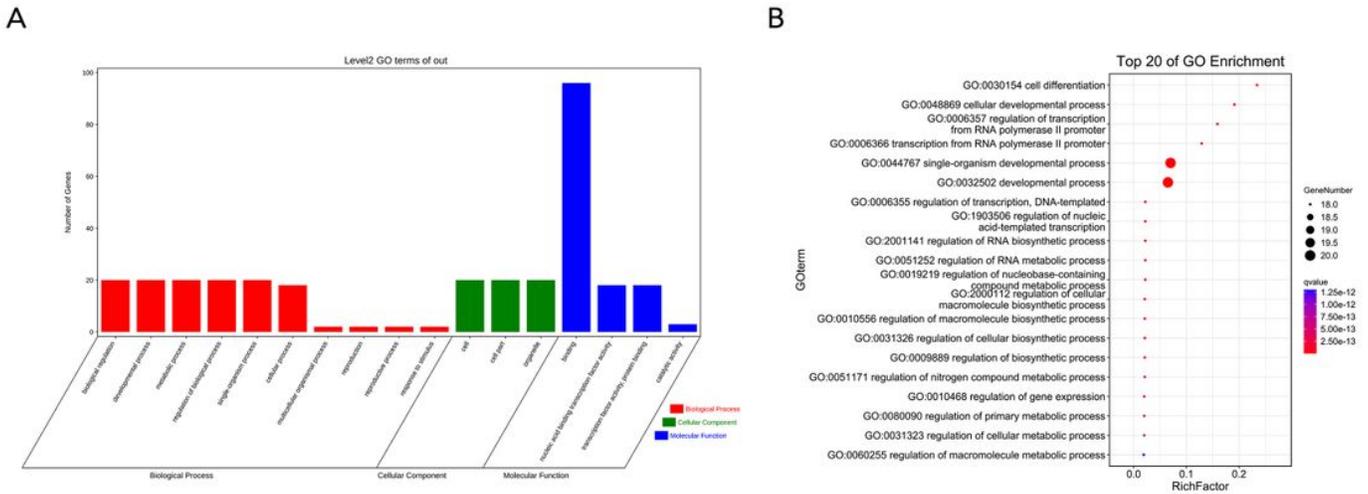
**Figure 5**

Neighbor-joining (NJ) tree showing relationships among R2R3-MYB proteins from *B. nivea* (Bn), *A. thaliana* (At), *O. sativa* (Os), *Z. mays* (Zm), *S. lycopersicum* (Sl) and *A. cosmosus* (Ac). MYB proteins from the six species were designated as C1 to C35. The table on the right indicated the number of the subgroup members and function of clades in each species.



