Planta (2021) 254:64
https://doi.org/10.1007/s00425-021-03713-1

ORIGINAL ARTICLE

Genome-wide analysis and characterization of R2R3-MYB family in pigeon pea (*Cajanus cajan*) and their functional identification in phenylpropanoids biosynthesis

Jie Yang[1,2] · Su Zhang[3,4] · Hongquan Li[1,2] · Litao Wang[4] · Ying Liu[1,2] · Lili Niu[4] · Qing Yang[3,4] · Dong Meng[3,4] · Yujie Fu[2,3,4]

Received: 19 June 2021 / Accepted: 28 August 2021 / Published online: 6 September 2021

Abstract

Main conclusion **Thirty CcMYB were identified to involve in flavonoid and lignin biosynthesis in pigeon pea genome. A comprehensive analysis of gene structure, phylogenetic relationships, distribution on chromosomes, gene duplication, and expression patterns was performed.**

Abstract MYB transcription factor is one of the largest gene families in plants and plays critical roles in plant growth and development, as well as resistance to biotic and abiotic stress. However, the function of *MYB* genes in pigeon pea (*Cajanus cajan*) remains largely unknown. Here, 30 *R2R3-MYB* which involved flavonoid and lignin biosynthesis were identified in the pigeon pea genome and were classified into five groups based on phylogenetic analysis. Simultaneously, another 122 key enzyme genes from biosynthetic pathways of flavonoid and lignin were identified and all of them were mapped on 11 chromosomes with the co-linearity relationship. Among these genes, the intron/exon organization and motif compositions were conserved and they have undergone a strong purifying selection and tandem duplications during evolution. Expression profile analysis demonstrated most of these genes were expressed in different tissues and responded significantly to MeJA, RNA-seq analysis revealed clear details of genes varied with time of induction. Ten key genes from the phenylpropanoid pathway were selected to further verify whether they responded to induction under different abiotic stress conditions (UV-B, cold, heat, salt, drought, and GA_3). This study elaborates on potential regulatory relationships between *R2R3-MYB* genes and some key genes involved in flavonoid and lignin biosynthesis under MeJA treatment, as well as adding to the understanding of improving abiotic stress tolerance and regulating the secondary metabolism in woody crops. A simplified discussion model for the different regulation networks involved with flavonoid and lignin biosynthesis in pigeon pea is proposed.

Keywords R2R3-MYB · Flavonoid · Lignin · Genome analysis · Abiotic stress · Pigeon pea

Introduction

Cajanus cajan (L.) Millsp. (Family: Fabaceae), also known as pigeon pea, is a multipurpose, hardy grain legume crop grown in semiarid and subtropical areas of the world. The

Communicated by Anastasios Melis.

Jie Yang and Su Zhang have contributed equally to this work.

✉ Yujie Fu
yujie_fu@163.com

[1] College of Chemistry, Chemical Engineering and Resource Utilization, Northeast Forestry University, Harbin 150040, China

[2] Key Laboratory of Forest Plant Ecology, Ministry of Education, Northeast Forestry University, Harbin 150040, China

[3] Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing Forestry University, Beijing 100083, China

[4] College of Forestry, Beijing Forestry University, Beijing 100083, China

crop can be described as unique because it is a legume and a woody shrub. Among the leguminous crops, pigeon pea ranks fifth in the area after soybean, common bean, peanut, and chickpea. In addition to being used as a food crop, pigeon pea has been widely utilized as forage, fuel plant, and medicinal material with many significant activities. As a folk medicine, pigeon pea leaves have been used to treat various ailments worldwide such as wound healing, arrest blood, pain relief. In recent years, pigeon pea has been widely brought to market as a special traditional Chinese medicine for the therapy of osteonecrosis of the femoral head. (Fu et al. 2008, 2006). In this regard, a large number of phytochemical studies reveal that these important activities are attributed to its abundance of secondary metabolites.

Pigeon pea contains several classes of interesting bioactive secondary metabolites including flavonoids, stilbenes, isocoumarins, and stilbene carboxylates (Liu et al. 2010; Rinthong and Maneechai 2018; Nix et al. 2015). Phytochemical investigations reveal that flavonoids and their derivates are the main bioactive compounds presented in pigeon pea. Based on the position and the modifications to the benzene rings, the main classes of these flavonoid derivatives include chalcones, flavanones, flavan 3-ols, flavonols, flavones, isoflavones, and anthocyanins (Lepiniec et al. 2006). Nutrient and health benefits have been reported for nearly all classes of flavonoids from pigeon pea, especially pinostrobin, isovitexin, genistin, apigenin, and luteolin (Cui et al. 2015; Duan et al. 2013; Zhang et al. 2012). The biosynthetic pathways leading to these bioactive compounds originated from the general phenylpropanoid pathway (GPP). This GPP mainly involves phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), and 4-coumarate: CoA ligase (4CL), and redirects carbon flow from primary metabolism to phenylpropanoid metabolism (Fraser and Chapple 2011). The end-product of the GPP pathway is catalyzed by a series of key enzymes (Koes et al. 1994), including chalcone synthase, chalcone isomerase, flavanone 3-hydroxylase, isoflavone synthase, flavone Synthase II, flavonoid B-ring hydroxylases, flavonol synthase, leucoanthocyanidin reductase, dihydroflavonol 4-reductase, leucoanthocyanidin dioxygenase, UDP flavonoid glucosyltransferase, which ultimately leads to the synthesis of flavonoids. It is known to all that lignin also originates from the GPP. The key structural genes of the lignin pathway have been identified in many species (Voelker et al. 2011), the cooperative regulations of flavonoid and lignin biosynthesis have been recognized to be broadly correlated (Kang et al. 2019). Among these regulations, TFs play an important role in plant growth, development, and stress response (Liu et al. 2015; Li et al. 2020a). A large number of TFs have been known to directly regulate the key genes involved in flavonoid and lignin biosynthesis in plants.

The MYB protein as one of the largest transcription regulators in the plant is highly conserved with the so called MYB domain at the N-terminal region, while the C-terminal is highly variable containing specific regulatory domains for transcriptional activation or repression (Jin and Martin 1999). (Matus et al. 2008; Karamysheva et al. 2004). The MYB protein usually contains 1–4 repeats (R1, R2, R3 and R4) and each repeat sequence encodes three α-helices, of which the second and third α-helices are further folded into a helix-turn-helix structure (HTH) (Ogata et al. 1992). From this, MYB proteins are classified into four major types such as 1R-MYB, 2R-MYB, 3R-MYB, and 4R-MYB proteins, of which the R2R3-MYBs are the most common type in plants.

Since the first MYB transcription factor was isolated and identified from plants (Paz-Ares et al. 1987), a large number of studies have been conducted on MYB transcription factors in plants (Dubos et al. 2010). Many of them were confirmed to be involved in the biosynthesis of flavonoids and lignin (Ma and Constabel 2019). Some R2R3-MYB members were identified as positive regulators involved in secondary metabolites biosynthesis in the phenylpropanoid pathway. Such as, *PtrMYB3*, *PtrMYB20* (McCarthy et al. 2010), *EgMYB2* (Goicoechea et al. 2005), *AtMYB83* (McCarthy et al. 2009), and *AtMYB85* (Zhou et al. 2009) promoted the accumulation of lignin in plants by binding to the cis-acting element in the promoter of the structural gene from the lignin synthesis pathway. *AtMYB11*, *AtMYB12*, and *AtMYB111* are the star genes regulating the biosynthesis of flavonol by activating transcription level of chalcone synthase, chalcone isomerase, flavanone 3-hydroxylase, and flavonol synthase (Mehrtens et al. 2005; Stracke et al. 2007; Luo et al. 2008; Misra et al. 2010; Pandey et al. 2012, 2014). The expression of both *AtMYB14* and *AtMYB15* can induce the accumulation of stilbenes in the phenylpropanoid pathway (Höll et al. 2013). In addition, certain R2R3-MYB TFs have been confirmed as repressors to negatively regulate lignin synthesis in plants, such as, *A. thaliana* (*AtMYB32*) (Preston et al. 2004), *Zea mays* (*ZmMYB31*, *ZmMYB42*) (Fornalé et al. 2010) (Sonbol et al. 2009), *Eucalyptus gunnii* (*EgMYB1*) (Legay et al. 2007), *Panicum virgatum* (*PvMYB4a*) (Shen et al. 2012), *Leucaena leucocephala* (*LlMYB1*) (Omer et al. 2013), and *Chrysanthemum morifolium* (*CmMYB1*) (Zhu et al. 2013) are similarly able to repress lignin synthesis. All the R2R3-MYB proteins encoded by these genes belong to R2R3-MYB subgroup 4 (Liu et al. 2015). At present, the function of *MYB* genes in the biosynthesis of flavonoids and lignin has been identified and functionally characterized only in Arabidopsis, herbaceous species, and fewer woody species, while the research of *MYB* genes was little known in medicinal or commercial crops. Up to our knowledge, the function of the MYB family in pigeon pea has not been well investigated until now.

In the present study, 30 *CcMYB* genes and 122 key enzyme genes which are involved in flavonoid and lignin biosynthesis were identified from the pigeon pea genome. Further, gene structures (intron/exon distribution), cis-acting elements, chromosomal locations, motif compositions and phylogenetic analysis, duplication events were also investigated. The synteny analysis of *CcMYB* genes and phenylpropanoid pathway-related genes from *Cajanus cajan* and *Glycine max*, *Oryza sativa*, *Arabidopsis thaliana* were compared together. In addition, as MeJA is an important regulator in the biosynthesis of flavonoid and lignin (Li et al. 2020b; Chen et al. 2020b; Cao et al. 2010), the possible regulation of *CcMYB* genes which are involved in flavonoid and lignin biosynthesis were analyzed according to the RNA-seq of pigeon pea under MeJA treatment. The expression levels of the above-mentioned genes in roots, stems, leaves, flowers, pods, beans, and six different abiotic stress conditions were measured. Moreover, 16 flavonoids in 6 different tissues of pigeon pea were quantitatively analyzed. This is the first report on genome-wide and transcriptome identification of the *CcMYB* genes family and their contribution to flavonoid and lignin biosynthesis in pigeon pea. This study also serves as a valuable reference for further analysis of the regulatory mechanisms involving flavonoid and lignin biosynthesis in plants.

Materials and methods

Identification and sequence analysis

Genome sequences, gene sequences, and general feature format (GFF) files of pigeon pea were downloaded from the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/genome/?term=Cajanus+cajan). The Hidden Markov model (HMM) profile of the MYB domain (PF00249) was downloaded from the Pfam protein family database (http://pfam.xfam.org/) (Liu et al. 2017). All CcMYB protein sequences were searched with a default E-value by HMMER software (version 3.0). To verify its accuracy, the 126 *AtMYB* in *Arabidopsis* were selected as query sequences to further identify all members of the CcMYB family in the pigeon pea genome. The key genes from the phenylpropanoid pathway in pigeon pea were obtained from homology searches using the NCBI database (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) and Uniport database (http://www.uniprot.org/). The MEME (http://meme-suite.org/) (Bailey et al. 2009) and Batch CD-search (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) were used to validate the CcMYB and phenylpropanoid pathway-related proteins domains. The ExPASy (https://web.expasy.org/compute_pi/) was used to calculate the molecular weight and isoelectric point values of the CcMYB and phenylpropanoid pathway-related protein sequences (Gasteiger et al. 2005).

Phylogenetic tree and multiple alignment

A neighbor-joining (NJ) phylogenetic tree of CcMYB and other MYBs that have been identified as functional in flavonoids and lignin biosynthetic pathway was constructed using MEGA 7 software and with 1000 bootstrap replicates for reliability (Kumar et al. 2016). The phylogenetic tree was displayed and annotated by the iTOL online tool (https://itol.embl.de/) (Letunic and Bork 2007). The CcMYB proteins were aligned by BioEdit software to visualize and analyze the sequences of conserved domains in CcMYB proteins. For the CcMYB interaction network, MYB interactions experiment data in *Arabidopsis* were constructed through the STRING website (http://stringdb.org/), and the homolog proteins were identified by BLASTp analysis in pigeon pea.

Gene structure and promoter analysis

The CcMYB and phenylpropanoid pathway-related gene structures were visualized by TBtools software (Chen et al. 2020a). The 2 kb genomic DNA sequences upstream of the initiation codon of the candidate gene were retrieved through the Plant CARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) to identify the cis-acting elements and functional sites in the promoter regions (Lescot et al. 2002).

Chromosomal locations, gene duplication and synteny analysis

The chromosomal location information of the genes was obtained from the pigeon pea genome. The location images of all genes were drawn by MapChart software and gene replication events were detected by multiple collinear scanning toolkits. (Wang et al. 2012). The synteny relationship of all genes between pigeon pea, *Arabidopsis*, soybean, and rice were drawn up by the TBtools software (Chen et al. 2020a).

