BMC Plant Biology

RESEARCH

Open Access

Systematic identification of R2R3-MYB S6 subfamily genes in Brassicaceae and its role in anthocyanin biosynthesis in *Brassica* crops

Daozong Chen[1†], Chenchen Wang[1†], Yi Liu[1], Wenjie Shen[1], Qiushi Cuimu[1], Dawei Zhang[3], Bo Zhu[1], Lunlin Chen[2] and Chen Tan[1*]

Abstract

The Brassicaceae family includes *Arabidopsis thaliana*, various vegetables and oil crops. The R2R3-MYB genes of the S6 subfamily are crucial for regulating anthocyanin biosynthesis, however, their systematic identification in Brassicaceae plants is still incomplete. Here, we systematically identified homologous genes of R2R3-MYB transcription factors from the S6 subfamily across 31 Brassicaceae species. A total of 92 homologous genes were identified, with species representation ranging from 0 to 10 genes per species. Phylogenetic analysis classified these homologous genes into six distinct groups. Notably, approximately 70% of the homologous genes were found within the G6 group, indicating a high degree of evolutionary conservation. Furthermore, a phylogenetic analysis was conducted on 35 homologous genes obtained from six species within the U's triangle *Brassica* plants. The findings provided evidence of significant conservation among orthologous genes across species and demonstrated strong collinearity on sub-genomic chromosomes, with notable tandem duplications observed on chromosomes A7 and C6. Subsequently, we predicted the *cis*-acting elements of these 35 homologous genes, and analyzed their structures, conserved motifs, and characteristic conserved domains, confirming the significant similarities between orthologous genes. Additionally, we employed white and purple flower rapeseed specimens to conduct qRT-PCR validation of the key genes and transcriptional regulators associated with the anthocyanin synthesis pathway. The results revealed significant differential expression of *BnaPAP2.A7.b* in purple flowers, alongside the differential expression of *BnaPAP2.C6.d*. Ultimately, based on previous research and the findings of this study, we propose a transcriptional regulatory framework to govern anthocyanin accumulation in distinct tissues or organs of *B. napus*. Our findings offer a novel perspective on the functional diversification of R2R3-MYB transcription factors within the S6 subfamily homologous genes, while also shedding light on the regulatory network governing anthocyanin biosynthesis in Brassicaceae species.

Keywords Brassicaceae, R2R3-MYB S6 subfamily genes, Anthocyanin biosynthesis, Homologous gene, *Brassica napus*

†Daozong Chen and Chenchen Wang contributed equally to this work.

*Correspondence:
Chen Tan
tanchen2020@gnnu.edu.cn
Full list of author information is available at the end of the article

Introduction

The Brassicaceae family, formerly known as Cruciferous, includes a variety of plants that hold significant importance in agriculture and research [1]. This family encompasses widely cultivated oil crops such as rapeseed (*Brassica napus*) and mustards (*B. juncea*), as well as vegetable species like cabbage (*B. oleracea*), Chinese cabbage (*B. rapa*), radish (*Raphanus sativus*), cut-flower stock (*Matthiola incana*), and model plants such as *Arabidopsis thaliana* [1]. It is estimated that *Brassica* crops and *A. thaliana* diverged approximately 12–20 million years ago [1], followed by a genome-wide tripling event in *Brassica* species that occurred around 5 to 15 million years ago [2]. *Brassica* crops are highly valued both as vegetables and as sources of oil, leading to their widespread cultivation globally and attracting the attention of researchers worldwide. Through a prolonged evolutionary process, *Brassica* crops have given rise to three distinct diploid species: *B. rapa* (AA, $2n=10$), *B. oleracea* (CC, $2n=18$), and *B. nigra* (BB, $2n=16$). Additionally, three allotetraploid species, *B. napus* (AACC, $2n=38$), *B. juncea* (AABB, $2n=36$), and *B. carinata* (BBCC, $2n=34$) have naturally emerged from the hybridization of the three diploid parents in pairs. To date, the genomes of six *Brassica* crops have been sequenced, providing an ideal model and rich information for investigating the evolution of gene functions within this group [3–8].

MYB transcription factors are widely distributed in plants, playing crucial roles in regulating plant development, metabolism, and responses to biotic and abiotic stresses [9–15]. In plants, MYB transcription factors can be categorized into four groups based on the number of adjacent repeat sequences: 1R-MYB (R1/2-MYB, R3-MYB), R2R3-MYB, 3R-MYB (R1R2R3-MYB), and 4R-MYB (R1R2R3R4-MYB) [9–11]. Among these categories, R2R3-MYB has experienced significant expansion in plant lineages, making it the largest MYB gene family [9, 16]. The R2R3-MYB protein sequence can be divided into 25 [9], 24 [10], 27 [17], 28 [12, 18], or 79 [16] subgroups based on the conservation of amino acid motifs in the DNA binding domain and C-terminal domain. The four R2R3-MYB genes, namely *MYB75* (*PAP1*), *MYB90* (*PAP2*), *MYB113*, and *MYB114*, found in *A. thaliana* are universally acknowledged as members of the S6 subfamily due to their possession of distinctive KPRPR[S/T]F domain [9, 10, 12, 16, 18]. Then, recent research findings indicate that a total of 96 genes belonging to the S6 subfamily of R2R3-MYB transcription factors have been identified across various plant species. These genes primarily participate in plant metabolic pathways, including phenylpropanoid metabolism, flavonoid biosynthesis, and anthocyanin biosynthesis [16]. In addition, the changes in expression patterns caused by the variation of *cis*-acting elements in the promoter region of these genes are also worthy of our in-depth exploration.

Anthocyanins, ubiquitous natural pigments, impart red, purple, and blue hues to plant tissues or organs. In plants, the accumulation of anthocyanins plays a pivotal role in bolstering resistance against biotic and abiotic stresses [19]. Anthocyanins, ubiquitous natural pigments, impart red, purple, and blue hues to plant tissues or organs. In plants, the accumulation of anthocyanins assumes a pivotal role in bolstering resistance against biotic and abiotic stresses [20]. The anthocyanin biosynthetic pathway and its transcriptional regulation have been extensively documented in *A. thaliana*. This includes the phenylpropanoid synthesis pathway, the flavonoid synthesis pathway, and the anthocyanin synthesis pathway. The catalytic enzymes involved in these metabolic pathways are primarily regulated by the R2R3-MYB transcription factor [17, 21–23]. Notably, the biosynthesis of anthocyanins is predominantly controlled by the MBW transcriptional regulatory complex, which consists of *PAP1*/*PAP2*/*MYB113*/*MYB114* (S6 subfamily of R2R3-MYB genes) and bHLH transcription factors (*TT8*, *GL3* and *EGL3*) and WD40 protein (TTG1) [17, 21–24]. The initial discovery of the *anthocyanin2* (*an2*) gene within the first S6 subfamily of R2R3-MYB genes in *Petunia hybrida* highlighted its crucial role in regulating the anthocyanin biosynthetic pathway and determining color variations [25]. Subsequently, numerous other members of the S6 subfamily of R2R3-MYB genes have been identified and implicated in the regulation of plant anthocyanin biosynthesis [16]. However, a comprehensive investigation and analysis of the S6 subfamily of R2R3-MYB genes in Brassicaceae species, excluding *A. thaliana*, has yet to be reported. In *Brassica* plants, the genome has experienced significant variation in copy number and potential differentiation of functions due to polyploidization and hybridization. This process contributes to a marked increase in the number of S6 subfamily copies of R2R3-MYB genes, and the functional redundancy that arises during evolution may result in functional differentiation among these copies. Nonetheless, there is a lack of systematic identification and research on this matter.