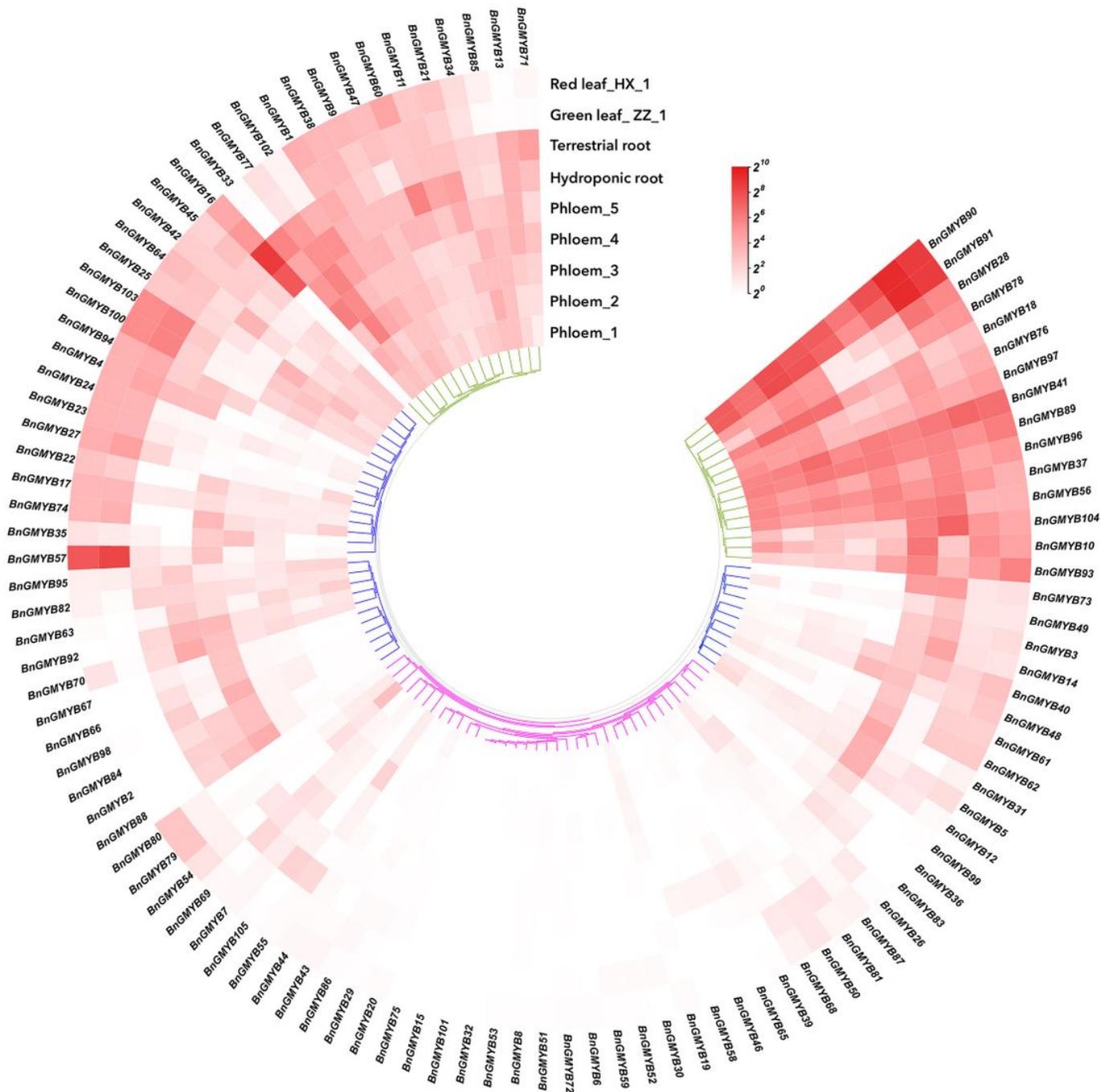
**Figure 6**

Analysis of cis-elements from promoter region of the BnGR2R3-MYB genes. The figure on the left shows the distribution of four cis-acting elements upstream promoter regions (1000bp). On the right is the heat map of the quantity of plant growth, stress responsive and phytohormone responsive cis-elements, represented in brown, blue and green, respectively, circle sizes indicate the number of cis-elements.



