

Genome-wide identification of R2R3-MYB gene family and association with anthocyanin biosynthesis in Brassica species

Daozong Chen

College of Life Sciences, Ganzhou Key Laboratory of Greenhouse Vegetable, Gannan Normal University

Haidong Chen

College of Life Sciences, Ganzhou Key Laboratory of Greenhouse Vegetable, Gannan Normal University

Guoqiang Dai

College of Life Sciences, Ganzhou Key Laboratory of Greenhouse Vegetable, Gannan Normal University

Haimei Zhang

College of Life Sciences, Ganzhou Key Laboratory of Greenhouse Vegetable, Gannan Normal University

Yi Liu

College of Life Sciences, Ganzhou Key Laboratory of Greenhouse Vegetable, Gannan Normal University

Wenjie Shen

College of Life Sciences, Ganzhou Key Laboratory of Greenhouse Vegetable, Gannan Normal University

Bo Zhu

College of Life Sciences, Ganzhou Key Laboratory of Greenhouse Vegetable, Gannan Normal University

Cheng Cui

Crop Research Institute, Sichuan Academy of Agricultural Sciences

Chen Tan (✉ tanchen2020@gnnu.edu.cn)

College of Life Sciences, Ganzhou Key Laboratory of Greenhouse Vegetable, Gannan Normal University

Research Article

Keywords: Brassica species, R2R3-MYB family, subgenome, anthocyanin biosynthesis, co-expression

Posted Date: March 24th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1454390/v1>

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

[Read Full License](#)

Abstract

Brassica species include important oil crops and vegetables in the world. The R2R3-MYB gene participates in a variety of plant functions, including the activation or inhibition of anthocyanin biosynthesis. Although previous studies have reported its phylogenetic relationships, gene structures, and expression patterns in *Arabidopsis*, the number and sequence variation of this gene family in *Brassica* crops and its involvement in the natural quantitative variation in anthocyanin biosynthesis regulation are still largely unknown. In this study, by using whole genome sequences and comprehensive genome-wide comparative analysis among the six cultivated *Brassica* species, 1990 R2R3-MYB genes were identified in six *Brassica* species, respectively. These R2R3-MYB genes were phylogenetically clustered into 12 groups. The R2R3-MYB family between A and C subgenomes showed better collinearity than between B and C and between A and B. From comparing transcriptional changes of five *Brassica* species with the purple and green leaves for the detection of the R2R3-MYB genes associated with anthocyanin biosynthesis, 7 R2R3-MYB genes were co-differentially expressed. The promoter and structure analysis of these genes showed that some variations between non-coding region, but they were highly conserved at the protein level and spatial structure. Co-expression analysis of anthocyanin-related genes and R2R3-MYBs indicated that *MYB90* was strongly co-expressed with *TT8*, and they were co-expressed with structural genes *F3H*, *LDOX*, *ANS* and *UF3GT* at the same time. These results further clarified the roles of the R2R3-MYBs for leaf coloration in *Brassica* species, which provided new insights into the functions of the R2R3-MYB gene family in *Brassica* species.

Introduction

The roles of transcription factors (TFs) in the regulation of transcript levels of structural genes controls many crucial biological processes (Allan et al. 2008). The TFs usually recognize target DNA in a sequence-specific manner and regulate the frequency of initiation of transcription upon binding to specific sites/motifs in the promoter of target genes. Generally, TFs can work as activators, repressors, or both to regulate the expression of target genes. MYB transcription factors have since been found in all plants and comprised one of the largest of transcription factors family (Lipsick, 1996). MYB transcription factor contains a conserved DBD (DNA-binding domain), which is generally composed of 1–4 imperfect repeats, named R repeat (Lipsick, 1996; Dubos et al., 2010). Each R repeat covers 50–55 amino acids in length that fold into three α -helices, when bound to target gene specific promoter sequences, the second and third helices were form a helix-turn-helix (HTH) structure (Ogata et al., 1992; Lipsick, 1996), and the third α -helix usually play a recognition role in binding to a short DNA sequence (Stracke et al., 2001).

For decades, it has been identified and reported in many plants that the R2R3-MYB transcription factor is widely involved in plant development and metabolic regulation (Stracke et al., 2001; Du et al., 2012; Wang et al., 2019; Wang et al., 2020; Liu et al., 2021). Recently, the increasing availability of plant genome sequences have facilitated a better understanding of this large gene family (Wang et al., 2019; Wang et al., 2020; Liu et al., 2021). The first plant R3R3-MYB gene *COLORED1* (*C1*) was isolated from maize (*Zea mays*) which encode a regulatory protein involved in anthocyanin biosynthesis (Paz-Ares et al., 1987).

Subsequently, the R2R3-MYB-mediated anthocyanin synthesis pathway was revealed in the model plant *Arabidopsis*, *PRODUCTION OF ANTHOCYANIN PIGMENT 1/2* (*PAP1/2*, *MYB75/90*) are reported to be involved in the transcriptional regulation of anthocyanins in vegetative tissues, while *TRANSPARENT TESTA 2* (*TT2*, *MYB123*) is involved in the biosynthesis and accumulation of anthocyanins in seed coats (Gonzalez et al., 2008; Xu et al., 2014). Nevertheless, compared to *Arabidopsis*, relatively few members of the *Brassica* species R2R3-MYB gene family have been well identified and functional characterized. Moreover, the R2R3-MYBs characterized in *Brassica* species are limited in the transcriptional regulatory mechanism of anthocyanin metabolism pathway to date.

R2R3-MYB transcription factors have been shown to be involved in phenylpropanoid metabolism (Hichri et al., 2011; Shi & Xie, 2014) and activate the structural genes in the anthocyanin biosynthetic pathway in many plants (Albert et al. 2014; Xie et al. 2014; Zhang et al. 2017). In *Brassica* species, the R2R3-MYB transcription factors play a key role in anthocyanin biosynthesis pathway. In *B. oleracea*, *BoMYB2* with different types of mutations in the promoter region determines the different color types (Yan et al., 2019). Similarly, the activated expression of the *BrMYB2* gene in purple-leaf *B. rapa* and the *BjPur* gene in purple-leaf *B. juncea* is due to the large deletion of the first intron (He et al., 2020; Heng et al., 2020). In *B. napus*, *BnaPAP2.A7* has been found to work as a key transcription factor regulating the synthesis of anthocyanins in the leaves (Chen et al., 2020). However, both *BoMYB2*, *BrMYB2*, *BjPur* and *BnaPAP2.A7* are homologous of *Arabidopsis AtPAP2*, while other R2R3-MYB transcription factors regulate the synthesis of anthocyanins in *Brassica* crops are rarely need to be explored.

In *Brassica*, three genomes (designated A, B and C) share mesohexaploid ancestry and occur both singly and in each pairwise combination to define the *Brassica* species, which is the excellent model for studying the evolution of gene families. In this study, we performed genome-wide identification of MYB transcription factors of the six cultivated *Brassica* species. R2R3-MYB and R1R2R3-MYB subfamilies were identified and analyzed, including chromosomal distribution, synteny and evolutionary relationships. We also analyzed the spatial and temporal expression profiles as well as differential expression profiles of R2R3-MYB genes in five *Brassica* species (except *B. nigra*) with green and purple leaves, and then the gene structure, conserved domains and original promoters. These findings will provide a comprehensive understanding of novel R2R3-MYB genes involved in anthocyanin biosynthesis and assist further exploration of the functional characteristics of R2R3-MYB transcription factors in *Brassica* species.

Materials And Methods

Identification of R2R3-MYB transcription factors in Brassica

In this study, the genome and protein sequences of the *B. rapa* (Chiifu), *B. oleracea* (TO1000), *B. nigra* (Ni100), *B. napus* (ZS11) were downloaded from the BRAD database (Chen et al., 2021), *B. juncea* (SY) genome sequence from NCBI PRJNA615316 (Kang et al., 2021), *B. carinata* (zd-1) genome sequence from GenBank JAAMPC000000000 (Song et al., 2021), and the corresponding R2R3-MYB genome and protein sequences from the *Arabidopsis* database (TAIR; <http://www.Arabidopsis.org/>).