In this study, we conducted a systematic identification and developmental analysis of the S6 subfamily of R2R3-MYB genes in Brassicaceae species, utilizing data from31 published reference genomes. We subsequently, focused on six *Brassica* species to analyze the evolution, synteny, conserved domains, and *cis*-acting elements of the R2R3-MYB gene S6 subfamily. Our findings revealed significant tandem duplications on certain chromosomes. Additionally, we employed transcriptome sequencing data from

Brassica crop leaves to perform an expression pattern analysis, which indicated that the functions of different gene copies may have undergone substantial changes. Furthermore, we utilized transcriptome sequencing data from various tissues and petals exhibiting diverse flower colors in *B. napus* to examine the expression patterns of ten S6 subfamily R2R3-MYB genes. Our results indicate that *BnaPAP1.A07.b* and *BnaPAP1.C06.d* are significantly up-regulated in tissues associated with anthocyanin synthesis and accumulation. Overall, our study systematically identified members of the R2R3-MYB gene S6 subfamily in Brassicaceae species and analyzed expression patterns in *Brassica* crops providing new insights into the anthocyanin transcriptional regulatory network in *Brassica* plants.

Materials and methods

Systematic identification of R2R3-MYB transcription factors of S6 subfamily in Brassicaceae plants

In this study, all sequences of 31 Brassicaceae species were downloaded from the TBGR database (http://www.tbgr.org.cn/). The nucleic acid sequence and protein sequence of *PAP1*, *PAP2*, *MYB113* and *MYB114* from *A. thaliana* were used as seed sequences for BLASTN and BLASTP (coverage = 60%, identity = 60%, e-value = 1.00E-20) comparison among 31 species of Brassicaceae, including six *Brassica* species (Supplementary Table 1), and homologous genes were screened. The protein sequence of homologous gene from 31 species of Brassicaceae was extracted and submitted to BatchCDD (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) analyses conservative structure domain, determine the members of the family of genes in combination with the results of the BLASTP. For detailed identification methods, please refer to Chen et al. [26] and Pucker [27].

Construction of a phylogenetic tree for 92 R2R3-MYB transcription factors of S6 subfamily homologous genes

To construct a phylogenetic tree of 92 R2R3-MYB transcription factors of S6 subfamily homologous genes, we utilized the *CER1* (*AT1G02205.1*) of *A. thaliana* as the outgroup. For the construction of the phylogenetic tree, we employed IQ-TREE2 software (v1.6.12) [28], with the -m parameter selecting MFP to automatically determined the optimal model and generate the tree. To validate the findings of the IQ-TREE2 phylogenetic tree, we employed the RAxML-NG software (v1.1) [29] for constructing a phylogenetic tree. The LG model was chosen for tree construction, with the –bs-trees parameter set to 1000 and the -all parameters selected to facilitate multiple search aimed at identifying the optimal tree topology and model parameters. Subsequently, the phylogenetic tree was visualized using the iTOL online website (https://itol.embl.de/, v6.8) [30].

Synteny analysis of 35 R2R3-MYB transcription factors of S6 subfamily homologous genes in six *Brassica* crops

In order to analyze the collinear relationship of 35 R2R3-MYB transcription factors of S6 subfamily homologous genes in six species of *Brassica*, the collinearity of intraspecific and interspecific genes was determined using the BLASTP (E-value: 1e-20, max_target_seqs:10) and SynOrthos (v1.0, http://brassicadb.cn/#/Download/). Then, the collinearity diagram of 35 homologous genes on each chromosome was drawn by TBtools-II (v2.096) software [31].

Cis-acting elements, gene structure, conserved motif, and conservative domain analysis

Based on the reference genome annotation information of six *Brassica* species, all 35 R2R3-MYB transcription factors of S6 subfamily homologous genes in six *Brassica* crops upstream of the start codon 2 Kb genome sequences were extracted and used for the prediction of *cis*-acting elements in the promoter region. PlantCARE website (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used to predict the *cis*-acting elements. Then, all predicted *cis*-acting elements were classified and analyzed. The MEME online tool (https://meme-suite.org/meme/) was used to investigate conserved domains 35 R2R3-MYB transcription factors of S6 subfamily homologous proteins in six *Brassica* crops. Then, the identified 35 R2R3-MYB transcription factors of S6 subfamily homologous protein sequences were extracted and submitted to CDD online website (https://www.ncbi.nlm.nih.gov/cdd/) to obtain conserved domain information, and the parameters were default parameters. Phylogenetic trees, motifs and *cis*-acting elements were mapped by TBtools-II (v2.096) software [31] using phylogenetic files (from RAxML-NG), *cis*-acting elements prediction files (from PlantCARE), gene structure (from gff3 files of genome annotation), conserved motifs (from MEME), and conversed domain (from CDD).

Analysis of the expression pattern of R2R3-MYB transcription factors of S6 subfamily homologous genes in *Brassica* crops

To further investigate the expression patterns of R2R3-MYB transcription factors of S6 subfamily homologous genes in various plants and tissues of *Brassica* crops, purple and green leaf transcriptome data of *B. rapa*, *B. oleracea*, *B.*

juncea and *B. napus*, and transcriptome data from *B. napus* leaves stems (ZS11, green; ZG-1, purple), flower petals (CQY311, orange-red; ZS11, yellow, ZH-13, purple; BH-27, white; CQY40, beige; CQY439, apricot), siliques (ZS11, green; ZG-1, purple), and seed coat (No2127, yellow; ZS11, black) were collected and analyzed. All RNA-seq data were obtained from the NCBI (additional Center for Biotechnology Information (NCBI) at https://www.ncbi.nlm.nih.gov/ through the following biological projects PRJNA554517 for leaves, PRJAN855492 and PRJNA1054216 for stems and flower petals, PRJNA734925 for siliques, and PRJNA597958 for seeds. The original transcriptome data were initially examined and statistically analyzed using FastQC software. Following this, the transcriptome data were aligned and compared with Hisat2 software, allowing for the selection of uniquely aligned reads for further analysis. Subsequently, htseq-count and StringTie software were employed to perform count statistics and calculate FPKM values for expression quantities. Finally, DESeq2 was utilized for differential expression analysis. For transcriptome sequencing data analysis methods, please refer to Chen et al. [26]. The expression histogram was generated using Excel, while the heat map was generated using TBtools-II (v2.096) software [31].

Plant materials, RNA extraction, and qRT-PCR analysis

The purple (purple flower inbred lines, ZH-13) and white (white flower inbred lines, BH-27) *B. napus* flower buds waiting to open with three biological replicates were collected, all samples were collected and immediately frozen in liquid nitrogen for RNA extraction and qRT-PCR analysis.Eastep® Super Total RNA Extraction Kit of Promega (Beijing) Biotech Co., Ltd was used to extract total RNA, and HiScript III 1st Strand cDNA Synthesis Kit (+gDNA wiper) of Nanjing Vazyme Biotech Co., Ltd was used to reverse transcription synthesis of cDNA. QRT-PCR was performed using SYBR qPCR Master Mix from Nanjing Vazyme Biotech Co., Ltd. The reaction system was 20.0 μL: SYBR qPCR Master Mix 10.0 μL, the primers: positive and negative primers 0.5 μL, cDNA template 4.0 μL, ddH_2O 5.0 μL. Amplification procedure: predenaturation at 95℃ for 30 s; 40 cycles were performed at 95℃ for 10 s, 60℃ for 30 s; 95℃ 15 s, 60℃ 60 s, 95℃ 15 s, melting curve. Each sample contains three technical replicates. The relative expression level was calculated by $2^{-\Delta\Delta Ct}$ method, the difference multiple was calculated according to the relative expression level, and finally the average value was obtained, which was visualized by Prism9 (v9.5.1) software. The primer information for qRT-PCR analysis were listed in Supplementary Table 6. For detailed methods, please refer to Chen et al. [32]. All materials of this research were cultivated in a basic greenhouse at Gannan Normal University (N25°47′, E114°52′).