Expression analysis based on high-throughput mRNA sequencing

For RNAseq library construction, we collected three biological replicates of pigeon pea seedlings from 4-week-old plants. Each sample was collected at 0, 3, 6, and 12 h after MeJA treatment at a concentration of 10 mg/L. Total RNA was prepared using the TRIZOL (Takara) following

the manufacturer's instructions. RNA samples were treated with RNase-free DNase Set (Takara). Total RNA was submitted to the BGI Genomics (Shenzhen, China) for library construction and sequencing. The libraries were sequenced by the 150 bp paired-end reads. The raw data were tested by FastQC (Brown et al. 2017). Clean reads were obtained by removal of low-quality reads (Q value < 20) by Cutadapt (Martin 2011). The clean reads mapped to the pigeon pea genome. Paired-end clean reads were aligned to the pigeon pea genome using TopHat v2.0.9 (Brueffer et al. 2016). To construct transcriptome, the mapped reads were assembled de novo using Cufflinks (Trapnell et al. 2012). The transcript abundance of all genes was denoted as FPKM (fragments per kilobase per million) and the log2 (FPKM) were used for hierarchical clustering, and the results were visualized by TBtools.

Plant materials and abiotic stress treatments

The seeds of pigeon pea (ICPL87119) were cultivated in the growth chamber of Northeast Forestry University. The indoor temperature was 25 °C, and the photoperiod consisted of 16 h of light and 8 h of darkness. Tissue material of pigeon pea from roots, stems, leaves, flowers, pods, and beans were used for tissue specificity experiments. Each tissue was collected from three different plants, and the collected samples were immediately stored at − 80 °C. Plants with similar growth performance (4-week-old plants) were selected for treatments. The pigeon pea seedlings were transferred with consistent growth status to aerated hydroponics with Hoagland solution for pre-cultivation for 1 week before stress treatment. Six stress treatments (including UV-B, cold, heat, drought, salt, and hormones) were applied to the pigeon pea. The pigeon pea seedlings were transferred to the 311 nm UV-B radiation plant incubator to simulate UV-B treatment. The temperature of the plant incubator was set at 40 °C and 4 °C, respectively, to simulate high and low-temperature conditions. In addition, 150 mM NaCl, 10% PEG-6000, and 50 mg/L GA_3 were added to the standard Hoagland solution for salt, drought, and hormone stresses. Plant samples of pigeon pea were collected at 0, 3, 6, and 12 h after treatment, and the expression levels of flavonoid and lignin synthesis-related genes were analyzed after the six treatments. Three biological replicates were set for each sample and control. All collected samples were frozen and stored at − 80 °C.

Total RNA isolation and qPCR analysis

Total RNA was extracted with the TRIZOL reagent (Rio et al. 2010). cDNA was prepared using the SuperScript™ III Reverse Transcriptase kit (Invitrogen) and used as the template for RT-PCR. Quantitative PCR of related genes was performed on a Light Cycler 9600 system (Roche, Switzerland) with SYBR Premix Ex Taq Kit (TAKARA). Gene-specific primers were designed using Primer 5.0 and listed in Supplementary Table S7. The relative gene expression levels were quantitatively analyzed by the $2^{-\Delta\Delta CQ}$ method. The *CcActin* (GenBank Accession No. LOC109798310) gene from pigeon pea was used as the endogenous reference gene (Meng et al. 2019). Three biological and technical replicates for each sample were completed based on qRT-PCR.

Plant harvesting and metabolite extraction

Three-month-old pigeon pea plants were subjected to metabolite analyses, plant tissues from 12 individual plants were collected. Samples were then freeze dried and kept at − 80 °C until subjected to solvent extraction. 0.1 g dry weight (DW) of roots, stems, leaves, flowers, pods, and beans powder of pigeon pea was accurately weighed. Subsequently, 80% ethanol aqueous solution (5 mL) was added as the extraction solvent of pigeon pea flavonoids, and the ultrasonic wave was continuously sonicated at room temperature (25 °C) with 100 W power for 30 min. After centrifugation at 10,000 for 10 min, the supernatant was collected, filtered using a syringe filter with 0.22 μm PVDF membrane, and analyzed by UPLC-MS/MS.

UPLC-MS/MS for accurate quantification

An Agilent ULTIVO triple quadrupole mass spectrometer (Agilent, Palo Alto, CA, USA) coupled to an Agilent 1290 liquid chromatography and autosampler was used for analysis. The ion source was electrospray ionization (ESI) with Agilent Jet Stream Technology used in positive or negative ion mode for all analytes. The data system was MassHunter software version B08 (Agilent). Separations were performed using the Agilent SB-C18 (50 × 2.1 mm, 1.8 μm) column operated at 30 °C. The flow rate of the mobile phase was 0.4 mL/min. The mobile phase consisted of a linear gradient of acetonitrile (A) and 0.1% (v/v) aqueous formic acid (B): 0–2.0 min, 25–35% A (v/v); 2.0–3.5 min, 35–90% A (v/v); 3.5–5.0 min, 90% A (v/v); 5.0–5.1 min, 90–25% A (v/v); 5.1–6 min, 25% A (v/v). The column was reconditioned for 3 min prior to the next injection. A 2 μL sample was injected for each run, with a total run time of 6.0 min. The optimum operating ESI conditions were: gas temperature 350 °C, gas flow rate 10 L/min, nebulizer pressure 50 psi, cell acceleration voltage 4 V. The capillary voltages were optimized to 4000 V in positive mode and 3500 V in negative mode, with equal nozzle voltages (0 V). All metabolites were quantified based on a calibration curve generated by authentic standards.

Statistical analysis

Statistical Product and Service Solutions program (SPSS, version 19) was used for all statistical analyses. Student's *t* test and one-way ANOVA were conducted for group comparisons. Data were presented as means of three biological replicates ± standard deviation. Three biological and technical replicates for each sample were completed.

Results

Identification of CcMYB family and key enzyme genes from phenylpropanoid pathway in pigeon pea

To identify *CcMYB* genes in pigeon pea, the MYB domain (PF00249), 126 AtMYBs were used as probes to screen all the members of the CcMYB gene family in pigeon pea, and further identify 221CcMYB transcription factors by analyzing the conserved domains. A near-neighbor (NJ) evolutionary system method was used to construct a phylogenetic tree of MYB proteins from pigeon pea (221 CcMYBs), *Arabidopsis* (10 AtMYBs) and, soybeans (10 GmMYBs). All selected MYB proteins were divided into nine groups based on phylogenetic analysis (Supplementary Fig. S1). In addition, 122 genes from the phenylpropanoid pathway were identified from the pigeon pea genome, including 19 general phenylpropanoid pathway genes, 46 genes from the flavonoid pathway, and 57 genes involved in lignin synthesis pathways (Supplementary Fig. S2, S3, S4). Detailed information of all identified *CcMYB* genes and phenylpropanoid pathway-related genes were provided in Supplementary Table S1. The number of amino acids encoded by the pigeon pea *CcMYBs* gene from 80 to 1037, and the protein molecular weight (MW) and isoelectric point (pI) of predicted CcMYB proteins ranged from 9.26 to 113.69 kD and 4.55–11.01, respectively. The length of key enzyme genes we focused on in pigeon pea ranged from 160 to 1501 amino acid, and the predicted protein MW and pI range were 17.76 kD to 170.37 kDa, 5.16–9.44 (Supplementary Table S1).

Phylogenetic analysis and classification of the *CcMYB* genes involved in flavonoid and lignin biosynthesis in pigeon pea

To investigate the phylogenetic relationship of the pigeon pea CcMYB proteins involved in flavonoid and lignin biosynthesis, the phylogenetic tree consisting of pigeon pea and other plants (28 R2R3 MYB proteins which involved in the biosynthesis of flavonoid and lignin has been confirmed) was constructed (Fig. 1). 30 *CcMYB* members of *CcMYB* gene family have high identity with the known R2R3-MYB which involved in the biosynthesis of flavonoids and lignin in many species (Supplementary Table S2). The selected 58 *MYB* genes were classified into five groups by phylogenetic analysis, including phenylpropane regulator, monolignol activator, monolignol repressor, flavonol activator, and stilbene activator (Fig. 1a). Sequence alignment showed that 30 *CcMYB* belong to the R2R3-MYB cluster with conserved sequence (-W-(X19)-W-(X19)-W-……-F/I-(X18)-W-(X18)-W-) (Fig. 1b).

Homologous proteins with similar sequences might have similar functions (Zhang et al. 2017). In this work, the amino acid sequence identity of CcMYB114a and AtMYB75, CcMYB5 and AtMYB5 were 60.67% and 56.82%, respectively. AtMYB75, and AtMYB5 have been verified to regulate the biosynthesis of flavonoids and lignin (Zuluaga et al. 2008). It suggests that CcMYB114a and CcMYB5 may cooperatively regulate multiple branches of the phenylpropanoid metabolic pathway. The identity of CcMYB83b, CcMYB83e, and the lignin synthesis pathway activator AtMYB83 (McCarthy et al. 2009) was 69.23% and 87.69%, respectively (Fig. 1a). EgMYB1 is an inhibitor of the lignin biosynthesis pathway and the identity between CcMYB308f and EgMYB1 was 69.23% (Fig. 1a) (Legay et al. 2007). These results provide evidence that CcMYB308f, CcMYB83b, and CcMYB83e may be involved in lignin biosynthesis. The identity between CcMYB12 and AtMYB12 (flavonol synthesis activator) (Mehrtens et al. 2005; Stracke et al. 2007; Luo et al. 2008; Misra et al. 2010; Pandey et al. 2012, 2014) was 56.10%. So CcMYB12 was probably involved in flavonol biosynthesis and promoted the accumulation of flavonol in pigeon pea. The results showed CcMYB14a and CcMYB14b were close to AtMYB15 and AtMYB14 based on multiple sequence alignment, with 48.83% and 51.20% similarity. AtMYB15 and AtMYB14 have been verified to regulate stilbene biosynthesis in recent studies (Höll et al. 2013) (Fig. 1a) indicated that CcMYB14a and CcMYB14b were likely to perform the same function in pigeon pea. It is not hard to see from the results, the selected 30 CcMYB transcription factors might be involved in the phenylpropanoid pathway. However, the functions of these transcription factors need to be further studied.

Gene structure and motif composition of *CcMYB* genes

The structural composition of 30 *CcMYB* genes (the number and distribution of introns and exons) were analyzed to understand the evolutionary imprint of MYB proteins. Gene structure analysis showed that the number of introns in different *CcMYB* genes was not the same. Most CcMYB genes contain two introns, and the six *CcMYB* genes (*CcMYB46*, *CcMYB83a*, *CcMYB83b*, *CcMYB83c*, *CcMYB83d*, and *CcMYB83e*) only contained one intron (Fig. 2a). The length

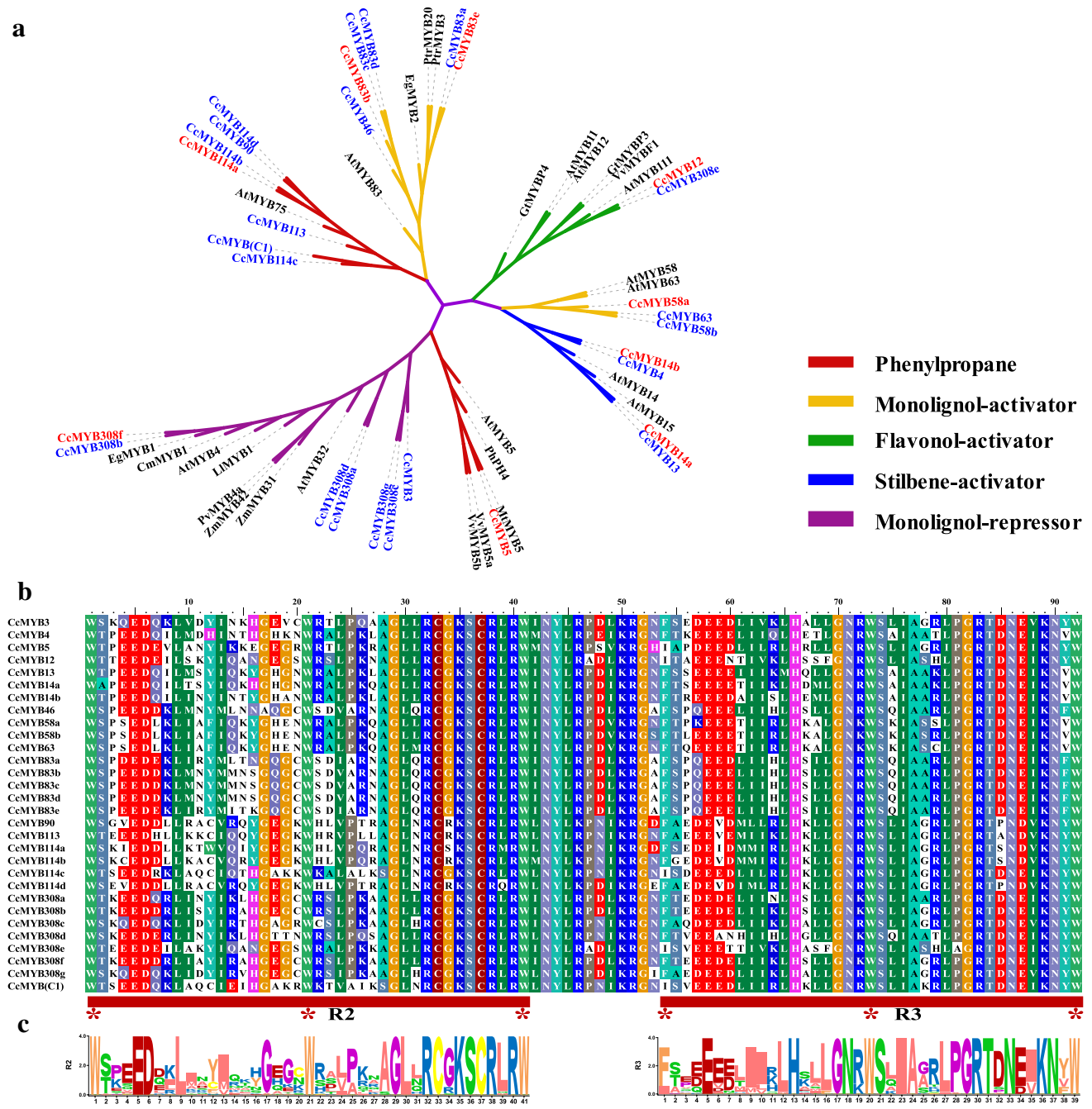

Fig. 1 Phylogenetic analysis and classification of the *R2R3-CcMYB* genes. **a** Phylogenetic analysis of R2R3-CcMYB protein sequences. The developmental evolutionary tree was constructed using the neighbour-Joining (NJ) algorithm of MEGA7 software. Bootstrapping with 1,000 replications was performed. R2R3-MYB marked in red label represents its higher similarity to MYB that the function has been determined in other plants. Detailed information of these *MYB* genes was provided in Supplementary Table S2. **b** Multiple sequence alignment of MYB proteins in pigeon pea plant. Sequences were aligned using BioEdit software. **c** Conserved motif logo of R2 and R3

of exons was similar among closely related *CcMYBs*, but the length of the introns was significantly different. The intron and exon characteristics of the *CcMYB* gene are consistent with the results of the phylogenetic tree (Fig. 2a). In addition, ten conserved motifs of CcMYB proteins were identified by MEME analysis (Fig. 2a, b). The diversity of gene structure and motifs compositions of pigeon pea *CcMYB* may lead to the diversification of its functions.