The candidate R2R3-MYB members in the six *Brassica* species were identified by a local BLASTP search with 125 R2R3-MYB genes from *Arabidopsis* to identify candidates with E-value < 1e-10, then all putative R2R3-MYB genes were identified in the PFAM protein family database using the HMMER software version 3.0 (Eddy, 2011) and using the hidden Markov model of the MYB-DNA-binding domain (PF00249) and Myb_DNA-bind_6 domain (PF13921) to search against the six *Brassica* species genome to identify candidates with E-value < 1e-10 (Finn et al., 2013). All protein sequences were further investigated using different online tools, including the motif search (<https://www.genome.jp/tools/motif/>), SMART (<http://smart.embl-heidelberg.de/>), ScanProsite (<https://prosite.expasy.org/scanprosite/>), NCBI-Conserved Domains Database (CDD) web server (<https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>), MEME Suite (<https://meme-suite.org/meme/doc/meme-format.html>) (Wang et al., 2020; Liu et al., 2021).

Phylogenetic analysis

The R2R3-MYB protein sequences of the six *Brassica* species and *Arabidopsis* were used to generate phylogenetic trees via ClustalX (Larkin et al., 2007) and MAFFT software (Kato and Standley, 2013) multiple sequence alignments with the default parameters. A maximum likelihood (ML) phylogenetic tree was constructed using FastTree2 software (v2.1.11), in which JTT (Jones-Taylor-Thornton) model was the best substitution model (Price et al., 2010). Additionally, itol (<https://itol.embl.de/>) and Adobe Illustrator CS6 software (v16.0.0) were used to modify the evolutionary tree.

Chromosomal location and synteny analysis

The genome annotation data were collected and mapped on the chromosomes using the TBtools software (v0.67) to identify the physical chromosomal location of all R2R3-MYB genes in six *Brassica* crops and *Arabidopsis* (Chen et al., 2020a). The collinearity of intraspecific and interspecific genes was determined using the BLASTP (E-value: 1e-10, max_target_seqs:1) and Multiple Collinearity Scan toolkit (MCSscanX, gap_penalty: -1, E-value: 1e-10) (Wang et al., 2012), TBtools software (v0.67) was used to drop the collinearity genes on each chromosome (Chen et al., 2020).

Expression profiles analysis based on RNA-seq data

The raw data of RNA-seq were downloaded from NCBI SRA database with item number PRJNA298501, PRJNA430791, PRJNA359160, PRJNA474411, PRJNA558197, PRJNA312129 and PRJNA560282. Low-quality reads were removed from the raw reads using Cutadapt and Trimmomatic software to get clean reads (Martin, 2011; Bolger et al., 2014). Clean reads were mapped to the corresponding reference genome using HISAT2 software (Pertea et al., 2016). Gene expression levels of each gene were calculated using StringTie and Ballgown software (Pertea et al., 2016). The read counts of each gene were calculated using the htseq-count function in htseq software (Anders et al., 2015). The R package DESeq2 (v1.16.1) was used to identify the differentially expressed genes (DEGs) between leaves of different colors based on the following criteria: padj < 0.01 & log2FoldChange > 2 (Chen et al., 2020b).

Analysis of the promoter characteristics, gene structure, conserved motif and co-different expression of R2R3-MYBs

The co-different expression R2R3-MYB promoter regions of 2000 bp regions upstream of the translational start sites ATG were examined based on their positions in the genomes of six *Brassica* species and *Arabidopsis* using Samtools software (v 1.8), which was used to identify the cis-elements in the promoters according to the online PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>). The gene structures of all co-different expression R2R3-MYB genes were analyzed according to the GFF annotation file of the gene position information in the six *Brassica* plants and *Arabidopsis* database. TBtools software (v0.67) was used to locate the co-different expression R2R3-MYB to the different chromosomes of each species, including exon and intron numbers and lengths (Chen et al., 2020). The MEME online tool (<https://meme-suite.org/meme/>) was used to investigate conserved domains, and the WEBLogo online tool (<https://weblogo.berkeley.edu/>) and AlphaFold Protein Structure Database (<https://www.alphafold.ebi.ac.uk/>) were used to analyze the R2 and R3 motif conserved sequence and SWISS-MODEL online tool (<https://swissmodel.expasy.org/>) was used to draw spatial structure. The RNA-seq data were used to perform co-expression network analysis using R language (v3.6.1). In order to calculate the adjacent order function formed by the gene network and the difference coefficients of different nodes, the TOM similarity algorithm calculates the co-expression correlation matrix to express the gene correlation in the network. The correlation network diagram was drawn by extracting the non-weight coefficients (weight) of R2R3-MYB and anthocyanin-related genes in the matrix. STRING software (<https://version-11-5.string-db.org/>) was used to reveal a co-expression plot (Wang et al., 2020; Liu et al., 2021).

Plant materials, RNA extraction, and qRT-PCR analysis

The *B. rapa*, *B. oleracea*, *B. napus*, *B. juncea* and *B. carinata* green and purple leaves of seven-leaf stage with three biological replicates were collected, all samples were collected and immediately frozen in liquid nitrogen for RNA extraction and qRT-PCR analysis. For detailed methods, please refer to Chen et al. (2020b). All materials of this research were collected from Gannan Normal University research base.

Results

Identification and characterization of R2R3-MYB family genes

The 125 R2R3-MYB genes protein sequences of *Arabidopsis* were downloaded and used as seed sequences to search the protein sequences database of six *Brassica* plants and *Arabidopsis* to identify homologous R2R3-MYB genes in six *Brassica* plants. Firstly, the candidate R2R3-type MYB members in the six *Brassica* plants were identified by a local BLASTP search with *Arabidopsis* to identify candidates with E-value < 1e-10. Secondly, Hidden Markov model (HMM) profile of the MYB domain Myb-DNA-

binding domain (PF00249) and Myb_DNA-bind_6 domain (PF13921) queried the hmmsearch program (HMMER3.0 package) against the protein database of six *Brassica* plants and *Arabidopsis*. Subsequently, all protein sequences were further investigated using hmmscan, SMART, ScanProsite, NCBI-CDD search and MEME Suite. A total of 2472 MYB genes were identified, these included 297 1R-MYB genes, 2120 R2R3-MYB genes, 33 R1R2R3-MYB genes, and 22 R1R2R3R4-MYB genes (Table 1, Supplementary Table S1). Among them, the R2R3-MYBs were the largest MYB subgroup containing 2120 genes, including 130 in *Arabidopsis*, 236 in *B. rapa*, 247 in *B. oleracea*, 248 in *B. nigra*, 425 in *B. napus*, 422 in *B. juncea* and 412 in *B. carinata* (Table 1).

Table 1
R2R3_MYB protein distribution in six Brassica crops

species	R2R3_MYB	3R_MYB	>=4R_MYB	R1_MYB	Total
<i>A.thaliana</i>	130	2	1	23	156
<i>B.carinata</i>	412	7	3	48	470
<i>B.juncea</i>	422	5	5	57	489
<i>B.napus</i>	425	9	4	63	501
<i>B.nigra</i>	248	2	3	36	289
<i>B.oleracea</i>	247	4	4	41	296
<i>B.rapa</i>	236	4	2	29	271
total	2120	33	22	297	2472

Phylogenetic analysis of R2R3-MYB genes in six Brassica species

To analyze the phylogenetic relationships and gene functions of the R2R3-MYB gene family members, a ML tree containing 2120 R2R3-MYB genes was constructed using mafft software for multiple sequence alignment of protein sequences, and using fasttree software a clustering tree was built. The 2120 R2R3-MYB genes could be divided into 12 subfamilies (T1-T12) and were drawn in different colors (Fig. 1). Compared with *Arabidopsis* 25 subfamilies, our T1 was the S21 subfamily, T2 contained the S22 and S23 subfamilies, T3 was the S25 subfamily, T5 contained the S6 and S15 subfamilies, T6 contained the S8, S18, S19 and S20 subfamilies, T7 contained the S13 and S16 subfamilies, T8 contained the S1, S2, S3, S5 and S12 subfamilies, T9 contained the S9, S10, S11 and S24 subfamilies, T10 was the S14 subfamily, T11 included the S4 and S7 subfamilies, T12 was the S17 subfamily (Supplementary Table S2). T4 subfamily contained one R2R3-MYB gene in *Arabidopsis* and 18 homologous in *Brassica* plants, which was separated into one subfamily (Supplementary Table S3). Previous research has shown that the group of R2R3-MYB genes in the same subclade might have a similar function. The R2R3-MYB gene

functions in the T5, T8 and T11 subgroups were known to be involved in the phenylalanine metabolism pathway, including the regulation of anthocyanin and procyanidin synthesis (Dubos et al., 2010). For example, in T11 subgroup, *AT1G22640* (*AtMYB3*), *AT4G38620* (*AtMYB4*), *AT2G16720* (*AtMYB7*), and *AT4G34990* (*AtMYB32*) inhibited phenylalanine biosynthesis; *AT2G47460* (*AtMYB11*), *AT3G62610* (*AtMYB12*) and *AT5G49330* (*AtMYB111*) had the ability to participate the PA biosynthesis. The T11 subfamily contained *AT1G56650* (*AtMYB75*), *AT1G66390* (*AtMYB90*), *AT1G66370* (*AtMYB113*), and *AT1G66380* (*AtMYB114*) that regulated anthocyanin biosynthesis in vegetative tissues.