Table 1 Genome version information and R2R3-MYB transcription factors of S6 subfamily homologous genes in 31 Brassicaceae species

Species	Version	BLAST	CDD	Result
Arabidopsis_thaliana	Tair10	4	4	4
Arabidopsis_halleri	A.halleri-v2.2	2	2	2
Arabidopsis_lyrata	Lyrate-v2.1	4	4	4
Arabis_alpina	Gray-v4.0	1	1	1
Aethionema_arabicum	A.arabicum-v1.0	0	0	0
Barbarea_vulgaris	Bittercress-v1.0	3	3	3
Boechera_retrofracta	Holboell-v1.0	3	3	3
Boechera_stricta	Drummond-v1.2	2	2	2
Brassica_carinata	Zd-1-v1.0	8	8	8
Brassica juncea	SCYZ	7	6	7
Brassica_napus	Darmor-v10.0	10	8	10
Brassica_nigra	Ni100_LR-v2.0	3	3	3
Brassica_oleracea	JZS-v2.0	4	4	4
Brassica_rapa	Chiifu-v3.5	3	3	3
Camelina_sativa	Camelina-v2.0	5	5	5
Capsella_grandiflora	C.grandiflora.v1.0	2	2	2
Capsella_rubella	Red_shepherd-v1.1	1	1	1
Cardamine_hirsuta	Hairy-v1.0	3	3	3
Eutrema_salsugineum	173-v1.0	1	1	1
Isatis_indigotica	Woad-v1.0	2	2	2
Leavenworthia_alabamica	Alabama-v1.0	2	2	2
Lepidium_meyenii	Peruvian-v1.0	1	1	1
Matthiola_incana	M.incana-v1.0	2	2	2
Microthlaspi_erraticum	M.erraticum-v1.0	3	3	3
Orychophragmus_violaceus	O.violaceus-v1.0	2	2	2
Raphanus_sativus	Radish-v1.0	3	3	3
Schrenkiella_parvula	Saltwater-v1.0	1	1	1
Sinapis_alba	S.alba-v1.0	4	4	4
Sinapis_arvensis	S.arvensis-v1.0	4	4	4
Sisymbrium_irio	London-v1.0	1	1	1
Thlaspi_arvense	Field-v1.1	1	1	1
Total	31	92	89	92

Results

Systematic identification and phylogenetic analysis of R2R3-MYB transcription factors of S6 subfamily in Brassicaceae

To systematically ascertain the constituents of the R2R3-MYB gene S6 subfamily within Brassicaceae species, we utilized the *PAP1*, *PAP2*, *MYB113*, and *MYB114* sequences from *A. thaliana* as seed sequences. By leveraging accessible genomic data, we conducted an investigation to identify potential candidate genes across 31 Brassicaceae species. Our analysis revealed a total of 92 R2R3-MYB S6 subfamily gene members, with copy numbers ranging from 1 to 10, except for *Aethionema*

arabicum, which can be accurately identified (Table 1, Supplementary Table S1). Specifically, 7 species contained one S6 subfamily homologous gene, 7 species contained two, another 7 species contained three, and the remaining 10 species possessed S6 subfamily homologous genes in four or more copies. Notably, in *B. napus* and *B. juncea*, discrepancies were observed between the BLAST results of protein sequences and nucleic acid sequences, as well as the predicted outcomes of CDD conserved domains. During the identification of conserved domains, it was noted that the conserved domains of *C06p39700.1.BnaDAR* and *BjuB05g10780S* transitioned from the PLN03212 superfamily to the PLN03091 superfamily. However, no conserved domain was detected in *C06p39710.1.BnaDAR* (Supplementary Fig. 1; Supplementary Table 1). Subsequently, further analysis was performed using genomic information from other strains of the two species, revealing that all three genes are homologous to R2R3-MYB genes of the S6 subfamily. Consequently, these three genes were retained for additional analysis. Interestingly, within the *A. thaliana* species, it has been observed that among the four R2R3-MYB S6 subfamily genes, *MYB113* (*AT1G66370*), *MYB114* (*AT1G66380*), and *PAP2* (*AT1G66390*), tandem duplications exist (Supplementary Table 1). These duplications are believed to have arisen during the evolutionary process of the *A. thaliana* genome. It is noteworthy that such phenomena have not been observed in *A. thaliana*. However, in *Brassica* crops, particularly in *B. napus*, the tandem duplication phenomenon is widespread on the A7 and C6 chromosomes. Notably, the C6 chromosome in *B. napus* exhibits four tandem duplications, which are hypothesized to be a result from chromosome rearrangement.

To investigate the evolutionary relationships among the homologous genes of the R2R3-MYB S6 subfamily in 31 Brassicaceae species, a maximum likelihood phylogenetic tree was constructed using their protein sequences. The 92 R2R3-MYB transcription factors of the S6 subfamily homologous genes were categorized into six Groups (G1-G6) based on their phylogenetic relationships (Fig. 1, Supplementary Table 1). Specifically, *g30510.t1* from *A. halleri*, *AL1G66590.t1* from *A. lyrata*, and *AT1G56650.1* from *A. thaliana* were assigned to separate groups. The G4, G5, and G6 groups consisted of 12, 14, and 63 genes, respectively. Approximately 70% of the homologous genes belonging to the R2R3-MYB transcription factors of the S6 subfamily were found in the G6 group, suggesting a high level of conservation in the R2R3-MYB genes of the S6 subfamily throughout evolution. Notably, all homologous genes of *Brassica* crops were found in the G6 group, providing evidence that *Brassica* crops diverged at a later stage compared to *A. thaliana*. Furthermore, the whole genome triplication events unique to *Brassica* crops and the allopolyploidization have resulted in the abundant quantitative variation of the R2R3-MYB genes of the S6 subfamily in *Brassica* crops and their relatively late separation in evolution [4, 33].

Phylogenetic and collinearity analysis of R2R3-MYB transcription factors of S6 subfamily genes in six *Brassica* crops

The *Brassica* genus is of considerable significance within the Brassicaceae family, encompassing a diverse array of agricultural and horticultural crops, and it possesses substantial research value [3]. Specifically, the six species within the U' s triangle serve as ideal models for investigating gene phylogeny and functional evolution [8]. To elucidate the evolutionary relationship of R2R3-MYB transcription factors belonging to the S6 subfamily across the six *Brassica* species, we conducted phylogenetic and collinearity analyses on 35 homologous genes of the S6 subfamily within *Brassica* crops (Fig. 2). These 35 R2R3-MYB transcription factors can be categorized into four distinct branches, which exhibit significant differentiation from the *AtPAP1* gene in *A. thaliana*. Notably, the first branch comprises three genes originating from three allotetraploid species of *Brassica*, while the remaining three branches encompass homologous genes from all six species of *Brassica* (Fig. 2A, Supplementary Table 2). Furthermore, the collinearity analysis of these 35 R2R3-MYB transcription factors reveals a consistent and orderly distribution across the chromosomes of *Brassica* plants. These homologous genes are mainly distributed on chromosomes A02, A03 and A07 in the A subgenome, on chromosomes C02, C03 and C06 in the C subgenome, and on chromosomes B03 and B05 in the B subgenome. Notably, the R2R3-MYB transcription factors of the S6 subfamily are most concentrated on the A7 and C6 chromosomes, particularly on the *B. napus* C6 chromosome, which contains four tandem repeats (Fig. 2B, Supplementary Table 3).