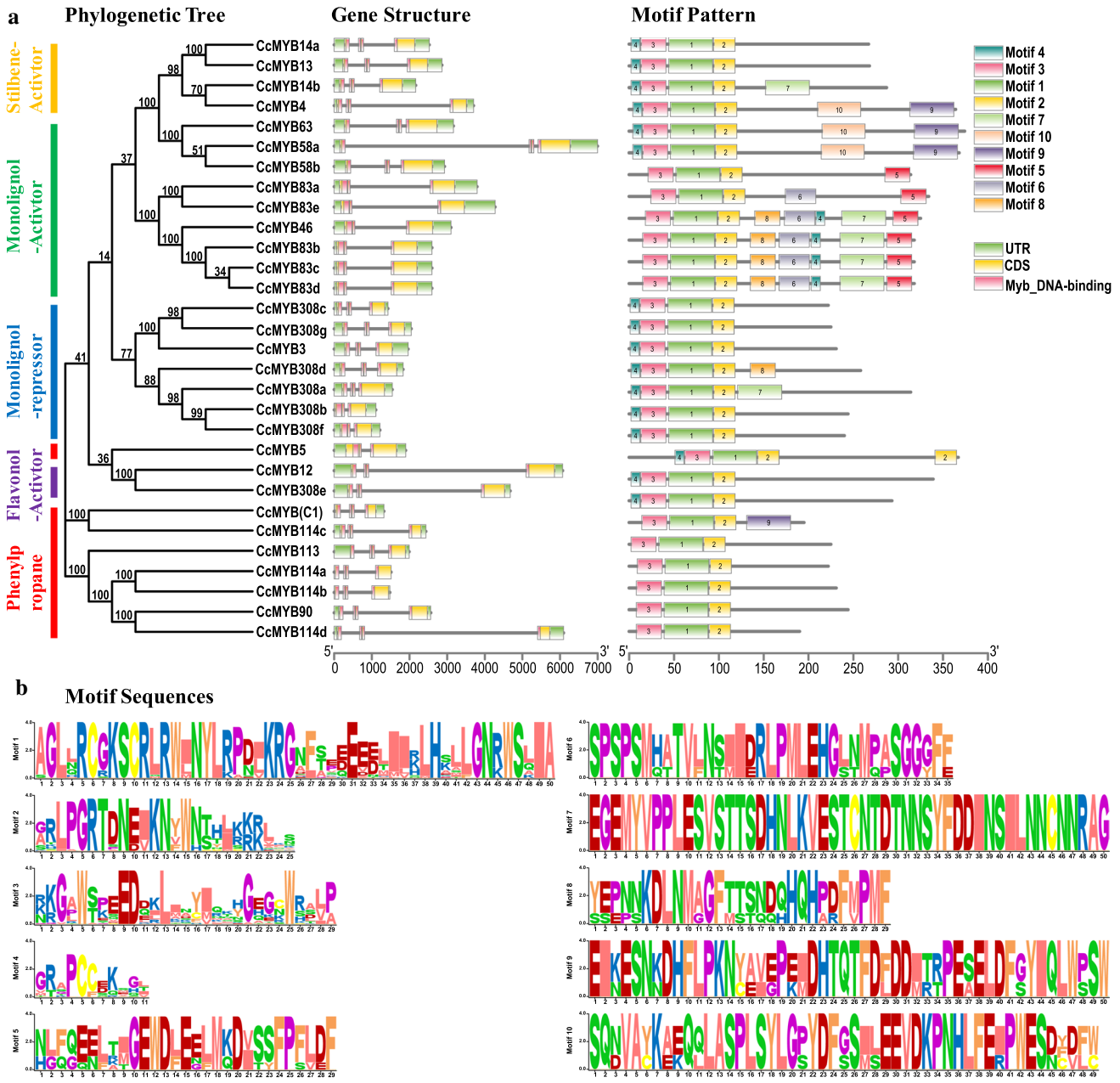

Fig. 2 Phylogenetic relationships, gene structure and architecture of conserved protein motifs in 30 *R2R3-MYB* genes involved in phenylpropanoid pathway from pigeon pea. **a** Phylogenetic tree, gene structure and conserved motifs of R2R3-MYB. The neighbor-joining (NJ) tree on the left includes 30 R2R3-MYB proteins from pigeon pea. According to the function, the 30 R2R3-MYB proteins were divided into clustered into fivecategories. The gene structure of *R2R3-MYBs* from pigeon pea are performed in the middle. Schematic representation on the right of conserved motifs (obtained using MEME) in 30 R2R3-MYB proteins. Different motifs are represented by boxes of different colors. **b** The logo of ten conserved motifs

Evolutionary patterns and divergence

To investigate the distribution of all genes we focused on in the genome of pigeon pea, the genes were plotted on the corresponding chromosomes (Fig. 3a). Most *CcMYB* genes and key genes from the phenylpropanoid pathway were located on chromosome 11, whereas chromosome 10 contained only two genes (Fig. 3a). Ten genes on chromosome 3,4, nine genes (*Cc4CL5*, *CcCAD8/9/12*, *CcCCoAOMT9*, *CcF5H1*, *CcPAL3*, *CcMYB(C1)*, *CcLDOX*) on chromosome 6, eight genes (*CcCCoAOMT4/5/6/7/8*, *CcCHS8*, *CcCAD7/11*) on chromosome 5, seven genes (*CcCCR11*, *CcC4H2*, *CcMYB308g*, *CcLAC1*, *CcFNSII4*, *CcMYB14a*, *CcCAD6)* on chromosome 2, six genes (*CcCHS9/11*, *CcMYB308c*,

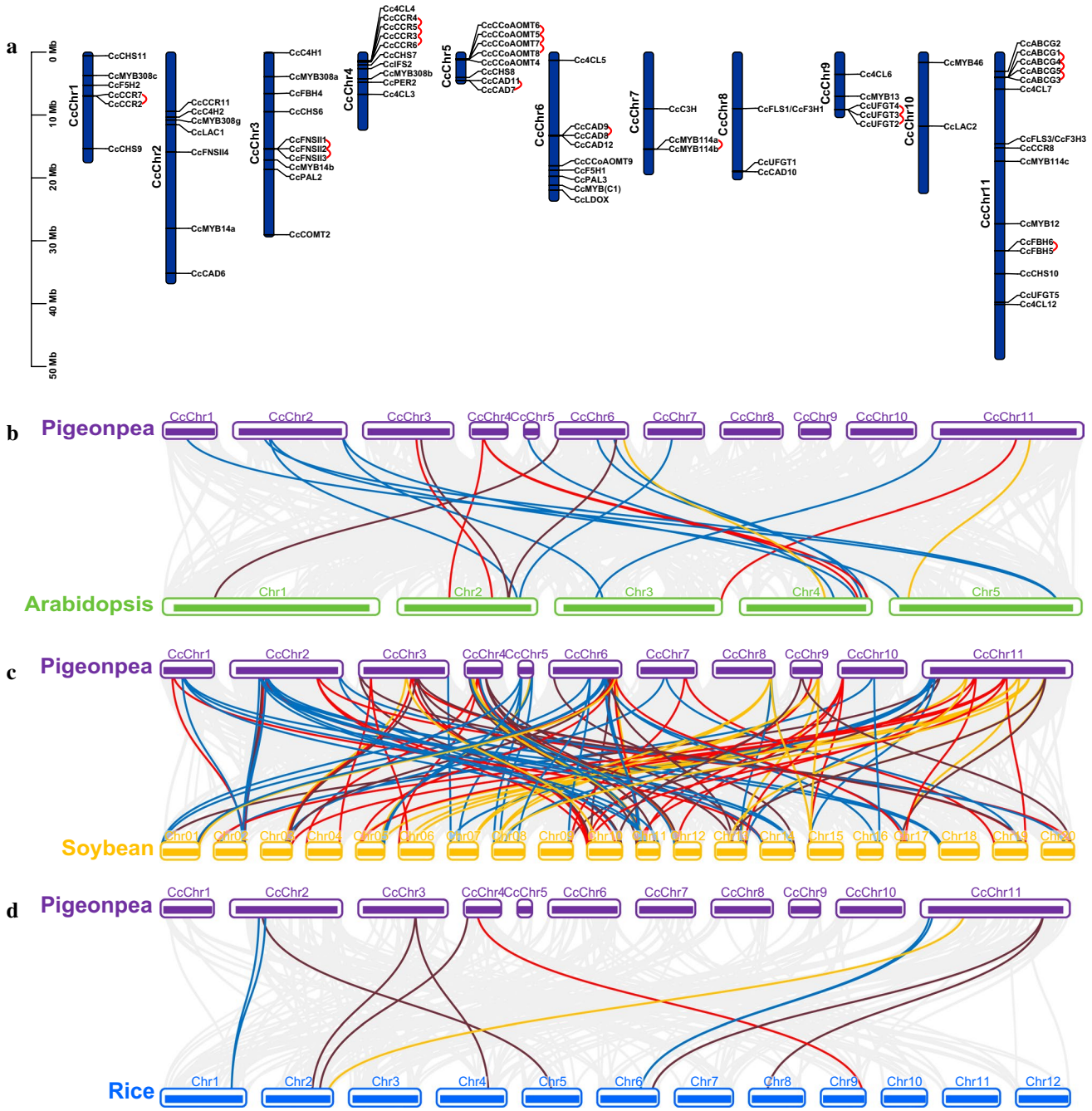

Fig. 3 Chromosomal locations, gene duplication and synteny analysis. **a** 30 *R2R3-MYBs*, 122 key enzyme genes from phenylpropane pathway distribution across 11 chromosomes of pigeon pea genome. Only 78 genes are mapped to the 11 chromosomes. The scale represents the length of pigeon pea chromosomes. Red lines represent the tandem duplication. **b–d** Gene duplication and synteny analysis of related genes involved in phenylpropanoid pathway between pigeon pea and three model species. Gray lines in the background indicated the collinear blocks, while the red, blue, yellow and brown lines highlight the R2R3-MYB, general phenylpropanoid pathway, flavonoid pathway and lignin pathway-related genes syntenic genes pairs

CcF5H2, *CcCCR2/7*) on chromosome 1, five genes (*Cc4CL6*, *CcMYB13*, *CcUFGT2/3/4*) on chromosome 9 and three genes on chromosome 7 and chromosome 8 (Fig. 3a). In this study, 28 tandem duplicated genes were identified. Ten genes were divided into five groups (*CcCCR2/7*, *CcCAD7/11*, *CcCAD8/9*, *CcMYB114a/b*, and *CcFBH5/6*). Six genes were divided into two groups (*CcFNSII1/2/3* and *CcUFGT2/3/4*). Twelve genes were divided into three groups (*CcCCR3/4/5/6*, *CcCCoAOMT5/6/7/8*, and *CcABCG1/3/4/5*) (Fig. 3a). From these results, it could be concluded that some

CcMYB genes and phenylpropanoid pathway-related genes were produced by gene duplication events.