Chromosomal distribution and collinearity analysis of R2R3-MYBs in six Brassica species

All 2120 R2R3-MYB genes were mapped onto the corresponding genomic chromosomes and were found to be located onto all chromosomes of each species, i.e., 130 genes in *Arabidopsis* to 5 chromosomes, 236 genes in *B. rapa* to 10 chromosomes, 247 genes in *B. oleracea* to 9 chromosomes, 248 genes in *B. nigra* to 8 chromosomes, 425 genes in *B. napus* to 19 chromosomes, 422 genes in *B. juncea* to 18 chromosomes and 412 genes in *B. carinata* to 17 chromosomes (Supplementary Fig. 1). The results showed that the R2R3-MYB gene was relatively evenly distributed in each subgenome of the six *Brassica* crops, and only relatively few in the B subgenome of *B. carinata*, with only 166 members (Table 2). Among the three subgenomes of A, B, and C, the A subgenome has the most distribution (Supplementary Fig. 1).

Table 2
Distribution of R2R3_MYB gene family in each subfamily

species	AT	B.ra	B.ol	B.ni	B.ju	B.na	B.ca	Total
T1	8	16	17	18	31	35	29	154
T2	8	14	16	15	28	32	24	137
T3	12	21	24	21	39	37	32	186
T4	1	2	2	2	4	4	4	19
T5	8	10	10	13	21	18	21	101
T6	23	40	39	39	70	76	67	354
T7	12	22	21	23	38	37	34	187
T8	22	48	52	52	90	81	93	438
T9	13	27	27	27	38	43	43	218
T10	8	15	16	15	29	23	24	130
T11	9	12	12	14	21	21	25	114
T12	6	9	11	8	13	18	17	82
all	130	236	247	247	422	425	413	2120

The comparison of interspecific synteny among six *Brassica* species and *Arabidopsis* were also analyzed to further explore the evolution of R2R3-MYB genes (Fig. 2). At the genome-wide level, three A subgenomes from *B. rapa*_(BraA), *B. napus*_(BnaA), *B. juncea*_(BjuA) had a good collinearity relationship, almost all collinearity blocks were distributed on the corresponding homologous chromosomes of the same number, showing that the A subgenome had the less degree of gene differentiation and better preservation of gene integrity during the evolution process (Fig. 2). In three C subgenomes from *B. oleracea*_(BolC), *B. napus*_(BnaC) and *B. carinata*_(BcaC), BolC and BnaC shared a good collinearity relationship, and all collinearity blocks were distributed in the corresponding homologous chromosomes of the same number. However, the chromosomal correspondence between the subgenome BolC and BcaC was poor, which indicated that the C subgenome had a less degree of gene differentiation and better preservation of gene integrity during the evolution of *B. napus*, while the chromosomal level variation was higher in *B. carinata*. Similarly, in the three B subgenomes from *B. nigra*_(BniB), *B. juncea*_(BjuB) and *B. carinata*_(BcaB), BniB and BjuB gave a good collinearity relationship, and most of the collinearity blocks were distributed onto the corresponding same numbers on homologous chromosomes. However, the chromosomal correspondence between BniB and BcaB subgenome was poor, indicating that the B subgenome had a lower degree of gene differentiation and better preservation of gene integrity during the evolution of *B. juncea*, while the *B. carinata* had a larger chromosome level variation (Fig. 2).

Analysis of the transcriptome and co-differentially expressed R2R3-MYBs

The different tissues with green, white or purple phenotype of *B. rapa*, *B. oleracea*, *B. napus*, *B. juncea* and *B. carinata* RNA-seq data were used to determine expression levels of R2R3-MYBs. Compared to the genome-wide analysis of the R2R3-MYB gene family, 41 R2R3-MYBs were found to be differentially expressed in transcriptome data. All these differentially expressed R2R3-MYBs (DE R2R3-MYBs) exhibited differential expressions between green and purple leaves. There were 9, 6, 11, 8, 7 DE R2R3-MYBs in *B. rapa*, *B. oleracea*, *B. napus*, *B. juncea* and *B. carinata*, respectively (Supplementary Table S4). Among them, 7 genes were differentially expressed in common, which were all homologous copies of *Arabidopsis* AT1G56650 (MYB75), AT1G66370 (MYB113), AT1G66380 (MYB114) and AT1G66390 (MYB90) genes (subfamily 6 of R2R3-MYBs). In five *Brassica* crops, they were *BraA07g032100.3C*, *Bo3g081880*, *Bo6g100940*, *BnaA07T0287000ZS*, *BjuB05g10740S*, *BcaB05g24263*, *BcaB03g15272* (Supplementary Table S3). In *B. napus*, the DE R2R3-MYBs *BnaA07T0287000ZS* and *BnaC06T0329100ZS* were all up-regulated in the purple leaf. Interestingly, of two differentially expressed genes in *B. carinata*, *BcaB05g24263* was down-regulated in purple leaves, while *BcaB03g15272* was up-regulated (Fig. 3).

Gene structural analysis and conserved motif identification of DE R2R3-MYBs

Gene structural analysis was helpful for the better understanding of its function and evolution. The intron numbers of the 7 DE R2R3-MYBs and *Arabidopsis* AT1G56650 (MYB75), AT1G66370 (MYB113), AT1G66380 (MYB114) and AT1G66390 (MYB90) were all three, except for four introns of *Bo3g081880* (Supplementary Fig. 2). However, there were large variations in the length of exons and introns, especially for *BraA07g032100.3C* with a large insert in the first intron.

The 2,000 base pairs (bp) upstream of R2R3-MYB gene sequences of the coding region were used to predict *cis* regulatory elements via the PlantCARE online tool (Supplementary Fig. 2). A total of 20 *cis*-regulatory elements of DE R2R3-MYBs were predicted, including 3 cellular development related *cis*-regulatory elements: cell cycle regulation, meristem expression, maximal elicitor-mediated activation, 5 hormone related *cis*-regulatory elements: abscisic acid responsiveness, auxin responsive, MeJA-responsiveness, gibberellin responsiveness, salicylic acid responsiveness. Similarly, 4 stress related *cis*-elements were also identified, including light responsiveness, defense and stress responsiveness, low-temperature responsiveness, dehydration, low-temp, salt stress. The MBS and MRE were specifically MYB binding sites involved in flavonoid biosynthetic, light responsiveness and drought-inducibility. AT-rich DNA binding protein, core promoter element around -30, MYBHv1 binding, protein binding site and maximal elicitor-mediated activation. The numbers of *cis*-regulatory elements for each DE R2R3-MYBs varied, *Bo6g100940* had only one and *BcaB05g24263* had nine.

All DE R2R3-MYBs contained two highly conserved MYB binding domains. The motif logo of DE R2R3-MYBs had 51 amino acid residues in the R2 repeat and 45 amino acid residues in the R3 repeat, respectively (Supplementary Fig. 3). The HTH structure of these two domains as revealed by three-dimensional (3D) protein structural models showed that DE R2R3-MYBs matched the typical characteristics of the R2R3-MYB family (Supplementary Fig. 3).

Co-expression analysis of anthocyanin-related genes and R2R3-MYBs

The DEGs between the green, white or purple tissues of five *Brassica* species were used to predict candidate R2R3-MYBs related to anthocyanin biosynthesis (Supplementary Table S4). *MYB90* (representing *MYB75*, *MYB90*, *MYB113*, *MYB114* and 7 DE R2R3-MYBs in *Brassica* plants) had a strong co-expression with *TT8*, and they were co-expressed with structural genes *F3H*, *LDOX*, *ANS* and *UF3GT* at the same time (Fig. 4). Simultaneously, the anthocyanin biosynthesis repressor *CPC* and *MYBL2* were also co-expressed with *TT8*. These results showed that these 7 DE R2R3-MYBs were the important regulation genes involved in anthocyanin biosynthesis processes.

Expression pattern of R2R3-MYB DEGs and anthocyanin related genes in five Brassica species

In order to confirm the transcriptional pattern revealed by the RNA-seq analysis, we performed qRT-PCR analysis for 7 DE R2R3-MYBs (Fig. 5). Six genes showed a similar relative expression level in green / white and purple tissues as revealed by RNA-seq analysis. The *MYB90* was hardly expressed in the green / white tissues of five *Brassica* species, but significantly upregulated in the purple/red tissues, which was consistent with the transcriptome and phenotype, respectively. However, *BcaB03g15272* is not expressed in purple and green leaves, which is different from the results of the transcriptome.