Conserved motifs, putative *cis*-acting elements, and gene structure analysis of R2R3-MYB transcription factors of S6 subfamily homologous genes in *Brassica* species

To conduct a comprehensive investigation into the evolutionary patterns of the 35 homologous genes belonging to the R2R3-MYB transcription factor S6 subfamily in *Brassica*, we analyzed the promoter regions of these genes to identify *cis*-acting elements and gene structures. Additionally, we utilized protein sequence information to identify conserved motifs and unique conserved domains (Fig. 3; Supplementary Table 4). The 2Kbp upstream sequences of the coding regions of the 35 R2R3-MYB transcription factor S6 subfamily

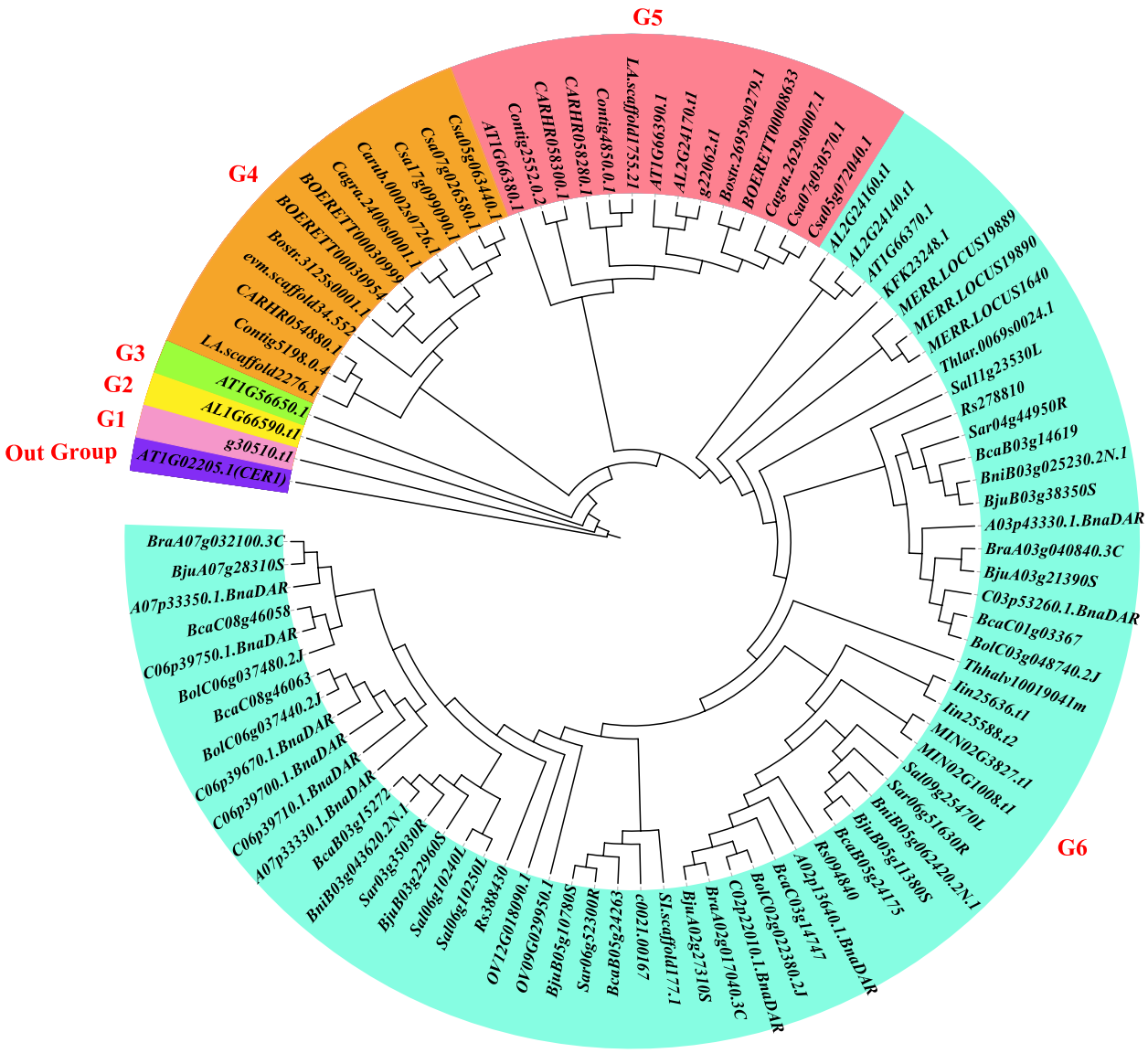

Fig. 1 Phylogenetic relationships of MYB5 proteins between 31 Brassicaceae species. According to the evolutionary distance, 92 genes were divided into 6 subfamilies, G1, > 1.5; G2, 1.5–0.3; G3, 0.3–0.2; G4, 0.2–0.1; G5, 0.1–0.01;G6, < 0.01. All six subfamilies of R2R3-MYB transcription factors of S6 subfamily proteins were well separated in different clades and represented by different colors. IQ-TREE software (v1.6.12) to build a phylogenetic tree, where the -m parameter selects MFP, automatically detects the best model and builds a tree

homologous genes were subjected to *cis*-regulatory element prediction using the PlantCARE online tool. A total of 11 types of *cis*-regulatory elements were predicted in the promoter regions of these genes, including abscisic acid responsiveness, anaerobic induction, auxin responsiveness, defense and stress responsiveness, gibberellin-responsiveness, light responsiveness, MeJA-responsiveness, meristem expression, and salicylic acid responsiveness, zein metabolism regulation, and MYB binding site (Fig. 3B). The *cis*-elements exhibit a high degree of conservation among closely related genes. For instance, the *cis*-acting elements of genes *BniPAP1.B03.a*, *BcaPAP1.B03.a* and *BjuPAP1.B03.b* share fundamental similarities, with the exception of an additional tandemly repeated light responsiveness component found in *BjuPAP1.B03.b* (Fig. 3A-B). The gene structure of the 35 homologous genes belonging to the R2R3-MYB transcription factor S6 subfamily is highly conserved, primarily consisting of three exons and two introns. However, certain genes within this subfamily have undergone structural variations. Notably, the third exon of *BnaPAP1.C06.b* and