To investigate the evolutionary mechanisms of the *CcMYB* genes and key genes from the phenylpropanoid pathway, three synteny analysis of all genes with the other three typical plants were constructed, including two dicotyledonous plants (*Arabidopsis thaliana* and *Glycine max*), and a monocotyledonous plant (*Oryza sativa*) (Fig. 3b–d). Finally, 22 collinear gene pairs between pigeon pea and *Arabidopsis* (Fig. 3b), 146 orthologs between pigeon pea (Fig. 3c) and soybean, and 15 orthologs between rice and pigeon pea were identified (Fig. 3d). The details of the gene pairs were shown in Supplementary Table S3, S4, S5. The number of orthologous events of pigeon pea-soybean was far greater than that of pigeon pea-*Arabidopsis* and pigeon pea-rice, and the closer evolutionary distance between pigeon peas and soybeans was confirmed. Our study may indicate that all genes examined in pigeon pea

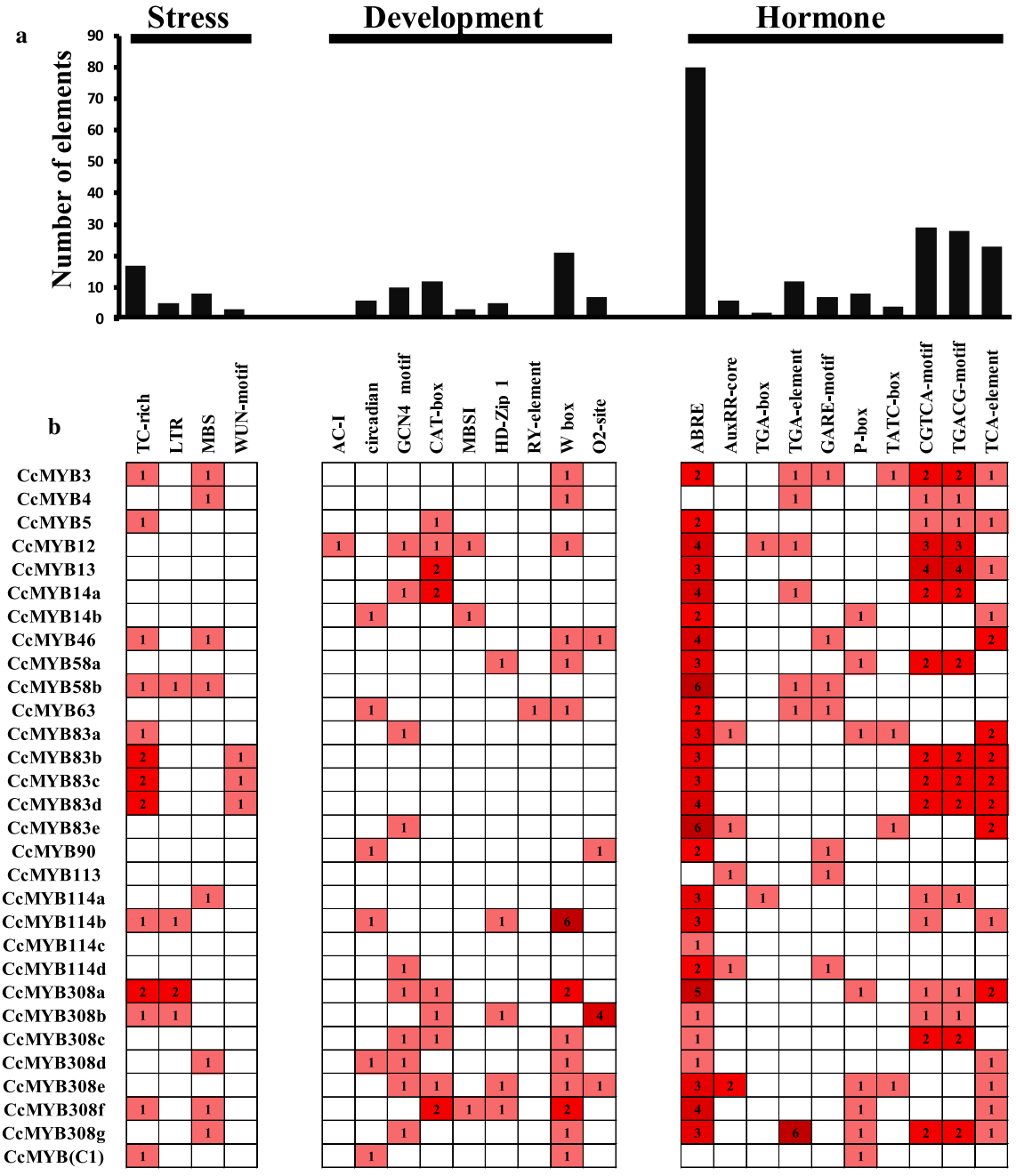

Fig. 4 Various cis-acting elements in 30 *R2R3-MYB* genes. **a** The number of *cis*-acting elements in response to various factors. **b** The number of occurrences of each *cis*-acting element

share a similar structure and function with *GmMYB* genes and the key genes from the phenylpropanoid pathway in soybean.

The analysis of the cis-elements in the promoter regions of *CcMYB* genes and the MYB-binding site of key genes from the phenylpropanoid pathway

To explore the regulatory mechanism of *CcMYB* genes, 2 kb upstream sequences from the translation initiation sites of *CcMYB* genes were analyzed using tools at the PlantCARE database to identify potential cis-acting elements. Various cis-acting elements, including stress, development, and hormone-responsive elements, were detected in the promoter regions of *CcMYB* genes (Fig. 4a, b). These results suggest that their expressions are controlled by complex regulatory networks.

For stress-related cis-acting elements, defense and stress-responsive element (TC-rich repeats), low-temperature responsive element (LTR), MYB binding sites (MBSs) involved in drought inducibility and, wound responsive element (WUN-motif) were detected in the promoters of 13, 4, 8, and 3 *CcMYB* genes, respectively (Fig. 4a, b). Among these stress-related cis-acting elements, TC-rich repeats were detected with the highest frequency (a total of 17 TC-rich repeats located in 13 *CcMYB* promoters), followed by MBSs, LTR, and WUN-motif. Various cis-elements related to hormone response elements were found in the *CcMYB* promoters (Fig. 4a, b). An abscisic acid responsive element (ABRE) was present in 27 of the *CcMYB* promoters (a total of 80 ABRE located in 27 *CcMYB* promoters). 16 *CcMYB* promoters contained a MeJA responsive element (CGTCA-motif, and TGACG-motif) that is involved in response to MeJA stress (a total of 57 MeJA responsive elements located in 16 CcMYB promoters). These suggested that *CcMYB* genes play important roles in abiotic stress responses. Additionally, auxin responsive elements (AuxRR-core, TGA-box and TGA-element) were found to be present in 13 *CcMYB* promoters, salicylic acid responsive elements (TCA-element) were found to be present in 16 *CcMYB* promoters and gibberellin responsive elements (GARE-motif) were found to be present in 7 *CcMYB* promoters (Fig. 4a, b).

To identify if the key genes from the phenylpropanoid pathway are regulated by the *CcMYB*, the cis-acting elements of the 2 kb promoter upstream of the start codon of related genes were analyzed and the genes with the MYB binding sites were identified (Supplementary Fig. S5). The results showed that 89 genes have MYB binding sites, *CcPAL2* was detected with the highest frequency of 9 MYB binding sites, *CcC3H* promoter has 7 MYB binding sites, *CcCHS2*, *CcUFGT3*, *CcCCR9*, and *CcCCoAOMT10* promoters have 6 MYB binding sites, *CcPAL3*, *CcCCR2*, *CcCCR12*, and *CcCOMT3* promoters have 5 MYB binding sites (Supplementary Fig. S5). These genes may be regulated by MYB transcription factors and participate in flavonoid and lignin biosynthesis.

Deep transcript abundance profiling of genes by RNA-seq

As MeJA is an important regulator in the biosynthesis of flavonoid and lignin (Li et al. 2020b; Chen et al. 2020b; Cao et al. 2010), the response mechanism under MeJA stress and the possible regulatory relationship between *CcMYB* and related genes were analyzed according to the RNA-seq of pigeon pea under MeJA treatment. The raw data included 18 M 150 bp paired-end reads. Each base was assigned a quality score using FastQC. The results showed that the data were highly credible with Q20 and showed that the quality of the data was very good. A heatmap of *CcMYB* and related genes from the phenylpropanoid pathway was generated with corresponding FPKM values of MeJA treatment using the TBtools (Fig. 5). The transcript abundance of 30 *CcMYB* genes, 19 general phenylpropanoid pathway structural genes, 46 flavonoid pathway structural genes, and 57 lignin pathway-related genes were evaluated. Most *CcMYBs* and key genes from the phenylpropanoid pathway were down-regulated in response to MeJA treatment, and the expression levels increased slightly as the stress treatment time prolonged. However, ten MYB transcription factors (*CcMYB4/CcMYB5/CcMYB12/CcMYB13/CcMYB14b/CcMYB308a/CcMYB308b/CcMYB308c/CcMYB308d/CcMYB308f/CcMYB308g*) have significantly responded to MeJA treatment. It is reported that genes have co-expression effects of the same metabolic pathway.

To further identify the related genes from flavonoid and lignin biosynthesis pathways that may be regulated by *CcMYB*, the co-expression trend from the *CcMYB* gene family and related genes were analyzed. The co-expression similarity ≥ 0.95 was shown in Fig. 6 to retain the target genes. It is noteworthy that the expression trend of *CcMYB12* under MeJA treatment is similar to those of *CcCCR12*, *CcCHS3*, *CcFBH6*, *CcPAL3*, and *CcLAC2* (Fig. 6). The expressions of *CcMYB14b*, *CcLAC1*, *CcCAD6*, *CcCCoAOMT1*, and *CcANR1* were slightly higher than that of the control group at 3 h after MeJA treatment (Fig. 6). However, their expression was down-regulated with the prolongation of MeJA treatment. In pigeon pea, the expressions of *CcMYB5* and *CcCAD10*, *CcMYB308a*, and *CcUFGT2*, *CcMYB308c* and *CcCAD8* were consistent with the above. They were up-regulated to varying degrees when induced by MeJA at 3 h, 6 h, and 12 h (Fig. 6). This suggested that there are co-expression effects among these ten *CcMYB* genes and the related

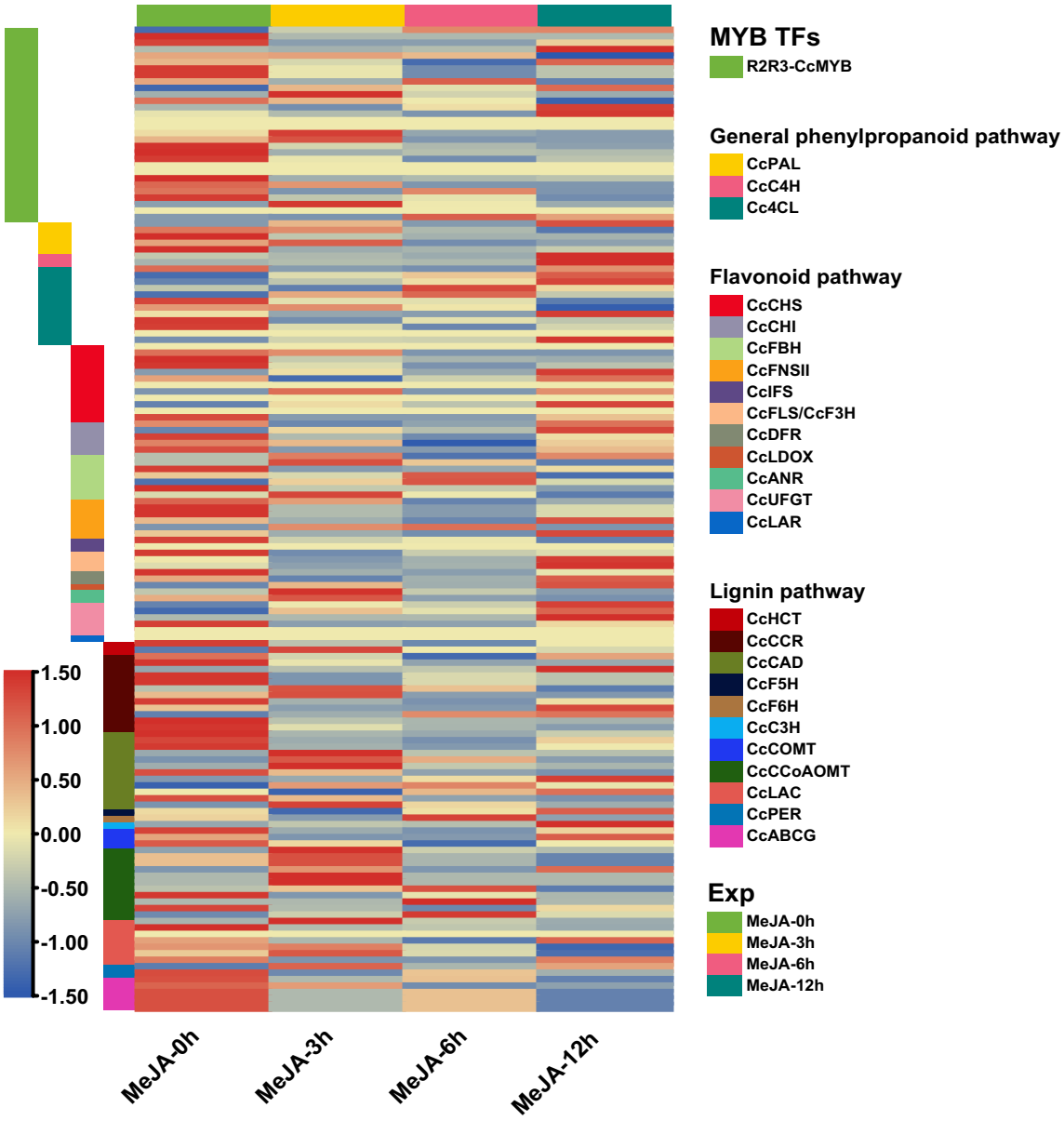

Fig. 5 The expression profile analysis of 122 key enzyme genes from phenylpropanoid pathway and 30 *CcMYB* under MeJA treatment in pigeon pea

genes from the phenylpropanoid pathway in response to MeJA treatment.