Discussion

The R2R3-MYB gene family in Brassica species

The R2R3-MYB is the largest subfamily of the MYB transcription factor family. To date, several R2R3-MYB families have been identified and analyzed. In the model plant *A. thaliana*, 125 R2R3-MYB genes were found (Stracke et al., 2001), and then 126 R2R3-MYB genes detected from the whole genome MYB transcription factor identification and divided into 26 subfamilies (Dubos et al., 2010), which provides a reference for the identification and analysis of R2R3-MYB transcription factors in other plants. In recent years, numerous R2R3-MYB transcription factors have been identified and functionally analyzed, such as in *Oryza sativa* (Chen et al., 2006), *Zea mays* (Du et al., 2012), *B. napus* (Hajiebrahimi et al., 2017) and horticultural plants *Petunia* (Chen et al., 2021), Strawberry (Liu et al., 2021) and *Capsicum annuum* (Wang et al., 2020). Importantly, more and more biological pathways involved in R2R3-MYB have been

revealed, such as the transcriptional regulation of anthocyanins (Zhao et al., 2013), fiber development (Wang et al., 2020) and root hair development (Slabaugh et al., 2011).

Six *Brassica* species of three diploids and three derived allotetraploids in U's triangle are excellent models for studying gene families. With the completion of the sequencing of the *B. carinata* genome, the genomes of all six species have been published, providing a useful tool for genome-wide analysis of the R2R3-MYB gene family. To date, no full analysis of the R2R3-MYB gene family is reported in these six *Brassica* crops and most functions remain unclear. In this study, we identified 2120 R2R3-MYB genes from the genomes of *Arabidopsis* and six *Brassica* crops, including 130 in *Arabidopsis*, 236 in *B. rapa*, 247 in *B. oleracea*, 248 in *B. nigra*, 425 in *B. napus*, 422 in *B. juncea* and 412 in *B. carinata* (Table 1), indicating the expansion of this gene family along with genome duplication in *Brassica* crops. The present numbers of R2R3-MYBs in *Arabidopsis*, *B. rapa* and *B. napus* were different from those of previous studies, 126 in *Arabidopsis* (Stracke et al., 2001), 256 in the *B. rapa* (Wang et al., 2015), and 249 in the *B. napus* (Hajiebrahimi et al., 2017). The main reason was probably that with the development of three-generation sequencing, genome information became more complete.

Evolution of the R2R3-MYB gene family

Gene duplication is a major factor responsible for the expansion of gene families and the generation of new genes, which has also led to changes in the number and function of duplicated genes, and also provides an excellent example for studying the inheritance and evolution of gene family. The six *Brassica* crops of U's triangle provides an excellent model for gene duplication and evolution research, because of the natural cross between the two diploids and the independent evolution of the parents and the allopolyploidy hybrids. To understand the evolutionary relationship of R2R3-MYB family genes in six *Brassica* crops, we constructed the phylogenetic tree with R2R3-MYB family genes from the six *Brassica* crops and *Arabidopsis*. All 2120 R2R3-MYB family genes from *Arabidopsis* and six *Brassica* crops were classified into 12 different classes with 130 R2R3-MYB family genes from *Arabidopsis* (Fig. 1), each cluster corresponds to one or more R2R3-MYB subfamily of *Arabidopsis* (Supplementary Table S3), which was in accordance with that the R2R3-MYB family was relatively conservative between different plants in evolution. Chromosomal distributions indicated that they were relatively evenly distributed on the A, B, and C subgenomes, but the abundance of the A subgenome was higher than that of the B and C subgenomes (Supplementary Fig. 1). Collinearity analysis of R2R3-MYBs in six *Brassica* species showed that the collinearity between A and C subgenomes was better than that between B and A or C subgenomes (Fig. 2), which was consistent that their genomic relationships (Schranz et al., 2006). Therefore, the results offered a useful framework for future research to understand the evolution of the R2R3-MYB gene family.

The expression level of R2R3-MYBs related to anthocyanin biosynthesis

Brassica crops present color variations caused by the accumulation of anthocyanins in tissues or organs, such as leaves, stems, flowers, siliques and seeds, from *B. campestris* L. var. *purpuraria* L. H. Bailey, purple cabbage, purple cauliflower, purple *B. juncea*, purple rapeseed. In our study, we analyzed the transcriptional profiles of co-differentially expressed R2R3-MYBs for green and purple leaves in five *Brassica* crops (except *B. nigra*), and a totally of 41 R2R3-MYBs were differentially expressed (Table 2). Of these, seven co-DE R2R3-MYBs were found in five *Brassica* crops (Fig. 3), and were homologous genes of *MYB75/MYB90/MYB113/MYB114* in *Arabidopsis*, and were the subfamily 6 of R2R3-MYB (Stracke et al., 2001; Dubos et al., 2010). The co-expression analysis of anthocyanin-related genes and R2R3-MYBs found that *MYB90* have a strong interaction with *TT8* to regulate the structural genes *F3H*, *LDOX*, *ANS* and *UF3GT* at the same time (Fig. 4). Interestingly, the anthocyanin biosynthesis repressor *CPC* and *MYBL2* were also co-expressed with *TT8*. In this research, comprehensive transcriptome results, phenotypic data and qRT-PCR analysis showed that these 6 DE R2R3-MYBs are the important regulation genes involved in anthocyanin biosynthesis processes (Fig. 5).

Previous studies have indicated that R2R3-MYB genes play key roles in the anthocyanin biosynthesis in various plants, such as *TT2*, *MYB75*, *MYB90*, *MYB113* and *MYB114* in *Arabidopsis* (Device et al., 1999; Xu et al., 2014), *Ruby1* and *Ruby2* in citrus (Butelli et al., 2012), *AcMYB123* in kiwifruit (Wang et al., 2019), *PpMYB10.1* in peaches (Rahim et al., 2014), *MdMYB10* and *MdMYB110a* in Apple (Espley et al., 2007; Chagnè et al., 2013), and *VvMYBA1* and *VvMYBA2* in grapes (Walker et al., 2007). In *Brassica*, previous studies have shown that the sequence variation of the R2R3-MYB transcription factor in the promoter region or the gene intron region to drive its active expression is the main reason for the cumulative variation of *Brassica* anthocyanins. Different types of transposon insertions or sequence variations in the *BoMYB2* promoter in purple *B. oleracea* drove its expression to increase significantly, resulting in different color phenotypes in different tissues or organs of different subspecies (Yan et al., 2019). The activated expression of the *BrMYB2* gene in purple leaf Chinese cabbage and the *BjPur* gene in purple leaf *B. juncea* was due to the deletion of the first large intron in the intergenic region (He et al., 2020; Heng et al., 2020). In *B. napus*, *BnaPAP2.A7* was previously found to be a key transcription factor regulating the synthesis of anthocyanins in the leaves (Chen et al., 2020). Interestingly, *BoMYB2*, *BrMYB2*, *BjPur*, and *BnaPAP2.A7* were homologous of *Bo6g100940*, *BraA07g032100.3C*, *BjuB05g10740S* and *BnaA07T0287000ZS*, which was highly consistent with our results. In addition, our results indicated that *MYB90* was the key transcription factor that combined with *TT8* transcription factor and TTG1 protein to form a MBW transcription regulatory complex to regulate the synthesis of anthocyanins in *Brassica* crops, when *MYB90* was not expressed or *MYBL2* inhibited *MYB90* expression by competing with *TT8* and TTG1, *Brassica* crops failed to complete the synthesis and accumulation of anthocyanins (Fig. 6).

Overall, the present study provided new insights into the roles of novel R2R3-MYB transcription factors in anthocyanin synthesis of *Brassica* crops and into anthocyanin transcriptional regulation in plants.

Declarations

Ethics approval and consent to participate:

This study including the collection on plants material complies with relevant institutional, national, and international guidelines and legislation.

Consent for publication:

Not applicable.

Competing interests:

The authors declare that there are no competing interests.

Statement on plant guidelines and permission:

All materials in this study are comply with relevant institutional, national, and international guidelines, legislation, and sub-section ethical approval and consent to participate.

Gannan Normal University agreed to the sampling of various experimental materials in this study.

Availability of data and materials:

All materials and related data in this study are available upon request. If you need these materials and related data, you can contact Daozong Chen(chendaozong61@163.com).

Funding:

This research was funded by the Natural Science Foundation of Jiangxi Province, grant number 20212BAB215002; the National Natural Science Foundation of China, grant number: 32160454; the Science and Technology Plan Project of Ganzhou, grant number [2020]60.