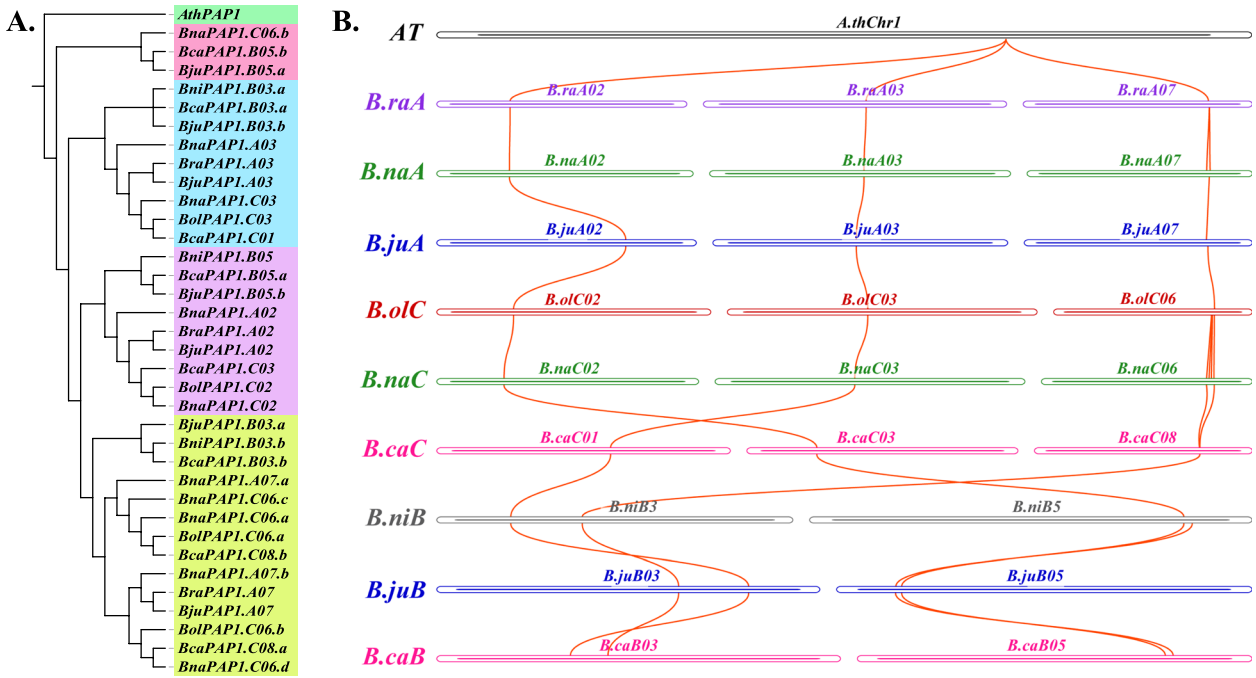

Fig. 2 Phylogenetic and collinear relationships among S6 subfamily genes R2R3-MYB transcription factors in six *Brassica* crops. **A** Phylogenetic relationships of R2R3-MYB transcription factors of S6 subfamily genes in six *Brassica* crops; **B** Collinear relationship among S6 subfamily genes R2R3-MYB transcription factors in six *Brassica* crops

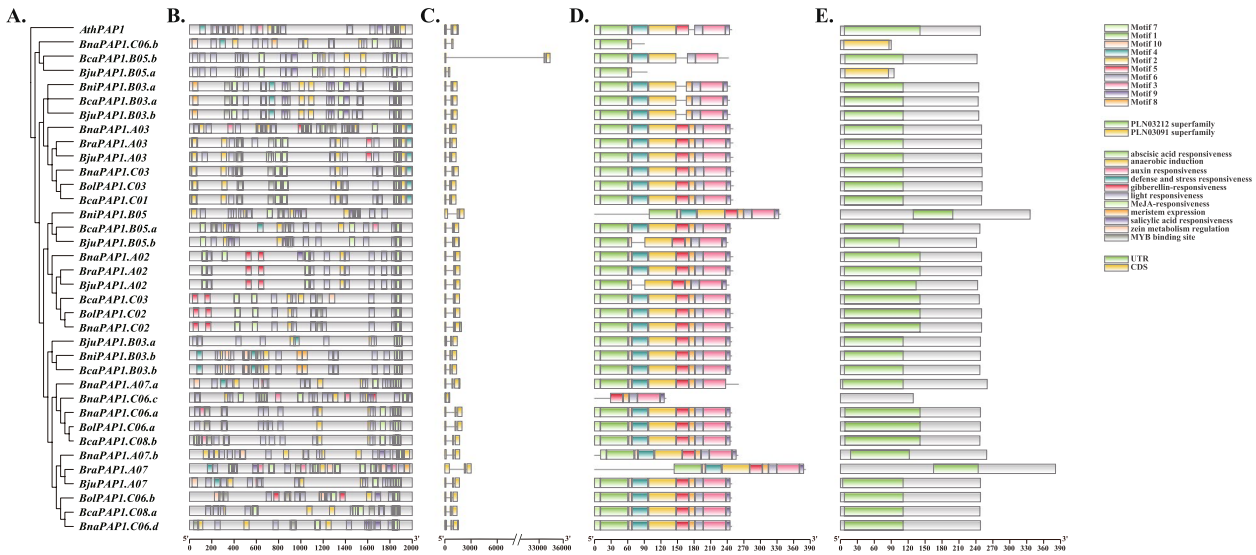

Fig. 3 Phylogenetic tree, motif, conserved domain, putative *cis*-elements and gene structure analysis of R2R3-MYB transcription factors of S6 subfamily homologous genes in Brassica species. **A** Phylogenetic tree; **B** Predicted *cis*-elements; **C** Gene structures; **D** Protein motifs; **E** Conserved domain

BjuPAP1.B05.a have been lost, while the first exon of *BnaPAP1.C06.c* is absent (Fig. 3C).

The analysis of conserved motifs revealed the presence of ten predominant motifs within a set of 35 homologous proteins belonging to the R2R3-MYB transcription factor S6 subfamily. These motifs were found to be relatively consistent across four distinct groups. Notably, certain conserved motifs in specific proteins underwent significant alterations. For instance, BnaPAP1.C06.b and BjuPAP1.B05.a lost seven conserved motifs, retaining only

motif1, motif7, and motif10. Consequently, their conserved domains transitioned from the PLN03212 superfamily to the PLN03091 superfamily. (Fig. 3D). BnaPAP1.C06.c exhibited the loss of five conserved motifs that possess the typical characteristics of R2R3-MYB transcription factors, resulting in the absence of its PLN03212 superfamily (Fig. 3E). This observation provides additional insight into the perplexing issue of successfully identify these three proteins through BLAST, while encountering difficulties in their identification through conserved domain identification in systematic analyses [10, 12]. The principal factor contributing to this discrepancy is the occurrence of tandem duplication, which induces substantial variations in the conserved domain of the protein [16].

Analysis of R2R3-MYB transcription factors of S6 subfamily homologous genes expression patterns related to anthocyanin biosynthesis in *Brassica* crops

In order to gain a deeper understanding of the expression pattern of R2R3-MYB transcription factor S6 subfamily homologous genes in *Brassica* crops, we conducted an analysis of gene expression utilizing available comparative transcriptome sequencing data. Specifically, we gathered transcriptome sequencing data from the purple and green leaves of four *Brassica* crops for this study. Our analysis revealed a diverse range of changes in the expression pattern of R2R3-MYB transcription factor S6 subfamily homologous genes (Fig. 4, Supplementary Table 5). Notably, in *B. juncea*, only *BjuPAP1.B05.b* exhibited significantly higher expression in purple leaves, while five copies were expressed in green leaves. In *B. oleracea*, the gene *BolPAP1.C02* displayed high expression levels in both purple and green leaves, whereas *BolPAP1.C06.b* was specifically highly expressed in purple leaves. In *B. rapa*, all three copies of the gene were expressed in purple leaves, with the expression level of *BraPAP1.A07* being significantly higher in purple leaves compared to green leaves. In *B. napus*, the gene *BnaPAP1.C02* was expressed in both purple and green leaves, while the expression levels of *BnaPAP1.A07.b* and *BnaPAP1.C06.d* were significantly higher in purple leaves than in green leaves (Fig. 4A).