Prediction of regulatory networks between CcMYB12, CcMYB14b, CcMYB5, CcMYB308f and related genes from phenylpropanoid pathway

In this study, four CcMYB proteins (CcMYB12, CcMYB14b, CcMYB5, and CcMYB308f) from phylogenetic trees (Fig. 1) and significantly responded to MeJA, were screened in pigeon pea and which shared close relationships with their homologs in *Arabidopsis*. To identify their functional and regulatory of CcMYBs, STRING software was used to draw a regulatory network map in pigeon pea and related genes in *Arabidopsis* (Fig. 7, Supplementary Table S6). The 4 CcMYB proteins exhibited strong regulation with many phenylpropanoid pathway-related proteins, such as flavonol synthase protein, basic helix-loop-helix (bHLH) DNA-binding protein, chalcone flavanone isomerase protein, chalcone and stilbene synthase protein, leucoanthocyanidin dioxygenase protein, and dihydroflavonol reductase.

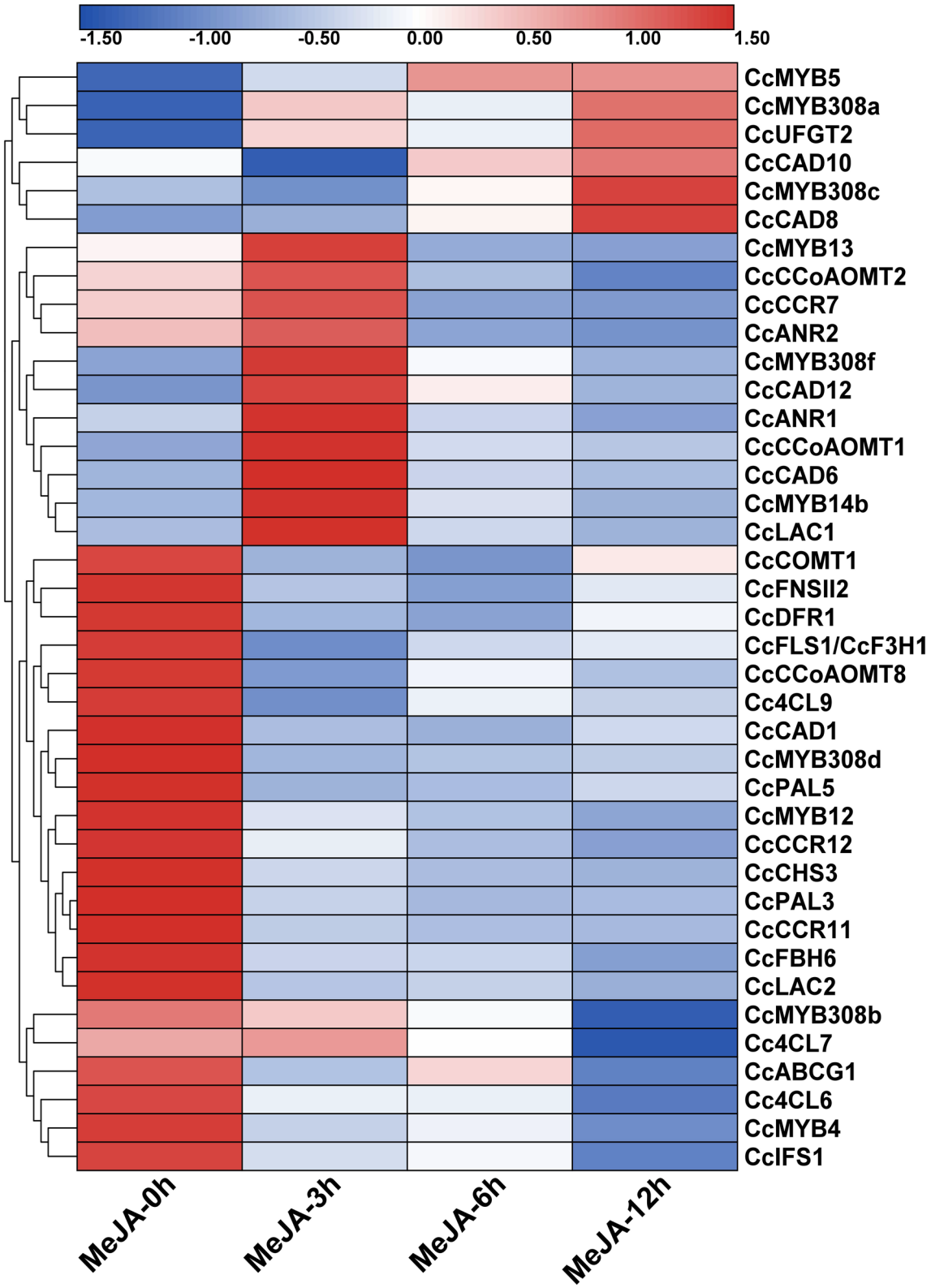

Fig. 6 The co-expression of key genes and *CcMYB* under MeJA treatment. Hierarchical clustering of expression profiles in response to MeJA hormone induction

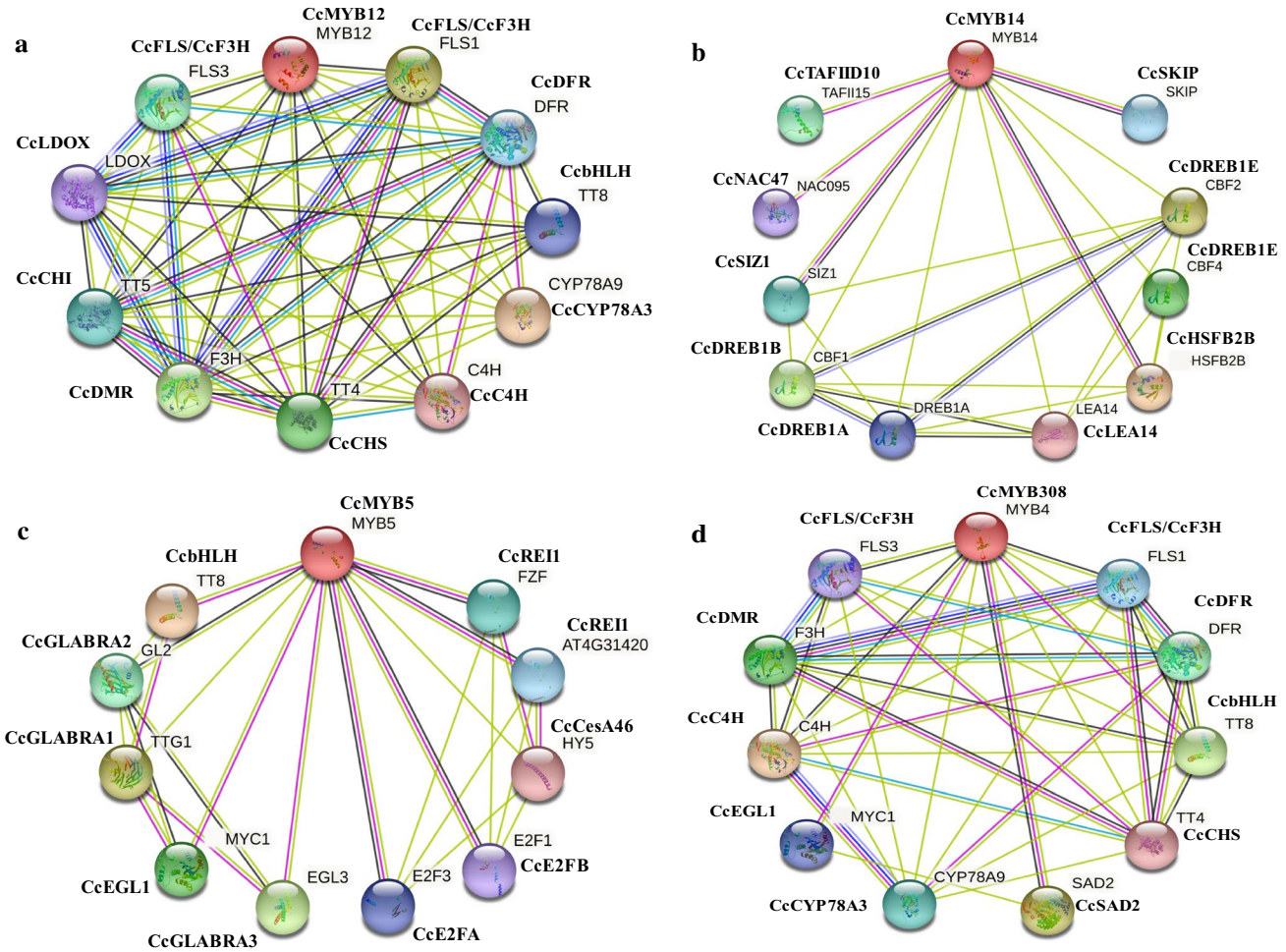

Fig. 7 Protein regulatory network of CcMYB12 (**a**), CcMYB14 (**b**), CcMYB5 (**c**) and CcMYB308 (**d**). A network of regulatory between the CcMYB and phenylpropanoid pathway related gene in pigeon pea. The purple lines represent the interaction from experiment results. The yellow-green, black and blue lines represent text mining, co-expression and protein homology respectively. Genes with a gray background are from *Arabidopsis*, and genes with a white background are from pigeon pea

Expression profiles of *CcMYB12*, *CcMYB14b*, *CcMYB5*, *CcMYB308f* and six key genes correlate with the flavonoid accumulations in pigeon pea

Six different tissues were selected to determine the flavonoids accumulation patterns in pigeon pea (Fig. 8a). Fifteen representative flavonoids, such as quercetin, naringenin, cajanolactone A, pinostrobin, orientin, isovitexin, vitexin, genistin, isorhamnetin, luteolin, biochanin A, calycosin, genistein, apigenin, and formononetin were quantitatively analyzed in different tissues (Fig. 8b). Moreover, cajaninstilbene acid, which belongs to stilbene was detected as well because it is a specific metabolite in pigeon pea and probably derived from the phenylpropanoid pathway. The accumulation patterns of flavonoids are significantly different in different tissues of pigeon pea. Genistin, luteolin, genistein apigenin, and cajanolactone A mainly accumulated in roots, whereas naringenin, biochanin A, and formononetin mainly accumulated in stems (Fig. 8b). Comparatively, cajaninstilbene acid, orientin, pinostrobin, isovitexin, vitexin, calycoin, and isorhamnetin were most highly accumulated in leaves in contrast to other tissues (Fig. 8b).

To unravel the molecular basis of flavonoids and lignin accumulation patterns, ten candidate genes including *CcMYB12*, *CcMYB14b*, *CcMYB5*, *CcMYB308f*, and six key enzyme genes from the phenylpropanoid pathway were selected, and their expression profiles were completely discussed (Fig. 9). As shown in Fig. 9, *CcMYB12* and *CcMYB308f* expression levels were higher in roots compared to other tissues, indicating that they might play more important roles in genistin, luteolin, genistein, and apigenin biosynthesis. *CcMYB5* expression levels were higher in stems compared to other tissues (Fig. 9), indicating that *CcMYB5* might play more important roles in flavonoids biosynthesis from stems. The expression profile is similar to the results of the phylogenetic analysis, indicating that *CcMYB5*