Authors' contributions:

CT, BZ and CC conceived and designed the experiments. DC, HC, GD, HZ, YL and WS analyzed the data. DC, CT, BZ and CC wrote the manuscript. All authors approved the manuscript and consent to publication this manuscript.

Acknowledgments:

Thanks to Professor Li Zaiyun for his help in the writing process of the thesis.

Conflicts of Interest:

The authors declare that the study was conducted in the absence of any commercial or financial relationships that could be envisaged and/or construed as a conflict of interest.

References

1. Albert N, Davies K, Lewis D, et al. 2014. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *Plant Cell*. 26:962–980.
2. Allan A, Hellens R, Laing W. 2008. MYB transcription factors that colour our fruit. *Trends Plant Sci*. 13:99–102.
3. Anders S, Pyl P, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 31:166–169.
4. Bolger A, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114–2120.
5. Butelli E, Titta L, Giorgio M, et al. 2008. Enrichment of tomato fruit with health-promoting anthocyanins by expression of select transcription factors. *Nat Biotechnol*. 26:1301–1308.
6. Chagné D, Lin-Wang K, Espley RV, et al. 2013. An ancient duplication of apple MYB transcription factors is responsible for novel red fruit-flesh phenotypes. *Plant Physiol* 161:225–239
7. Chen C, Chen H, Zhang Y, et al. 2020a. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant*. 13:1194–1202.
8. Chen D, Liu Y, Yin S, et al. 2020b. Alternatively Spliced *BnaPAP2.A7* Isoforms Play Opposing Roles in Anthocyanin Biosynthesis of *Brassica napus* L. *Front Plant Sci*. 11: 983.
9. Chen G, He W, Guo X, et al. 2021. Genome-wide identification, classification and expression analysis of the MYB transcription factor family in *Petunia*. *Int J Mol Sci*. 22:4838.
10. Chen H, Wang T, He X, et al. 2021. BRAD V3.0: an upgraded Brassicaceae database. *Nucleic Acids Res*. 10: 1057.
11. Chen Y, Yang X, He K, et al. 2006. The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol Biol*. 60:107–24.
12. Devic M, Guillemot J, Debeaujon I, et al. 1999. The BANYULS gene encodes a DFR-like protein and is a marker of early seed coat development. *Plant J*. 19:387–398.
13. Du H, Feng B, Yang S, et al. 2012. The R2R3-MYB transcription factor gene family in maize. *PLoS One*. 7: e37463.
14. Dubos C, Stracke R, Grotewold E, et al. 2010. MYB transcription factors in *Arabidopsis*. *Trends Plant Sci*. 15, 573–581.
15. Eddy S. 2011. Accelerated profile HMM searches. *PLoS Comput Biol*. 7: e1002195.
16. Espley R, Hellens R, Putterill J, et al. 2007. Red colouration in apple fruit is due to the activity of the MYB transcription factor, *MdMYB10*. *Plant J*. 49:414–427.
17. Finn R, Bateman A, Clements J, et al. 2013. Pfam: the protein families database. *Nucleic Acids Res*. 42: D222–D230.
18. Gonzalez A, Zhao M, Leavitt J, et al. 2008. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J* 53: 814–827.

19. Hajiebrahimi A, Owji H, Hemmati S. 2017. Genome-wide identification, functional prediction, and evolutionary analysis of the R2R3-MYB superfamily in *Brassica napus*. *Genome*. 60: 797–814.
20. He Q, Wu J, Xue Y, et al. 2020. The novel gene BrMYB2, located on chromosome A07, with a short intron 1 controls the purple-head trait of Chinese cabbage (*Brassica rapa* L.). *Hort Res*. 7: 97.
21. Heng S, Cheng Q, Zhang T, et al. 2020. Fine mapping of the BjPur gene for purple leaf color in *Brassica juncea*. *Theor Appl Genet*, 133: 2989–3000.
22. Hichri I, Barrieu F, Bogs J, et al. 2011. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J Exp Bot*. 62: 2465–2483.
23. Katoh K, Standley D. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30: 772–780.
24. Kim D, Pertea G, Trapnell C, et al. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 14: R36.
25. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 33: 1870–1874.
26. Larkin M, Blackshields G, Brown N, et al. 2007. Clustal W and clustal X version 2.0. *Bioinformatics*. 23: 2947–2948.
27. Lipsick J. 1996. One billion years of Myb. *Oncogene*. 13, 223–235.
28. Liu J, Wang J, Wang M, et al. 2021. Genome-wide analysis of the R2R3-MYB gene family in *Fragaria* × *Ananassa* and its function identification during anthocyanins biosynthesis in pink-flowered *Strawberry*. *Front Plant Sci*. 12: 702160.
29. Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 17: 10–12.
30. Ogata K, Hojo H, Aimoto S, et al. 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: 6428–32.
31. Paz-Ares J, Ghosal D, Wienand U, et al. 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*. 6:3553–3558.
32. Pertea M, Kim D, Pertea GM, et al. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 11:1650–1667.
33. Price M, Dehal P, Arkin A. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*. 5: e9490.
34. Rahim M, Busatto N, Trainotti L. 2014. Regulation of anthocyanin biosynthesis in peach fruits. *Planta*. 240: 913–929.
35. Ramsay N, Glover B. 2005. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci*. 10: 63–70.

36. Schranz M, Lysak M, Mitchell-Olds T. 2006. The ABC's of comparative genomics in the Brassicaceae: building blocks of *crucifer* genomes. *Trends Plant Sci.* 11: 535–542.
37. Shannon P, Markiel A, Ozier O, et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13: 2498–2504.
38. Shi M, Xie D. 2014. Biosynthesis and metabolic engineering of anthocyanins in *Arabidopsis thaliana*. *Recent Pat Biotechnol.* 8: 47–60.
39. Slabaugh E, Held M, Brandizzi F. 2011. Control of root hair development in *Arabidopsis thaliana* by an endoplasmic reticulum anchored member of the R2R3-MYB transcription factor family. *Plant J.* 67: 395–405.
40. Stracke R, Werber M, Weisshaar B. 2001. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr Opin Plant Biol.* 4:447–456.
41. Walker A, Lee E, Bogs J, et al. 2007. White grapes arose through the mutation of two similar and adjacent regulatory genes. *Plant J.* 49: 772–785.
42. Wang J, Liu Y, Tang B, et al. 2020. Genome-wide Identification and capsaicinoid biosynthesis-related expression analysis of the R2R3-MYB gene family in *Capsicum annuum* L. *Front Genet.* 11: 598183.
43. Wang L, Tang W, Hu Y, et al., 2019. A MYB/bHLH complex regulates tissue-specific anthocyanin biosynthesis in the inner pericarp of red-centered kiwifruit *Actinidia chinensis* cv. Hongyang. *Plant J.* 99: 359–378.
44. Wang N, Ma Q, Ma J et al. 2019. A comparative genome-wide analysis of the R2R3-MYB gene family among four *gossypium* species and their sequence variation and association with fiber quality traits in an interspecific *G. hirsutum* × *G. barbadense* population. *Front Genet.* 10: 741.
45. Wang Z, Tang J, Hu R, et al. 2015. Genome-wide analysis of the R2R3-MYB transcription factor genes in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*) reveals their stress and hormone responsive patterns. *BMC Genomics.* 16: 17.
46. Wang Y, Tang H, Debarry JD, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40: e49.
47. Wong M, Henry R, Ma B, et al. 1998. The large subunit of basal transcription factor SNAPc is a Myb domain protein that interacts with *Oct-1*. *Mol Cell Biol.* 18: 368–377.
48. Xie Q, Hu Z, Zhang Y, et al. 2014. Accumulation and molecular regulation of anthocyanin in purple tumorous stem mustard (*Brassica juncea* var. *tumida* Tsen et Lee). *J Agric Food Chem.* 62: 7813–7821.
49. Xu W, Grain D, Bobet S, et al. 2014. Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in *Arabidopsis* seed. *New Phytol.* 202: 132–144.
50. Yan C, An G, Zhu T, et al. 2019. Independent activation of the BoMYB2 gene leading to purple traits in *Brassica oleracea*. *Theor Appl Genet.* 132: 895–906.

51. Zhang Y, Hu Z, Chu G, et al. 2014. Anthocyanin accumulation and molecular analysis of anthocyanin biosynthesis-associated genes in eggplant (*Solanum melongena* L.). J Agric Food Chem. 62: 2906–2912.
52. Zhao L, Gao L, Wang H, et al. 2013. The R2R3-MYB, bHLH, WD40, and related transcription factors in flavonoid biosynthesis. Funct Integr Genomics. 13: 75–98.