Ulteriorly, we employed comparative transcriptome data of various tissues of *B. napus*, including leaves, stems, petals, siliques, and seed coats, to investigate the expression pattern of ten homologous genes belonging to the S6 subfamily of R2R3-MYB transcription factors in *B. napus* (Fig. 4B). The findings revealed that *BnaPAP1.C02* exhibited high expression levels in multiple purple and green tissues of *B. napus*, whereas *BnaPAP1.A07.b* and *BnaPAP1.C06.d* displayed significantly elevated expression levels specifically in purple tissues (Fig. 4B). Additionally, in order to provide further validation for the aforementioned findings, transcriptome sequencing was employed to analyze the petals of six different flower colors (yellow, beige, white, red, apricot, and purple) in *B. napus*. The obtained results demonstrated the presence of *BnaPAP1.A07.b* in three dark-colored flowers. Furthermore, the expression levels of *BnaPAP1.A02* and *BnaPAP1.C06.d* were significantly up-regulated in purple petals (Fig. 4C). These findings suggest that *BnaPAP1.A07.b*, along with its orthologous genes *BraPAP1.A07*, *BolPAP1.C06.b*, and *BjuPAP1.B05.b*, play a crucial role in the regulation of anthocyanin biosynthesis and accumulation. Moreover, it is evident that other copies of these genes exhibit a wider range of variations in their functions.

Analysis of expression patterns of anthocyanin biosynthesis related genes in different flower colors of *B. napus* by qRT-PCR

To corroborate the findings from the high-throughput sequencing analysis, we conducted additional investigation using real-time quantitative RT-PCR (qRT-PCR) to examine the transcriptional expression pattern of differentially expressed genes (DEGs) involved in anthocyanin synthesis. A total of ten genes associated with the anthocyanin biosynthetic pathway were selected for qRT-PCR analysis, including three early structural genes (EBGs), three late structural genes (LBGs), and four transcriptional regulators (Fig. 5). The results of this study indicate that there were significant differential expressions in six structural genes of the anthocyanin biosynthetic pathway (*BnaCHS.A02*, *BnaCHI.C04*, *BnaF3H.A09*, *BnaDFR.C09*, *BnaANS.C01*, and *BnaUGT.A08*) in purple petals. Additionally, among the four transcription factors examined, *BnaPAP1.A07.b* and *BnaLBD39.A01* exhibited significant differential expression in purple flowers, while *BnaPAP1.C06.d* and *BnaTT8.C09* demonstrated differential expression in the same flowers (Fig. 5). Furthermore, all genes displayed similar relative expression levels, consistent with the results from high-throughput sequencing, thereby confirming the reliability of our transcriptome data analysis.

Discussion

The Brassicaceae family comprises approximately 338 genera and around 3,700 species, primarily native to temperate regions in the northern hemisphere and have since been extensively cultivated worldwide [34]. Through a comprehensive evolutionary process, Brassicaceae species have been classified into five distinct lineages based on their phylogenetic relationships [35, 36]. Notably, the genome of *A. thaliana* was successfully sequenced over two decades ago, marking a significant milestone as the

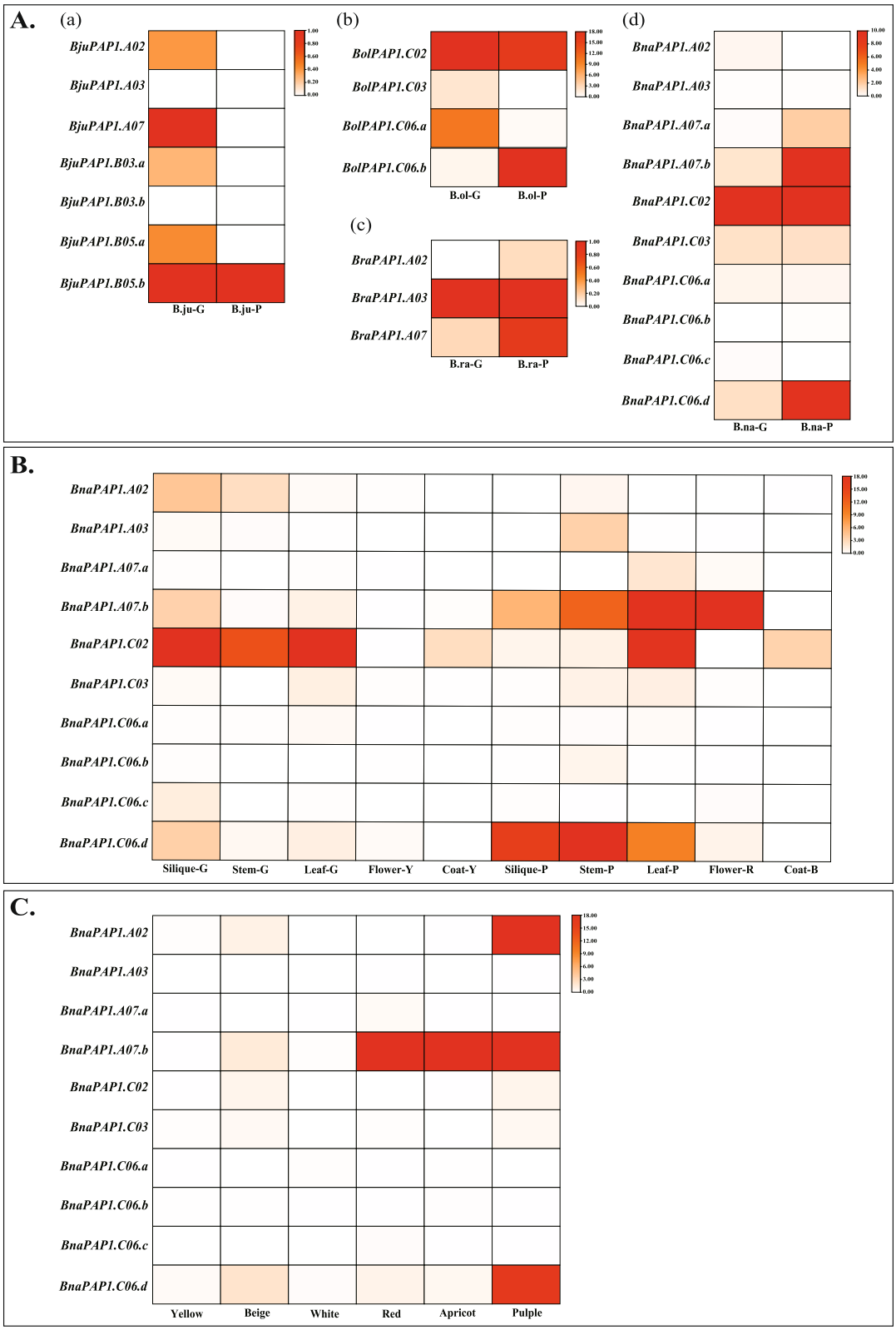

Fig. 4 The expression patterns of R2R3-MYB transcription factor S6 subfamily homologous genes in *Brassica* crops and in different tissues of *B. napus*. **A** The expression patterns in four *Brassica* crops; **B** The expression patterns in different tissues of *B. napus*; **C** The expression patterns in different colors of *B. napus* flowers