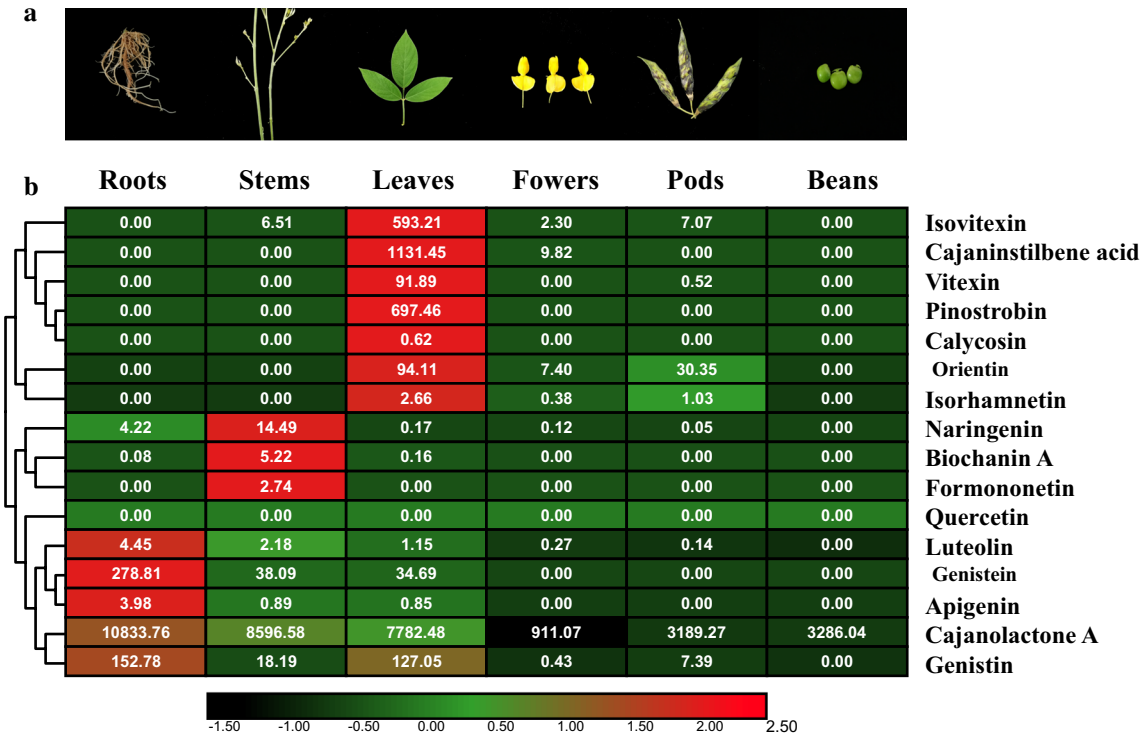

Fig. 8 The flavonoids accumulation in different tissues of pigeon pea. **a** The six different tissues from pigeon pea. **b** Flavonoids contents at different tissues. The value represents the flavonoids accumulation of per gram dry weight material (μg/g). Clustering based on the similarity matrix, whose practical development uses row-standardization

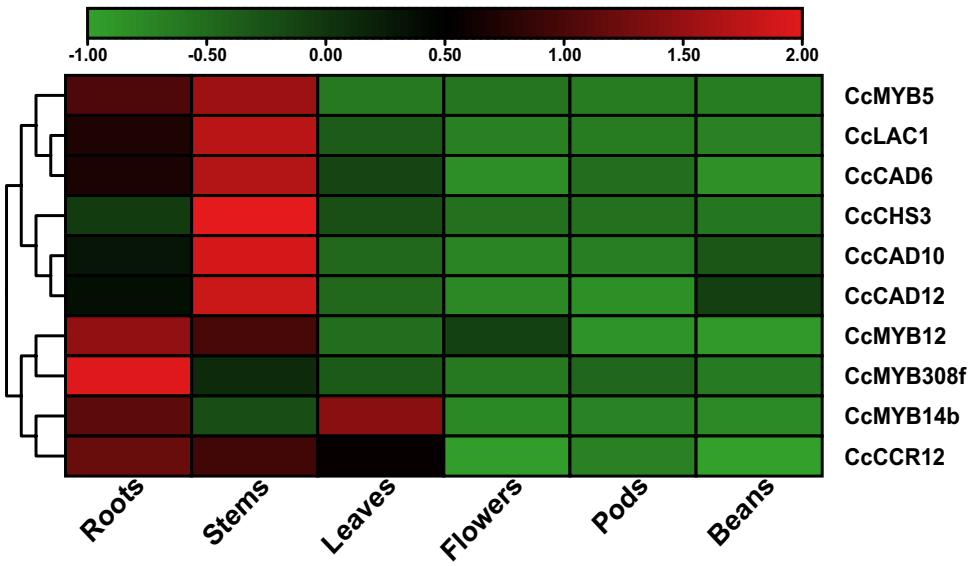

Fig. 9 Expression patterns of *CcMYB12*, *CcMYB14b*, *CcMYB5*, *CcMYB308f*, and six key genes from phenylpropanoid pathway in different tissues of pigeon pea. The data represent mean ± SD of three biological replicates

may cooperatively regulate the biosynthesis of flavonoids and lignin. Similarly, the results from the phylogenetic tree show that *CcMYB14b* may be involved in the biosynthesis of stilbene and *CcMYB14b* was a highly expressed level in the leaves (Fig. 9), suggesting that *CcMYB14b* is potentially involved in the biosynthesis of cajaninstilbene acid. From the results, we can find that the 4 *R2R3-CcMYB* and 6 key enzyme genes from the phenylpropanoid pathway have similar expression trends (Fig. 9). To summarize the above results, the biosynthesis of flavonoids and lignin is strictly regulated by the corresponding R2R3-MYB transcription factor. However, the mechanism of the CcMYB transcription

factor regulating the biosynthesis of flavonoids and lignin needs more studies.

Expression patterns of *CcMYB12*, *CcMYB14b*, *CcMYB5*, *CcMYB308f*, and six related genes in response to different abiotic stress

To characterize the abiotic stress-responsive *CcMYB* and phenylpropanoid pathway key enzyme genes we focused on, the expression levels of ten selected genes were determined by quantitative real-time PCR in different abiotic stress (UV-B, cold, heat, salt and drought, and hormone). From the results, we can find the expression of *CcMYB12, CcMYB14b, CcMYB5, CcMYB308f* were induced by various abiotic stresses (Fig. 10). *CcMYB5*, *CcMYB12*, *CcMYB14b*, showed low transcription levels under UV-B treatment and poor responses to UV-B. Conversely, *CcMYB308f* showed positive responses to UV-B treatment (Fig. 10). All *CcMYB* genes we focused on were down-regulated under cold stress. However, they were up-regulated under heat stress, indicating that the expression of these genes was significantly induced by heat stress. *CcMYB12* was responsive to GA_3 stress as well, but it was not changed in drought stress and salt stress. This suggests that *CcMYB12* may not participate in salt and drought stress, but is sensitive to high temperature (Fig. 10). *CcMYB5*, *CcMYB14b, CcMYB308f* were up-regulated under drought and salt stress, indicating the regulatory role of *CcMYB* genes has a wide array of abiotic stress responses.

CcCHS3, which is the first key enzyme gene from the flavonoid synthesis pathway was up-regulated under UV-B, heat, cold and GA_3 treatments, and showed low transcription levels under drought and salt stress. However, the *CAD6* gene of the lignin synthesis pathway was up-regulated under drought and salt stress, and other key enzyme genes (*CcCAD10*, *CcCAD12*, *CcCCR12*, and *CcLAC1*) from the lignin pathway are more sensitive to drought and salt stress (Fig. 10). The expression levels of *CcMYB5* and *CcMYB308f* genes were up-regulated as well. Consistent with the results of the developmental tree and promoter elements, suggesting that *CcMYB5* and *CcMYB308f* may participate in the biosynthesis of lignin in pigeon pea (Fig. 12).

Based on the above results, it is easy to find that the selected genes derived from the lignin and flavonoid pathway have different responses to abiotic stress, which was probably due to the specific expression exhibited using the same substrate to compete for carbon sources. Our results strongly supporting the role of the selected genes in abiotic stress tolerance in crop plants.

Discussion

In recent years, the whole genome sequencing of many plants has been completed based on the widespread application of high-throughput sequencing technology. MYB gene family members are known to play important roles in plant secondary metabolism such as the phenylpropanoid metabolism pathway (Ma and Constabel 2019; Liu et al. 2015; Borevitz et al. 2000). MYB gene families were systematically identified in a variety of plants by genomic analysis (Wei et al. 2020; Li et al. 2020a). However, to date, papers about the identification and functional role of the *MYB* genes involved in flavonoid and lignin biosynthesis in pigeon pea have not been reported yet. In the present investigation, 30 *CcMYB* genes involved in flavonoid and lignin biosynthesis and 122 key enzyme genes from the phenylpropanoid pathway were first identified in the pigeon pea genome. The gene structure, cis-acting elements analysis, chromosomal distribution, phylogenetic analysis, gene duplication events, synteny analysis, transcript abundance profiling, expression profile analysis in various tissues, and flavonoids accumulations characteristics were determined.

The amino acid sequence length differences of the 30 *CcMYB* genes were identified, and the complexity was confirmed in the pigeon pea genome. The evolution of the gene family was determined by the composition of the gene structure. (Xiao et al. 2017; Xu et al. 2012). Most members of the *CcMYB* genes in pigeon pea have two introns. Our results are consistent with the gene structure of identified *MYB* genes family in other plants. (Li et al. 2020a; Sun et al. 2019). The physical and chemical properties of CcMYB proteins (such as the molecular weight and isoelectric point) are significantly different, which may also be the reason for the functional diversity of CcMYB proteins. (Feng et al. 2017). In addition, the *CcMYB* gene has tandem duplication events and has closer collinearity with soybeans. These findings are consistent with the identification results of *MYB* genes in other plants. (Du et al. 2012).

In this work, we identified 122 key enzyme genes that may be involved in the biosynthesis of flavonoid and lignin in pigeon pea and analyzed the transcription abundance of these gene family members through RNA-seq (Fig. 11). As shown in Fig. 11, *CcPAL3*, *CcC4H2*, *Cc4CL9*, *CcCHS12*, *CcCHI3*, *CcIFS1*, *CcFNSII6*, *CcFBH4*, *CcDFR2*, *CcFLS1*, *CcFLS3*, *CcLDOX*, *CcUFGT4*, *CcANR2*, *CcCCR2*, *CcCAD1*, *CcHCT1*, *CcC3H*, *CcCCoAOMT1*, *CcCOMT1*, *CcF5H1*, *CcLAC5*, and *CcABCG5* have the highest transcription abundance compared to other genes from flavonoid and lignin biosynthesis pathway. In addition, 65 genes of

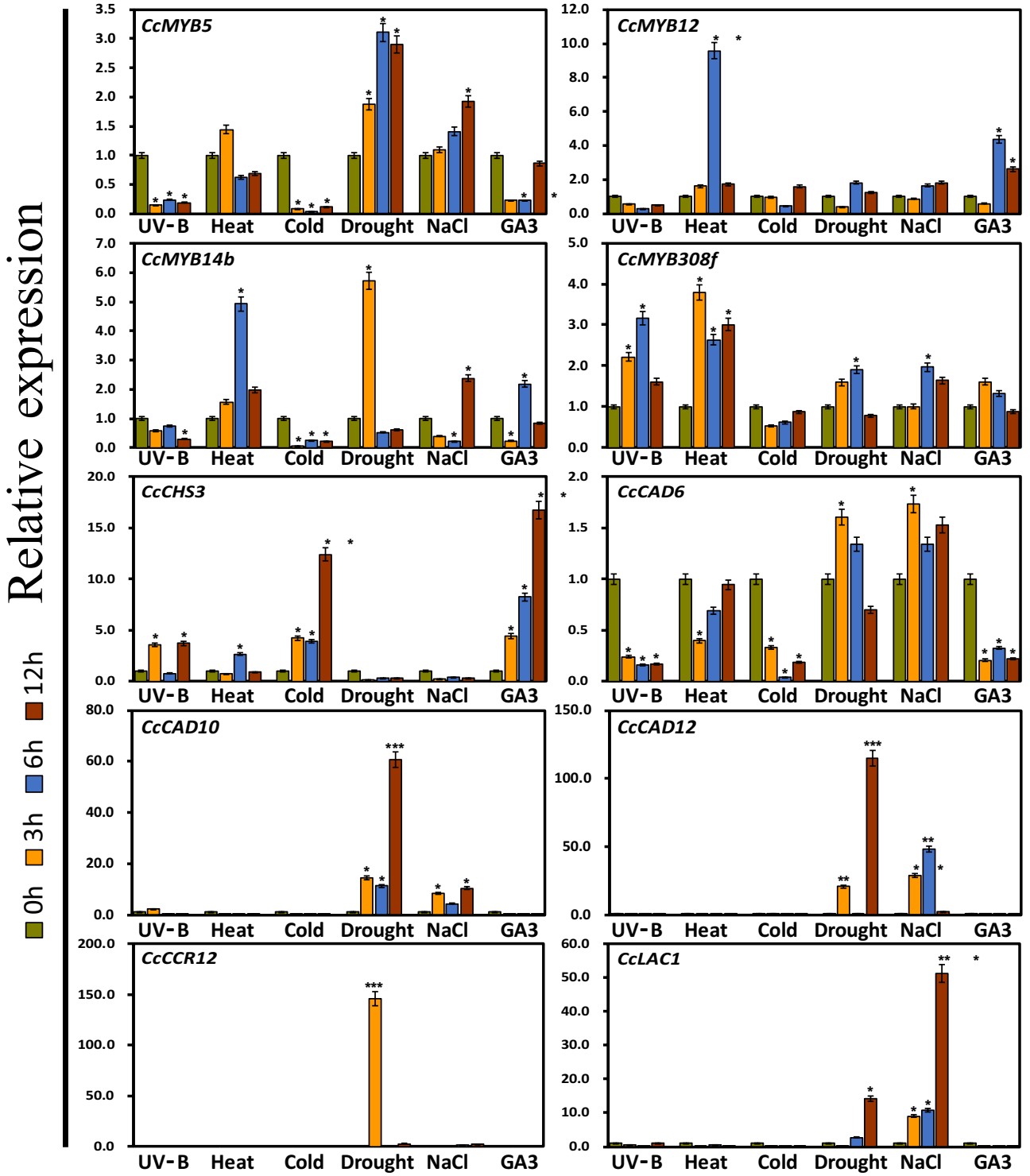

Fig. 10 The relative expression of *CcMYB12*, *CcMYB14b*, *CcMYB5*, *CcMYB308f*, and six related genes was determined by six stress treatments of pigeon pea. Data are shown as the mean value ± SD from three independent assays. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ by Student's *t* test

them were significantly induced by MeJA (Fig. 11). Pearson's correlation coefficient was used to quantitatively analyze the expression patterns correlation between these 65 key enzyme genes and the selected 30 *CcMYBs* under MeJA treatment (Fig. 11). From the results, we can find the co-expression effects of 10 *CcMYB* genes, and 28 key enzyme