Figures

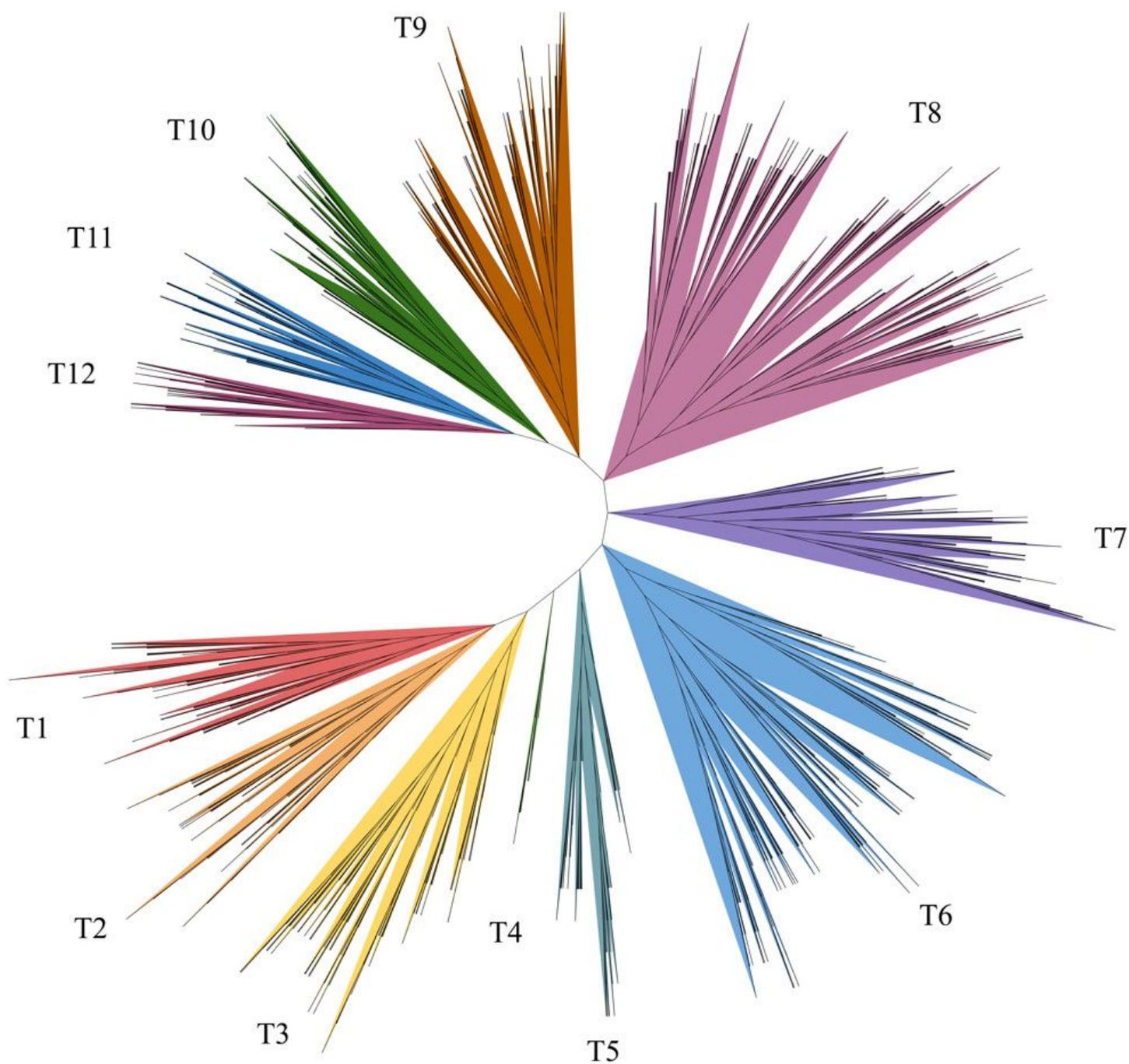


Figure 1

Phylogenetic relationships of R2R3-MYB proteins between six *Brassica* crops and *Arabidopsis*. The lines represented genes from six *Brassica* crops and *Arabidopsis*, respectively. All 12 subfamilies of R2R3-MYBs were well separated in different clades and represented by different colors. The ML phylogenetic tree was generated using JTT algorithm with 1,000 bootstrap value via FastTree2.

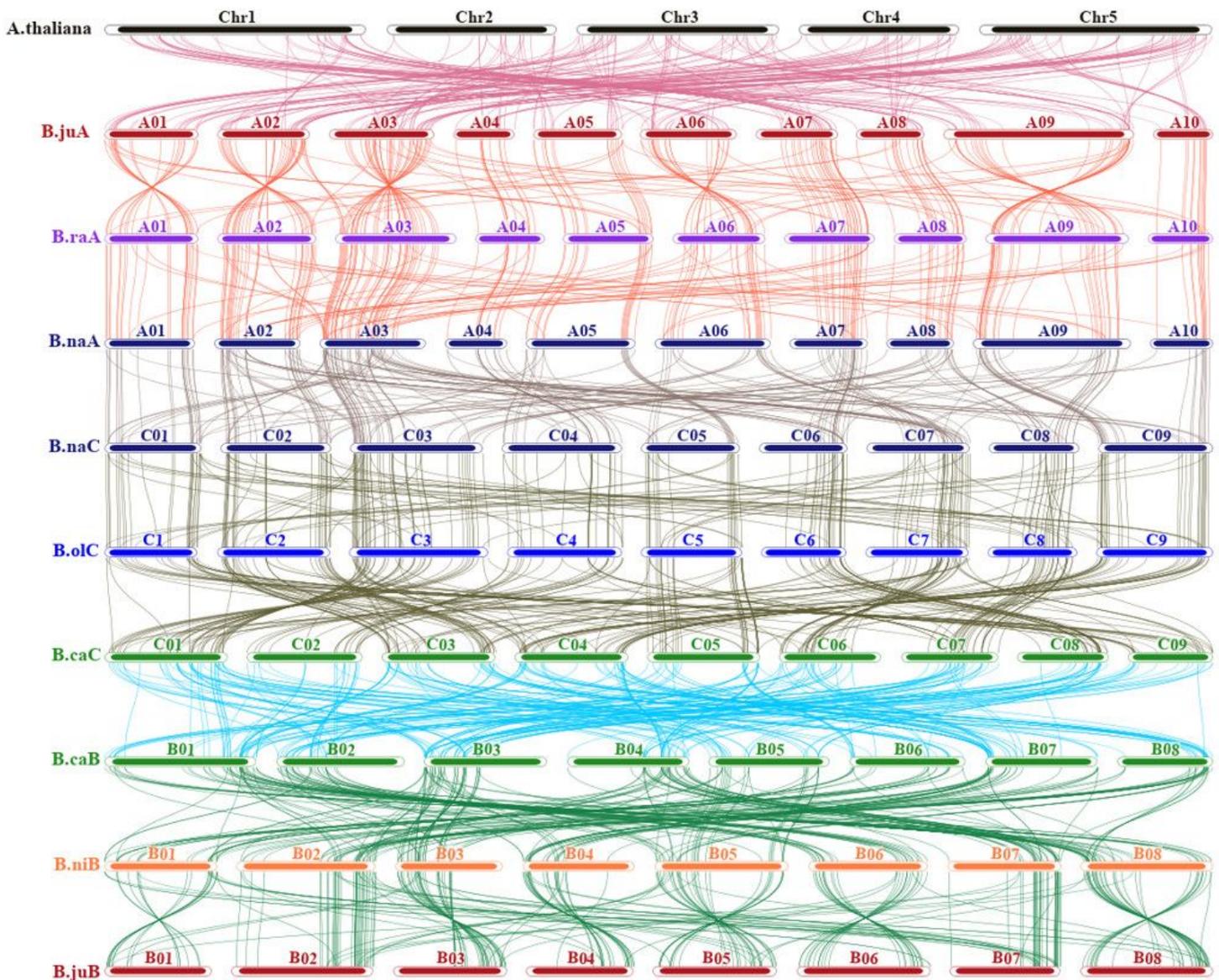


Figure 2

Interspecies synteny of R2R3-MYBs between the six species in U's triangle and *Arabidopsis*. R2R3-MYBs colinearity of the genomes of the six species in U's triangle and *Arabidopsis*, including three diploid species, *B. rapa* (A genome, BraA), *B. nigra* (B genome, BniB), and *B. oleracea* (C genome, BolC) and three tetraploid species, *B. napus* (AACC, BnaA, and BnaC subgenomes); *B. juncea* (AABB, BjuA and BjuB subgenomes); and *B. carinata* (BBCC, BcaB and BcaC subgenomes).