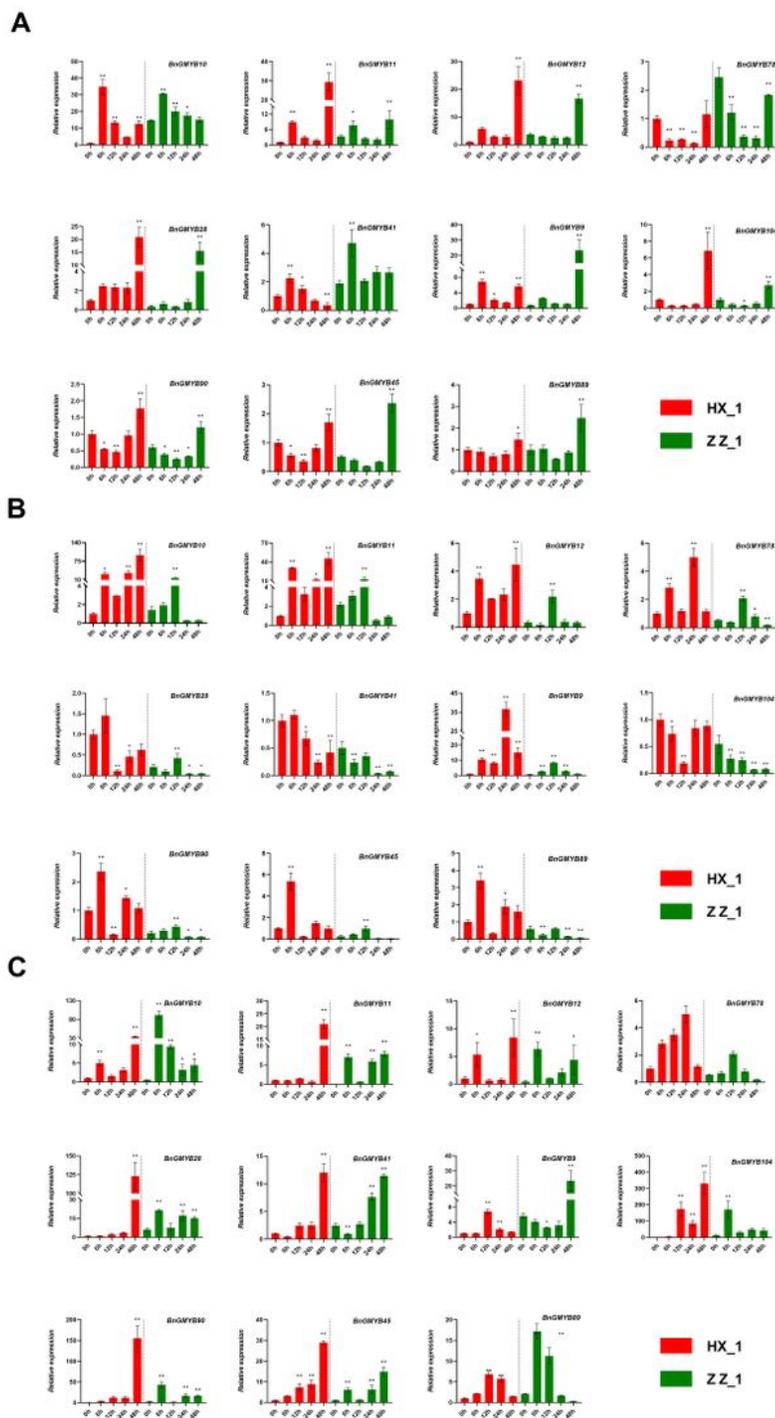
**Figure 7**

BnGMYB Gene Ontology (GO) functional classification. (A) BnGMYB TFs were divided into categories with three independent ontologies. (B) The Top 20 of GO enrichment in biological processes.



**Figure 8**

Hierarchical clustering of expression profiles of ramie R2R3-MYB genes in 9 samples including different tissues and developmental stages. Log<sub>2</sub>(FPKM + 1) values were displayed according to the color code. Detailed FPKM values were listed in Table S7.



**Figure 9**

The relative expression levels of selected MYB genes in two ramie varieties in different tissue sites (root, stem, leaf) at different periods under Cd+2 treatment (0h, 6h, 12h, 24h, 48h). A represents root; B represents stem; C represents leaf. The error bar represents the standard deviation of the three biological duplicates. \* represents significant difference ( $P < 0.05$ ), \*\* represents extremely significant difference ( $P < 0.01$ ).

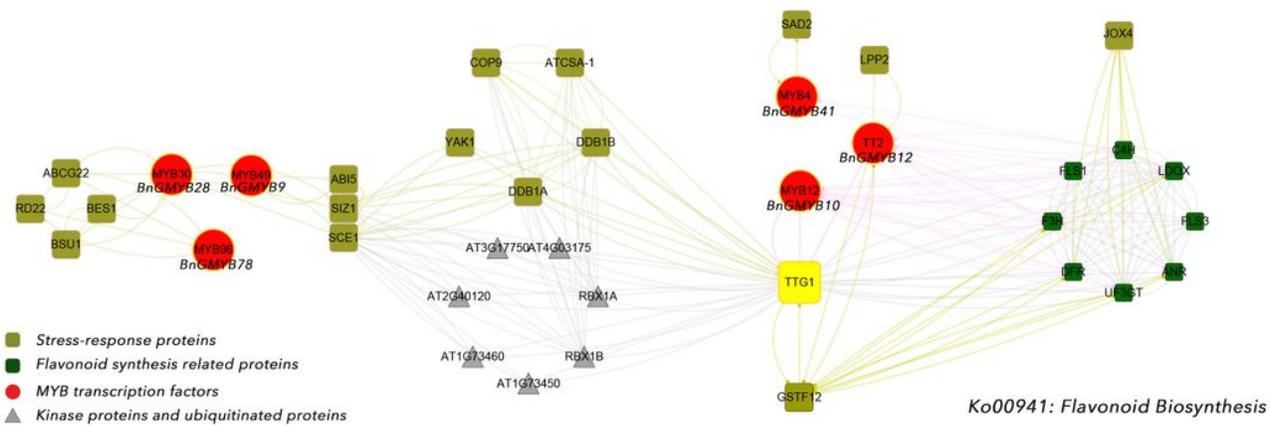


Figure 10

The protein-protein interaction network for BnGR2R3-MYB proteins based on their Arabidopsis orthologs. The red solid circles indicated cadmium stress-related MYB transcription factors, the olive squares represented stress-response proteins, and the green squares were flavonoid synthesis related proteins.

## Supplementary Files

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

- [Additionalfile1.xlsx](#)
- [Additionalfile2.txt](#)
- [Additionalfile3.xlsx](#)
- [Additionalfile4.xlsx](#)
- [Additionalfile5.xlsx](#)
- [Additionalfile6.xlsx](#)
- [Additionalfile7.xlsx](#)
- [Additionalfile8.xlsx](#)
- [Additionalfile9.docx](#)
- [Additionalfile10.pdf](#)
- [Additionalfile11.pdf](#)