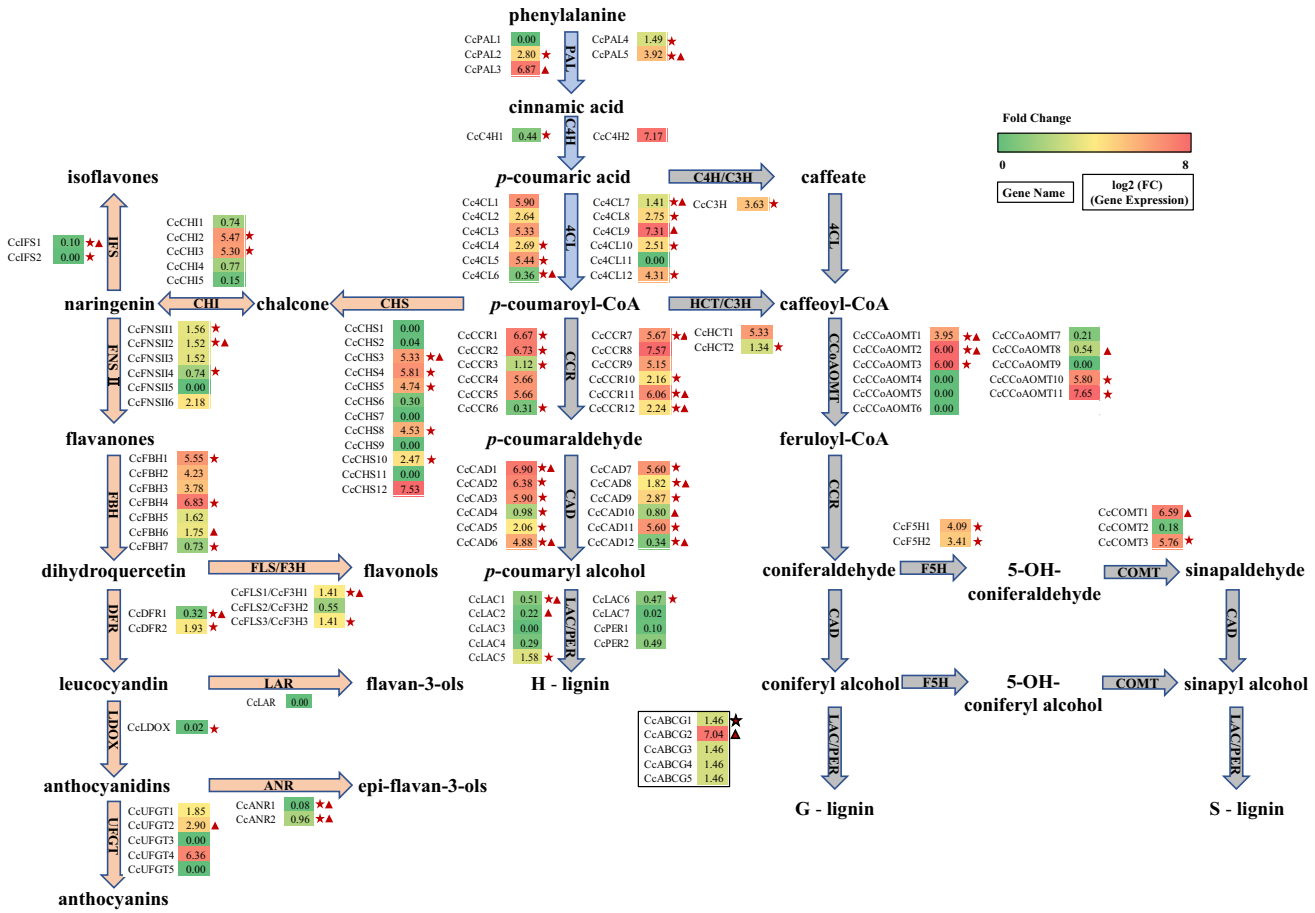

Fig. 11 Expression of structural genes in phenylpropanoid pathway. Enzyme gene highlighted with red is high expression, while enzyme gene highlighted with green is low expression in wild-type plant tissue. The red star represents the differential expression genes by MeJA stress. The red triangle represents the similar expression trend of structural gene and MYB transcription factor by MeJA stress. *PAL* phenylalanine ammonia-lyase, *C4H* cinnamate-4-hydroxylase, *4CL* 4-coumarate: CoA ligase, *CHS* chalcone synthase, *CHI* chalcone isomerase, *IFS* isoflavone synthase, *FNS II* flavone Synthase II, *F3H* flavanone 3-hydroxylase, *FLS* flavonol synthase, *LAR* leucoanthocyanidin reductase, *DFR* dihydroflavonol 4-reductase, *LDOX* leucoanthocyanidin dioxygenase, *UFGT* UDP flavonoid glucosyl transferase, *HCT p-hydroxycinnamoyl CoA* quinate/shikimate *p*-hydroxycinnamoyl transferase, *C3H* coumarate 3-hydroxylase, *CCoAOMT* caffeoyl-CoA *O*-methyltransferase, *CCR* cinnamoyl-CoA reductase, *F5H* ferulate 5-hydroxylase, *COMT* caffeic acid *O*-methyltransferase, *CAD* cinnamyl alcohol dehydrogenase, *ABCG* ATP-binding cassette transporters G, *LAC* laccase, *PER* peroxidase

genes were identified under MeJA treatment (Fig. 6). Among them, the identity in Pearson's correlation coefficient of *CcMYB12* and *CcCCR12*, *CcCHS3* were 0.998 and 0.995, respectively. The correlation coefficients between *CcMYB5* and *CcCCR10* were 0.971. Pearson's correlation coefficient of *CcMYB308f* and *CcCAD12* was 0.992. *CcMYB14b* and *CcLAC1* have a higher correlation coefficient, which was 0.999. They implied that *CcMYB12*, *CcMYB5*, *CcMYB308f*, and *CcMYB14b* may have co-expression effects with the key genes, and are involved in the flavonoid and lignin biosynthetic.

The specific expression of the MYB transcription factor allows it to regulate a variety of plant biological and physiological processes. (Ambawat et al. 2013). Studying gene expression profiles is crucial to discovering their growth, development, and metabolic processes in plants. In pigeon pea, the transcripts abundance of examed *CcMYB12*, *CcMYB14b*, *CcMYB5*, *CcMYB308f*, and six key genes were confirmed that associated with different tissues, and the regulatory mechanisms of flavonoid and lignin biosynthesis in different plant tissues are significantly different. In this study, the flavonoids accumulation in six different tissues is closely related to the tissue-specific *CcMYB*.

Recent studies demonstrated that the type and number of cis-acting elements in the promoter sequence can reflect the trend of gene response to different stress. The promoters of *CcMYBs* contain a variety of abiotic stress response elements, including drought, temperature, and light. (Fig. 4). Our study exhibited that *CcMYBs* showed different responses under UV-B, drought, temperature, and

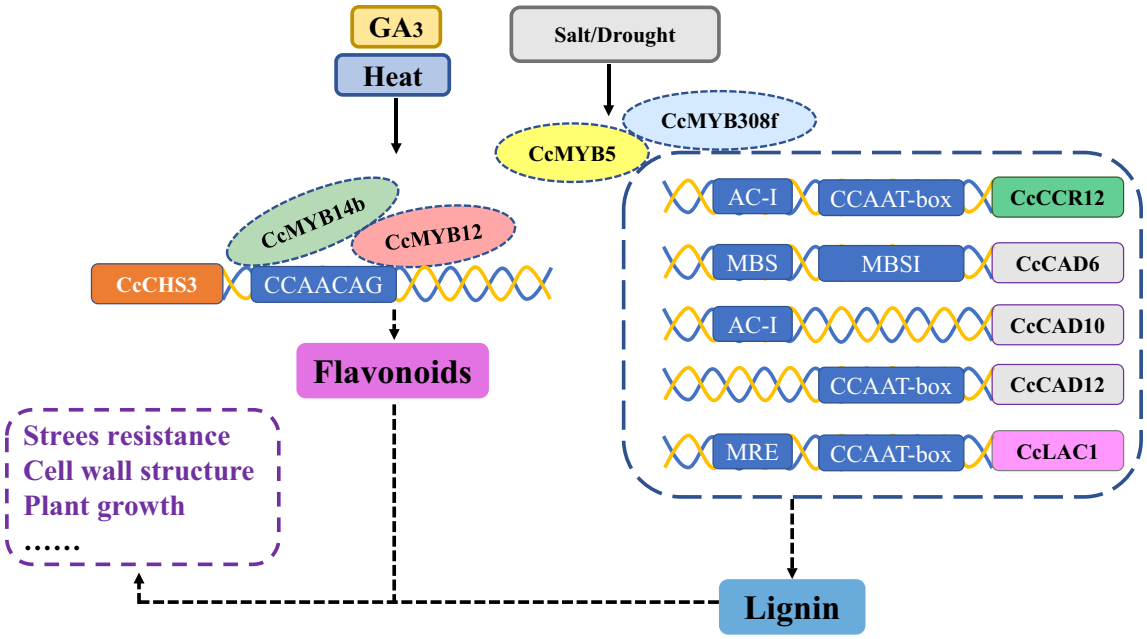

Fig. 12 Hypothetical regulatory pattern of the representative R2R3-MYB transcription factors in pigeon pea. AC-I, MBS, MBSI and MRE represent MYB-binding site

salt stress. This may be related to the number of defense and stress-responsive elements. In addition, *CcMYB* and key enzyme genes from the lignin pathway are significantly induced by salt and drought stress (Fig. 12). It is consistent with the previous reports that salt stress-induced *MYB* genes in wheat, sunflower, *Arabidopsis*, and rice (Li et al. 2020a; Nagaoka and Takano 2003; Hwang et al. 2001; Yu et al. 2017), indicating that the accumulation of lignin can improve drought resistance of plants. Furthermore, the promoter region of key genes from the lignin pathway contains a large number of MYB binding elements which are related to drought and temperature (Supplementary Fig. S5). It is implied that the significant response to drought induction of these genes is closely related to the MYB binding elements on the promoter sequences (Fig. 12). Collectively, these results indicated there may be many functional diversifications of *CcMYBs* and related genes from the phenylpropanoid pathway in pigeon pea. These representative *CcMYBs* and the key genes from the phenylpropanoid pathway might have ubiquitous functions to resist abiotic stress.

Conclusions

In conclusion, 30 *CcMYB* genes were identified to involve in flavonoid and lignin biosynthesis in pigeon pea genome via genome-wide screening. A comprehensive analysis of the intron–exon organization, phylogenetic relationships, distribution on chromosomes, gene duplication, conserved motifs, and expression levels under abiotic stress conditions was performed. RNA-seq analysis revealed ten *CcMYB* and key enzyme genes from the phenylpropane pathway significantly respond to the induction of MeJA and have a co-expression trend. The specific expression patterns of examined genes are closely related to the accumulation of flavonoids in six different tissues of pigeon pea. In addition, these candidate genes were significantly up-regulated under different abiotic stress conditions. Our study provides basic research on flavonoid and lignin biosynthesis in pigeon pea. Moreover, the R2R3-MYB family and the key enzyme genes from the phenylpropanoid pathway in pigeon pea were identified and analyzed, they would support a valuable reference for similar studies in other plant species.

Author contribution statement YJF and SZ conceived and designed research. JY conducted experiments. HQL, LTW and YL contributed new reagents or analytical tools. LLN, QY and DM analyzed data. JY and SZ wrote the manuscript. All authors read and approved the manuscript.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00425-021-03713-1.

Acknowledgements This work was supported by State Key Program of National Natural Science Foundation of China (31930076), the Fundamental Research Funds for the Central Universities (2572020AW03), the 111 Project (B20088), and Heilongjiang Touyan Innovation Team Program (Tree Genetics and Breeding Innovation Team).

Data availability statements The datasets generated during and analysed during the current study are available from the corresponding author on reasonable request.