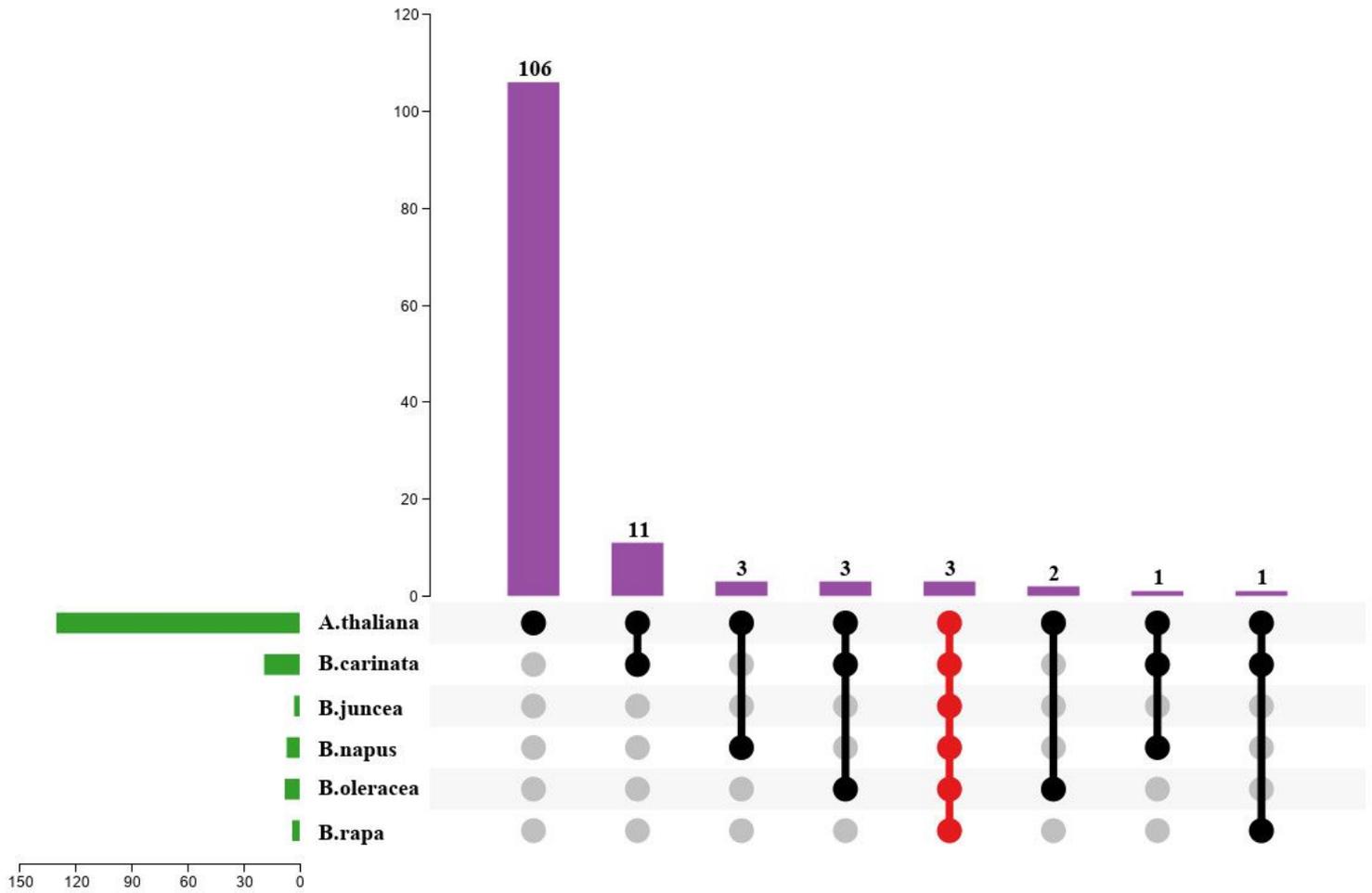


Figure 3

The differentially expressed R2R3-MYBs between five *Brassica* crops of green and purple leaves RNA-seq data. The red color indicates the co-differentially expressed R2R3-MYBs in five *Brassica* crops.

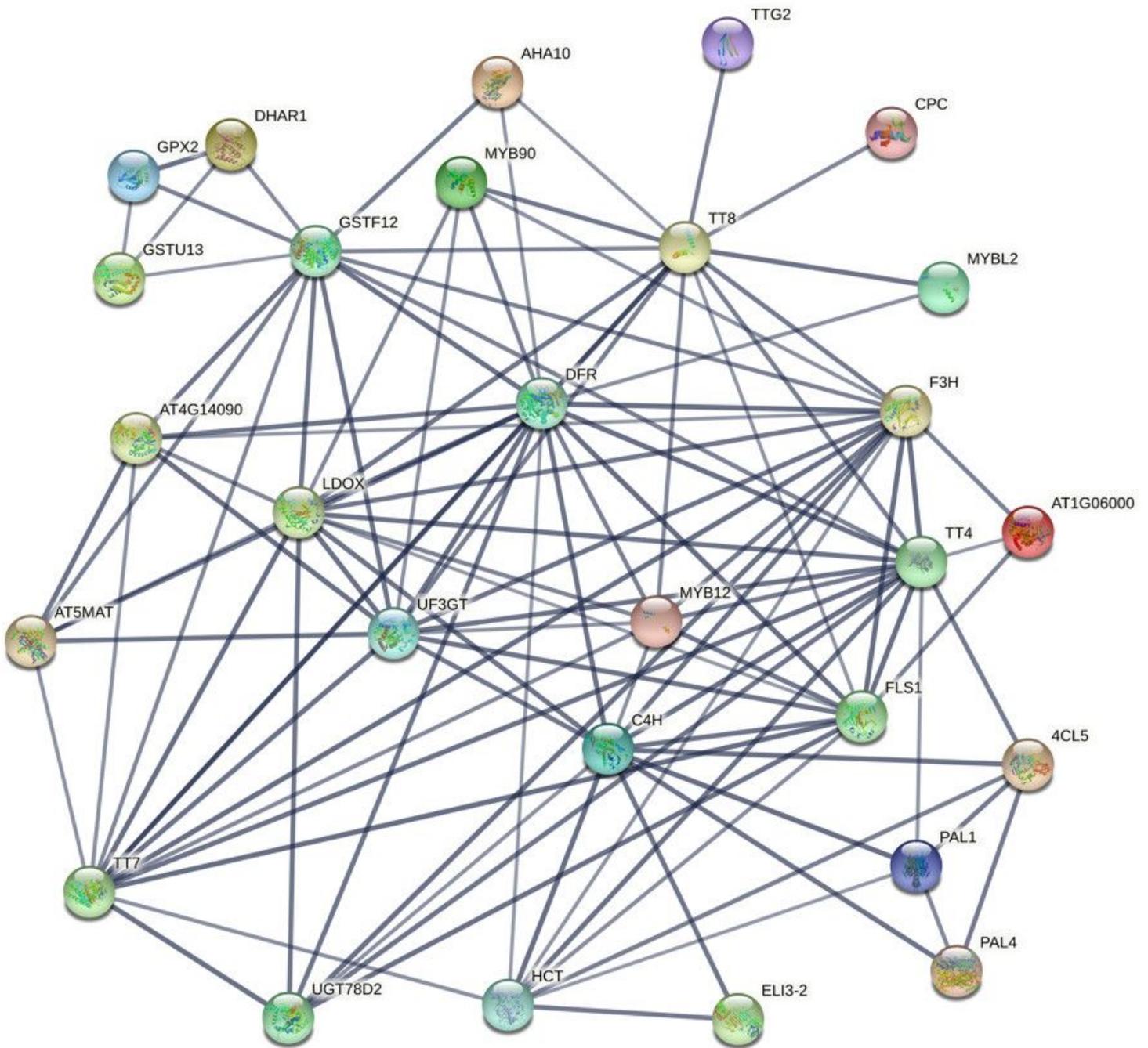


Figure 4

Co-expression networks of R2R3-MYB with anthocyanin biosynthesis pathway genes. The connecting line indicates that there is interaction between each other, and the thickness of the line represents the strength of the interaction.