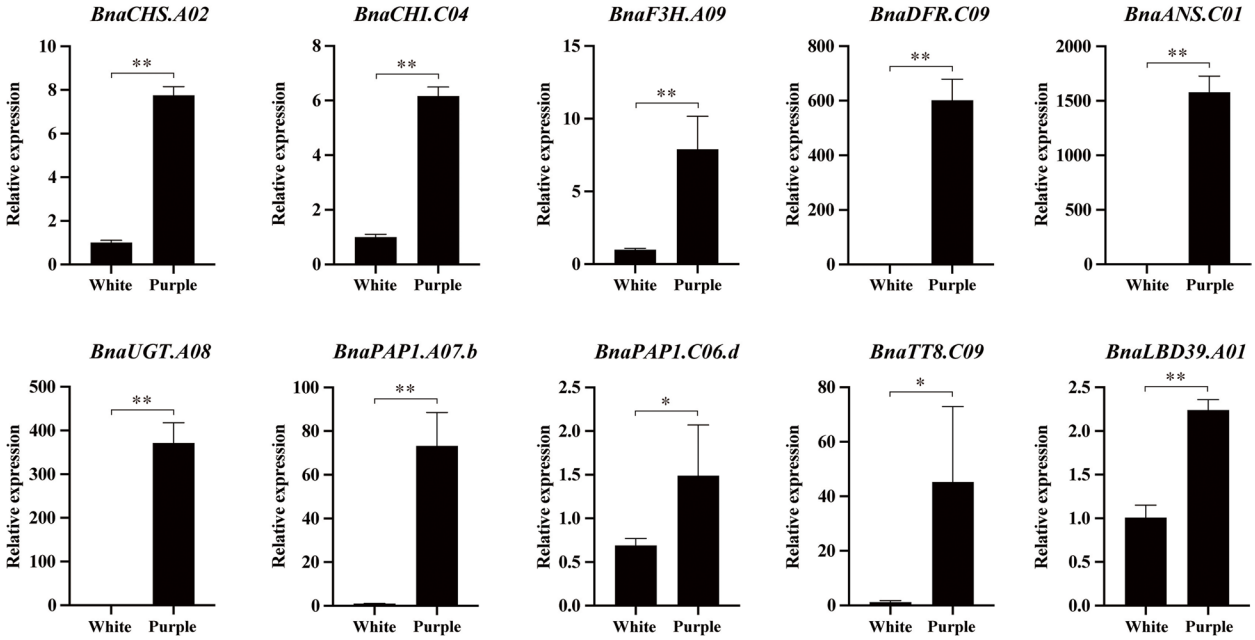

Fig. 5 qRT-PCR analysis of expression patterns of anthocyanin biosynthesis-related genes in white and purple flower *B. napus*. Ten genes were selected for qRT-PCR analysis, *BnaCHS.A02*, *BnaUGT.A08*, and *BnaF3H.A09* were categorized as early structural genes (EBGs), while *BnaDFR.C09*, *BnaANS.C01*, and *BnaF3H.A08* were classified as late structural genes (LBGs), *BnaPAP1.A07.b*, *BnaPAP1.C06.d*, *BnaTT8.C09*, and *BnaLBD39.A01* were identified as the primary regulatory genes, both white and purple flower samples with three biological replicates, at the same time, three technical repetitions were made for each repetition, the *BnaActin3* was used housekeeping gene for the normalization strategy of qRT-PCR analysis, the point represents the mean value of three technical replicates in a representative biological experiment, the error bars indicate s.d, student's t-test, $^{**}P < 0.01$, $^{*}P < 0.05$

first fully sequenced plant species and propelling investigations into the complexities of plant life to unprecedented heights [37]. In recent years, nearly one hundred genome versions of 31 Brassicaceae species have been deciphered, providing valuable insights for studying the systematic evolution and gene functions of Brassicaceae plants [38].

Evolution of R2R3-MYB transcription factors of S6 subfamily in Brassicaceae

The evolutionary process of the Brassicaceae family has a lengthy history, particularly evident in the discernible disparities between species belonging to different genera. Notably, Brassicaceae species have experienced an early instance of whole-genome duplication, followed by a distinct tripling event subsequent to the divergence of *Brassica* crops [3, 5]. As a result, *Brassica* crops can be considered representative triploid species in relation to *A. thaliana*. The formation of the six *Brassica* species within U's triangle model can be attributed to the natural duplication of three diploid species and their subsequent hybridization in pairs [8]. Genetic phenomena, such as genome shock and chromosomal rearrangement, occurring as a result of hybridization and duplication, contribute to an increased copy number of homologous genes [3–8]. Consequently, certain chromosome segments are lost while others are replicated multiple times. This intricate process provides a valuable model for investigating the phylogeny and evolution of conserved genes [3–8]. Here, we identified the homologous genes of the R2R3-MYB transcription factors of the S6 subfamily in Brassicaceae species by utilizing the latest reference genome information in conjunction with the BLAST model system, and we analyzed their serial duplication. This approach effectively addressed the issue of certain homologous genes being overlooked during BLASTP identification due to variations in protein sequences. In the present study, we conducted a systematic analysis to identify 31 homologous genes of R2R3-MYB transcription factors from the S6 subfamily, resulting in a total of 92 identified genes [26]. Interestingly, no homologous genes were detected in *A. arabicum*, whereas 10 homologous genes were found in *B. napus*, as inferred from the number of homologous genes (Table 1). With the exception of a limited number of homologous genes found in certain closely related species of *A. thaliana*, such as *A. lyrata*, the presence of homologous genes in other simple diploid species is typically restricted to 1–2. Conversely, the diploid species radish (*Raphanus sativus*), *B. rapa*, *B.oleracea*, and *B. nigra* exhibit four homologous

genes that have undergone whole genome triplication events, while the allotetraploid species *B. napus*, *B. juncea*, and *B. carinata* possess 7–10 homologous genes (Table 1). These results indicate that the number of homologous genes of R2R3-MYB transcription factors of the S6 subfamily is closely related to whole genome duplication events, the number of homologous genes increases significantly with genome polyploidization and hybridization, and the number also changes during later chromosome rearrangements [3–8]. Phylogenetic analysis revealed a noteworthy finding, approximately 70% of the homologous genes belonging to the S6 subfamily of R2R3-MYB transcription factors can be classified into a single group. Notably, all 35 homologous genes identified in *Brassica* crops were situated within this group, suggesting a high degree of conservation in the evolutionary trajectory of R2R3-MYB transcription factors within the S6 subfamily, which supports their ability to perform similar functions (Fig. 1). Previous studies have indicated that the A, B, and C subgenomes of *Brassica* are closely related evolutionarily, particularly the A and C subgenomes [5]. The collinearity analysis of homologous genes belonging to the S6 subfamily of R2R3-MYB transcription factors across six species of *Brassica* U's triangle provide further evidence for the strong collinearity observed among the chromosomes of *Brassica* crop subgenomes. The quantitative variation observed is primarily attributed to the tandem duplication of homologous genes on chromosomes A7 and C6 (Fig. 2). Furthermore, the analysis of conserved motifs and specific conserved domains of R2R3-MYB transcription factors reveals that three homologous genes, *BnaPAP1.C06.b*, *BjuPAP1.B05.a*, and *BnaPAP1.C06.c*, exhibit significant structural variations and are all located in tandem. Mutations occurring during the duplication process of tandemly repeated genes can give rise to novel gene variations (Fig. 3). This finding provides additional evidence for the high conservation of R2R3-MYB transcription factors among the S6 subfamily homologous genes, while also highlighting the generation of novel variations through tandem duplications. In particular, *Brassica* crops have experienced triploid events, natural hybridization, and natural doubling over an independent evolutionary timeline of 12 to 20 million years. These evolutionary occurrences have resulted in alterations to the anthocyanin transcriptional regulation mechanisms of *Brassica* crops when compared to *A. thaliana*. This is evidenced by an increase in gene copy number and functional differentiation, contributing to the complexity of the anthocyanin regulation mechanism. The functional divergence of these genes may contribute to the extensive color variation observed in *Brassica* crops. Our analysis aims to enhance the understanding of the phylogeny and evolution of Brassicaceae plants.