References

Ambawat S, Sharma P, Yadav NR, Yadav RC (2013) MYB transcription factor genes as regulators for plant responses: an overview. Physiol Mol Biol Plants 19(3):307–321

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37:W202–W208

Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C (2000) Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. Plant Cell 12(12):2383–2393

Brown J, Pirrung M, McCue LA (2017) FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics 33(19):3137–3139

Brueffer C, Saal LH (2016) TopHat-Recondition: a post-processor for TopHat unmapped reads. BMC Bioinform 17(1):1–3

Cao S, Zheng Y, Wang K, Rui H, Tang S (2010) Effect of methyl jasmonate on cell wall modification of loquat fruit in relation to chilling injury after harvest. Food Chem 118(3):641–647

Chen J, Wang J, Wang R, Xian B, Ren C, Liu Q, Wu Q, Pei J (2020b) Integrated metabolomics and transcriptome analysis on flavonoid biosynthesis in safflower (*Carthamus tinctorius* L.) under MeJA treatment. BMC Plant Biol 20(1):1–12

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020a) TBtools-an integrative toolkit developed for interactive analyses of big biological data. bioRxiv:289660

Cui Q, Peng X, Yao XH, Wei ZF, Luo M, Wang W, Zhao CJ, Fu YJ, Zu YG (2015) Deep eutectic solvent-based microwave-assisted extraction of genistin, genistein and apigenin from pigeon pea roots. Sep Purif Technol 150:63–72

Du H, Yang SS, Liang Z, Feng BR, Liu L, Huang YB, Tang YX (2012) Genome-wide analysis of the MYB transcription factor superfamily in soybean. BMC Plant Biol 12(1):106

Duan M-H, Luo M, Zhao CJ, Wang W, Zu YG, Zhang D-Y, Yao XH, Fu YJ (2013) Ionic liquid-based negative pressure cavitation-assisted extraction of three main flavonoids from the pigeonpea roots and its pilot-scale application. Sep Purif Technol 107:26–36

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) MYB transcription factors in Arabidopsis. Trends Plant Sci 15(10):573–581

Feng J, Dong Y, Liu W, He Q, Daud M, Chen J, Zhu S (2017) Genome-wide identification of membrane-bound fatty acid desaturase genes in *Gossypium hirsutum* and their expressions during abiotic stress. Sci Rep 7(1):1–12

Fornalé S, Shi X, Chai C, Encina A, Irar S, Capellades M, Fuguet E, Torres JL, Rovira P, Puigdomènech P (2010) ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux. Plant J 64(4):633–644

Fraser CM, Chapple C (2011) The phenylpropanoid pathway in Arabidopsis. Arabidopsis Book/am Soc Plant Biol 9:e0152

Fu Y, Zu Y, Liu W, Efferth T, Zhang N, Liu X, Kong Y (2006) Optimization of luteolin separation from pigeonpea [*Cajanus cajan* (L.) Millsp.] leaves by macroporous resins. J Chromatogr A 1137(2):145–152

Fu Y-J, Liu W, Zu YG, Tong M-H, Li SM, Yan MM, Efferth T, Luo H (2008) Enzyme assisted extraction of luteolin and apigenin from pigeonpea [*Cajanus cajan* (L.) Millsp.] leaves. Food Chem 111(2):508–512

Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook. Springer, pp 571–607

Goicoechea M, Lacombe E, Legay S, Mihaljevic S, Rech P, Jauneau A, Lapierre C, Pollet B, Verhaegen D, Chaubet-Gigot N (2005) EgMYB2, a new transcriptional activator from *Eucalyptus xylem*, regulates secondary cell wall formation and lignin biosynthesis. Plant J 43(4):553–567

Höll J, Vannozzi A, Czemmel S, D'Onofrio C, Walker AR, Rausch T, Lucchin M, Boss PK, Dry IB, Bogs J (2013) The R2R3-MYB transcription factors MYB14 and MYB15 regulate stilbene biosynthesis in *Vitis vinifera*. Plant Cell 25(10):4135–4149

Hwang MG, Chung IK, Kang BG, Cho MH (2001) Sequence-specific binding property of *Arabidopsis thaliana* telomeric DNA binding protein 1 (AtTBP1). FEBS Lett 503(1):35–40

Jin H, Martin C (1999) Multifunctionality and diversity within the plant MYB-gene family. Plant Mol Biol 41(5):577–585

Kang X, Kirui A, Widanage MCD, Mentink-Vigier F, Cosgrove DJ, Wang T (2019) Lignin-polysaccharide interactions in plant secondary cell walls revealed by solid-state NMR. Nat Commun 10(1):1–9

Karamysheva ZN, Surovtseva YV, Vespa L, Shakirov EV, Shippen DE (2004) A C-terminal Myb extension domain defines a novel family of double-strand telomeric DNA-binding proteins in Arabidopsis. J Biol Chem 279(46):47799–47807

Koes RE, Quattrocchio F, Mol JN (1994) The flavonoid biosynthetic pathway in plants: function and evolution. BioEssays 16(2):123–132

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870–1874

Legay S, Lacombe E, Goicoechea M, Briere C, Séguin A, Mackay J, Grima-Pettenati J (2007) Molecular characterization of EgMYB1, a putative transcriptional repressor of the lignin biosynthetic pathway. Plant Sci 173(5):542–549

Lepiniec L, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, Nesi N, Caboche M (2006) Genetics and biochemistry of seed flavonoids. Annu Rev Plant Biol 57:405–430

Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res 30(1):325–327

Letunic I, Bork P (2007) Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23(1):127–128

Li J, Liu H, Yang C, Wang J, Yan G, Si P, Bai Q, Lu Z, Zhou W, Xu L (2020a) Genome-wide identification of MYB genes and expression analysis under different biotic and abiotic stresses in *Helianthus annuus* L. Ind Crops Prod 143:111924

Li Y, Chen X, Wang J, Zou G, Wang L, Li X (2020b) Two responses to MeJA induction of R2R3-MYB transcription factors regulate flavonoid accumulation in *Glycyrrhiza uralensis* Fisch. PLoS ONE 15(7):e0236565

Liu W, Kong Y, Zu Y, Fu Y, Luo M, Zhang L, Li J (2010) Determination and quantification of active phenolic compounds in pigeon pea leaves and its medicinal product using liquid chromatography–tandem mass spectrometry. J Chromatogr A 1217(28):4723–4731

Liu J, Osbourn A, Ma P (2015) MYB transcription factors as regulators of phenylpropanoid metabolism in plants. Mol Plant 8(5):689–708

Liu C, Xie T, Chen C, Luan A, Long J, Li C, Ding Y, He Y (2017) Genome-wide organization and expression profiling of the R2R3-MYB transcription factor family in pineapple (*Ananas comosus*). BMC Genomics 18(1):1–16

Luo J, Butelli E, Hill L, Parr A, Niggeweg R, Bailey P, Weisshaar B, Martin C (2008) AtMYB12 regulates caffeoyl quinic acid and

flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol. Plant J 56(2):316–326
Ma D, Constabel CP (2019) MYB repressors as regulators of phenylpropanoid metabolism in plants. Trends Plant Sci 24(3):275–289
Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. Embnet Journal 17(1):10–12
Matus JT, Aquea F, Arce-Johnson P (2008) Analysis of the grape MYB R2R3 subfamily reveals expanded wine quality-related clades and conserved gene structure organization across Vitis and Arabidopsis genomes. BMC Plant Biol 8(1):1–15
McCarthy RL, Zhong R, Ye ZH (2009) MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in Arabidopsis. Plant Cell Physiol 50(11):1950–1964
McCarthy RL, Zhong R, Fowler S, Lyskowski D, Piyasena H, Carleton K, Spicer C, Ye ZH (2010) The poplar MYB transcription factors, PtrMYB3 and PtrMYB20, are involved in the regulation of secondary wall biosynthesis. Plant Cell Physiol 51(6):1084–1090
Mehrtens F, Kranz H, Bednarek P, Weisshaar B (2005) The Arabidopsis transcription factor MYB12 is a flavonol-specific regulator of phenylpropanoid biosynthesis. Plant Physiol 138(2):1083–1096
Meng D, Yang Q, Dong B, Song Z, Niu L, Wang L, Cao H, Li H, Fu Y (2019) Development of an efficient root transgenic system for pigeon pea and its application to other important economically plants. Plant Biotechnol J 17(9):1804–1813
Misra P, Pandey A, Tiwari M, Chandrashekar K, Sidhu OP, Asif MH, Chakrabarty D, Singh PK, Trivedi PK, Nath P (2010) Modulation of transcriptome and metabolome of tobacco by Arabidopsis transcription factor, *AtMYB12*, leads to insect resistance. Plant Physiol 152(4):2258–2268
Nagaoka S, Takano T (2003) Salt tolerance-related protein STO binds to a Myb transcription factor homologue and confers salt tolerance in Arabidopsis. J Exp Bot 54(391):2231–2237
Nix A, Paull CA, Colgrave M (2015) The flavonoid profile of pigeonpea *Cajanus cajan*: a review. Springerplus 4(1):125
Ogata K, Hojo H, Aimoto S, Nakai T, Nakamura H, Sarai A, Ishii S, Nishimura Y (1992) Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. Proc Natl Acad Sci 89(14):6428–6432
Omer S, Kumar S, Khan BM (2013) Over-expression of a subgroup 4 R2R3 type MYB transcription factor gene from *Leucaena leucocephala* reduces lignin content in transgenic tobacco. Plant Cell Rep 32(1):161–171
Pandey A, Misra P, Chandrashekar K, Trivedi PK (2012) Development of AtMYB12-expressing transgenic tobacco callus culture for production of rutin with biopesticidal potential. Plant Cell Rep 31(10):1867–1876
Pandey A, Misra P, Khan MP, Swarnkar G, Tewari MC, Bhambhani S, Trivedi R, Chattopadhyay N, Trivedi PK (2014) Co-expression of Arabidopsis transcription factor, At MYB 12, and soybean isoflavone synthase, Gm IFS 1, genes in tobacco leads to enhanced biosynthesis of isoflavones and flavonols resulting in osteoprotective activity. Plant Biotechnol J 12(1):69–80
Paz-Ares J, Ghosal D, Wienand U, Peterson P, Saedler H (1987) The regulatory c1 locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. EMBO J 6(12):3553–3558
Preston J, Wheeler J, Heazlewood J, Li SF, Parish RW (2004) AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*. Plant J 40(6):979–995
Rinthong P-o, Maneechai S (2018) Total phenolic content and tyrosinase inhibitory potential of extracts from *Cajanus cajan* (L.) Millsp. Pharmacogn J 10(6s):109–112
Rio DC, Ares M, Hannon GJ, Nilsen TW (2010) Purification of RNA using TRIzol (TRI reagent). Cold Spring Harbor Protoc 210(6):pdb.prot5439
Shen H, He X, Poovaiah CR, Wuddineh WA, Ma J, Mann DG, Wang H, Jackson L, Tang Y, Neal Stewart Jr C (2012) Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. New Phytol 193(1):121–136
Sonbol F-M, Fornalé S, Capellades M, Encina A, Tourino S, Torres JL, Rovira P, Ruel K, Puigdomenech P, Rigau J (2009) The maize ZmMYB42 represses the phenylpropanoid pathway and affects the cell wall structure, composition and degradability in *Arabidopsis thaliana*. Plant Mol Biol 70(3):283
Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B (2007) Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. Plant J 50(4):660–677
Sun W, Ma Z, Chen H, Liu M (2019) MYB gene family in potato (*Solanum tuberosum* L.): genome-wide identification of hormone-responsive reveals their potential functions in growth and development. Int J Mol Sci 20(19):4847
Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7(3):562–578
Voelker SL, Lachenbruch B, Meinzer FC, Strauss SH (2011) Reduced wood stiffness and strength, and altered stem form, in young antisense 4CL transgenic poplars with reduced lignin contents. New Phytol 189(4):1096–1109
Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40(7):e49–e49
Wei Q, Chen R, Wei X, Liu Y, Zhao S, Yin X, Xie T (2020) Genome-wide identification of R2R3-MYB family in wheat and functional characteristics of the abiotic stress responsive gene TaMYB344. BMC Genomics 21(1):1–16
Xiao Y, Zhou L, Lei X, Cao H, Wang Y, Dou Y, Tang W, Xia W (2017) Genome-wide identification of WRKY genes and their expression profiles under different abiotic stresses in *Elaeis guineensis*. PLoS ONE 12(12):e0189224
Xu G, Guo C, Shan H, Kong H (2012) Divergence of duplicate genes in exon-intron structure. Proc Natl Acad Sci 109(4):1187–1192
Yu Y, Ni Z, Chen Q, Qu Y (2017) The wheat salinity-induced R2R3-MYB transcription factor TaSIM confers salt stress tolerance in *Arabidopsis thaliana*. Biochem Biophys Res Commun 491(3):642–648
Zhang DY, Zu YG, Fu YJ, Luo M, Wang W, Gu CB, Zhao CJ, Jiao J, Efferth T (2012) Enzyme pretreatment and negative pressure cavitation extraction of genistein and apigenin from the roots of pigeon pea [*Cajanus cajan* (L.) Millsp.] and the evaluation of antioxidant activity. Ind Crops Prod 37(1):311–320
Zhang C, Freddolino PL, Zhang Y (2017) COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. Nucleic Acids Res 45(W1):W291–W299
Zhou J, Lee C, Zhong R, Ye ZH (2009) MYB58 and MYB63 are transcriptional activators of the lignin biosynthetic pathway during secondary cell wall formation in *Arabidopsis*. Plant Cell 21(1):248–266
Zhu L, Shan H, Chen S, Jiang J, Gu C, Zhou G, Chen Y, Song A, Chen F (2013) The heterologous expression of the chrysanthemum R2R3-MYB transcription factor CmMYB1 alters lignin composition and represses flavonoid synthesis in *Arabidopsis thaliana*. PLoS ONE 8(6):e65680
Zuluaga DL, Gonzali S, Loreti E, Pucciariello C, Degl'Innocenti E, Guidi L, Alpi A, Perata P (2008) *Arabidopsis thaliana* MYB75/

PAP1 transcription factor induces anthocyanin production in transgenic tomato plants. Funct Plant Biol 35(7):606–618