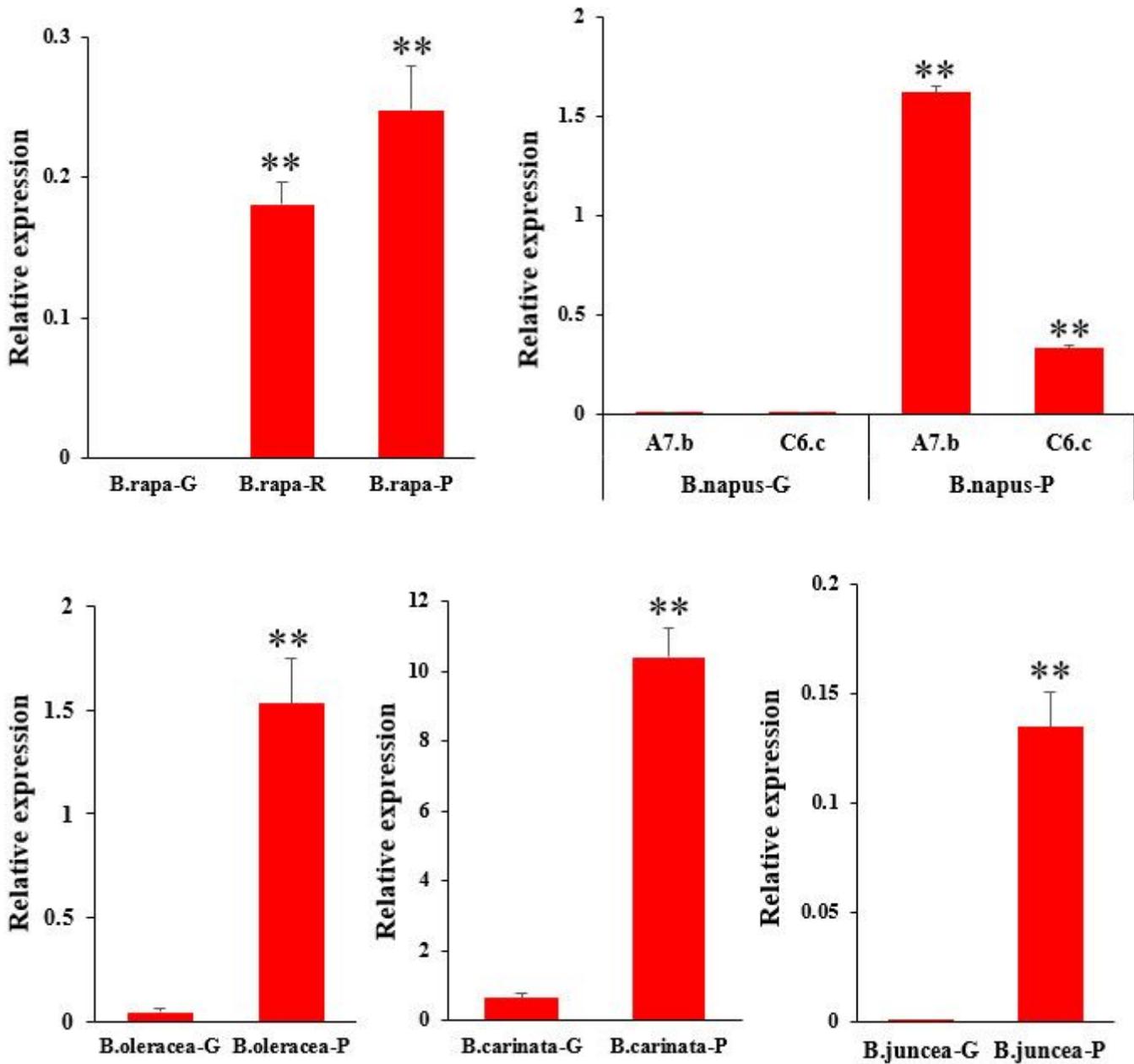


Figure 5

qRT-PCR analysis of *AtPAP2* homologous relative expression in five *Brassica* species with green and purple leaves. For qRT-PCR analysis, both five *Brassica* species green and purple samples are with three replicates, the error bars represent difference between three repetitions of anthocyanins content analysis, and the ** means statistically significant.

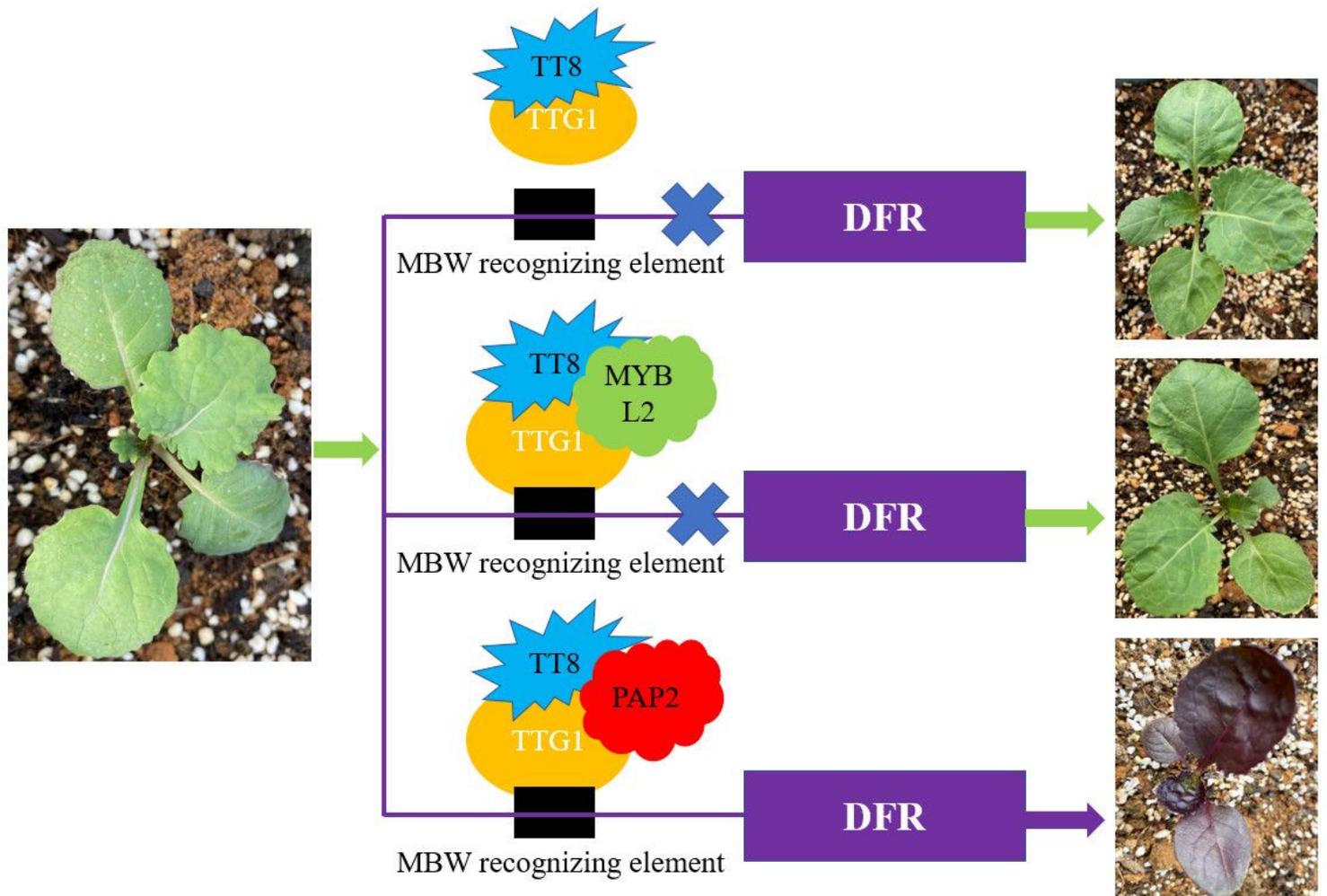


Figure 6

Proposed ‘activator-and-repressor’ loop model for the regulation of anthocyanin accumulation in *Brassica* species. *PAP2* acts as the activator by directly and positively regulating the expression of anthocyanin biosynthetic genes with *TTG1* and *TT8*, whereas *MYBL2* acts as the repressor by inhibiting the activity of the *PAP2-TT8-TTG1* complex. The arrow and error symbol lines indicate activation or repression, respectively.

Supplementary Files

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

- [SupplementaryFigure1.jpg](#)
- [SupplementaryFigure2.jpg](#)
- [SupplementaryFigure3.jpg](#)
- [SupplementaryTableS1.xlsx](#)
- [SupplementaryTableS2.xlsx](#)

- [SupplementaryTableS3.docx](#)
- [SupplementaryTableS4.docx](#)
- [Supportinginformation.docx](#)