Functional differentiation of R2R3-MYB transcription factors of S6 subfamily genes in regulating anthocyanin biosynthesis

The involvement of R2R3-MYB transcription factors from the S6 subfamily, specifically the genes *PAP1*, *PAP2*, *MYB113*, and *MYB114* in the transcriptional regulation of plant anthocyanin biosynthesis in plants has been extensively documented across various species. These transcription factors play a crucial role in modulating color variation in different plant tissues and organs, including fleshy roots [39], tubers [40], leaves [32, 41], stems [42], flowers [42, 43], as well as peel and pulp [44]. This indicates that *PAP1*, *PAP2*, *MYB113*, and *MYB114* exhibit similar functionalities across diverse plant tissues or organs. Notably, within the six *Brassica* crops located in the U's triangle, namely *B. rapa*, *B. oleracea*, *B. juncea*, and *B. napus*, a wide array of organ-specific variations, including rhizomes, leaves, stems, flowers, and seeds. These variations are accompanied by a diversity of colors, which confer multiple functionalities such as oil production, edibility, and ornamental value. The precise accumulation of anthocyanins in various tissues and organs of *Brassica* crops is pivotal in determining their diverse coloration, with *PAP1* and its homologous genes playing a significant role in this process. In *B. rapa*, the homologous gene *BrMYB2*, homologous to *PAP1*, serves as a critical regulator of anthocyanin biosynthesis in purple leaves [45]. Similarly, in *B. oleracea*, distinct variations in the promoter region of the *PAP1* homolog *BoMYB2* are key determinants of color variation in ornamental kale, kohlrabi, cabbage, cauliflower, broccoli, Chinese kale, and Brussels sprout [46]. The regulation of anthocyanin-mediated leaf color variation in *B. juncea* is predominantly governed by the *MYB113* gene [41, 47]. Conversely, in *B. napus*, the presence of ten homologous genes of *PAP1* has facilitated the identification of distinct regulatory genes for anthocyanin production across different tissues. Specifically, *BnaPAP2.A7* serves as the primary regulatory factor in leaves and petals [32, 42, 43, 48], while *BnaPAP2.C6* exerts the main regulatory function in stems and siliques [49]. In this study, we utilized RNA-seq data to conduct a comprehensive analysis of the transcriptional regulation mechanisms underlying anthocyanin biosynthesis across various tissues and organs of *B. napus*. Interestingly, *BnaPAP1.A07.b*, along with its orthologous genes *BraPAP1.A07*, *BolPAP1.C06.b*, and *BjuPAP1.B05.b*, exhibited significantly higher expression in tissues associated with anthocyanin synthesis in *Brassica* (Fig. 4). In contrast, the expression levels of other gene copies varied among different materials. For instance, *BnaPAP1.C02* was highly expressed in both green and purple tissues, suggesting a potential differentiation in its function. Notably, *PAP1* is characterized

by tandem repeats on the A07 and C06 chromosomes of *B. napus*. However, expression pattern analyses across multiple tissues and various flower colors revealed that *BnaPAP2.A7b* and *BnaPAP2.C6.d* are highly expressed and likely drive the primary function, while other copies exhibit certain expression levels that may contribute to auxiliary functions (Fig. 4). These multiple gene copies may have undergone sub-functionalization or become silent during evolution, leading to altered expression patterns in the regulation of the anthocyanin synthesis pathway. Building on our previous research, we aimed to elucidate the transcriptional regulation mechanism of anthocyanin in different tissues and organs of *B. napus*, and to reveal the expression pattern and function of different copies of *PAP1* (Fig. 6). Notably, in leaves, the *BnaPAP2.A7b* transcriptional regulator collaborates with *TT8* and *TTG1* to form an MBW complex, which plays a crucial role in regulating anthocyanin biosynthesis (Fig. 6A). This regulatory mechanism bears a striking resemblance to that observed in petals (Fig. 6C). Furthermore, we found that the *BnaPAP2.C6* transcriptional regulator interacts with *TT8* and *TTG1*, forming an MBW complex that exerts significant control over anthocyanin biosynthesis regulation (Fig. 6B). This finding is consistent with the regulatory mechanism observed in siliques (Fig. 6D). These results indicate that the expression of different copies of *PAP1* is tissue-specific, and the functions of certain copies may vary. However, further exploration is needed to fully understand this tissue-specific regulatory mechanism and the associated functional variations.

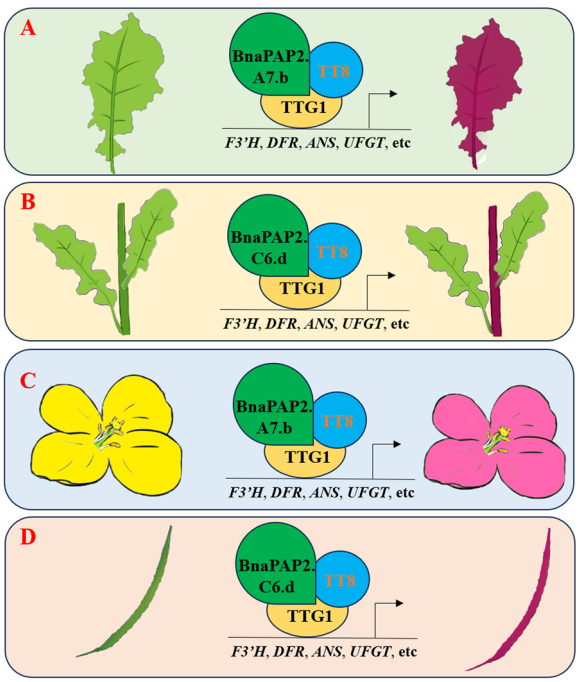

Fig. 6 Proposed transcriptional regulatory model for the regulation of anthocyanin accumulation in different tissues or organs of *B. napus*. The MBW transcriptional regulatory complex composed of *PAP2-TT8-TTG1* regulates the expression of anthocyanin synthesis pathway structural genes *F3'H*, *DFR*, *ANS*, *UFGT*, etc., thereby activating anthocyanin synthesis. **A** Leaves; **B** Stems; **C** Flowers; D: Siliques.11

Conclusion

In this study, we systematically identified homologous genes of the S6 subfamily of R2R3-MYB transcription factors across 31 Brassicaceae species, resulting in the identification of a total of 92 homologous genes, followed by a phylogenetic analysis. We then focused on the analysis of 35 homologous genes from the S6 subfamily of R2R3-MYB transcription factors in six *Brassica* species, examining the promoter *cis*-acting elements, gene structure, and conserved domains. Through collinearity analysis, we clarified the distribution and collinearity of these 35 genes within the A, B, and C subgenomes. Additionally, we investigated the expression patterns of these 35 homologous genes by integrating various leaf transcriptome datasets. Finally, we utilized transcriptome data from different tissues and organs of *B. napus* to analyze the expression patterns of 10 S6 subfamily R2R3-MYB transcription factors in *B. napus*. Building upon previous studies, we proposed a transcriptional regulation mechanism for anthocyanin mediated by *PAP1* homologous genes across different tissues and organs of *B. napus*. In the future, we will combine overexpression, gene editing and DNA affinity purification sequencing (DAP-seq) to further explore. Consequently, our findings provide novel insights into the anthocyanin transcriptional regulatory network of *Brassica* crops, highlighting the functional diversification of S6 subfamily R2R3-MYB transcription factors. These results will contribute new knowledge regarding anthocyanin-mediated traits, such as stress resistance, disease resistance, and nutritional value in *Brassica* crops. Furthermore, our research establishes a theoretical foundation for developing new *Brassica* varieties with enhanced anthocyanin content, alongside advancing molecular marker-assisted breeding for varieties enriched with anthocyanins, stress resistance, and disease resistance. Additionally, while further experimental validation is necessary, these findings enhance our understanding of the transcriptional regulation mechanisms of cross-species anthocyanins and serve as a reference for analyzing the molecular mechanisms underlying other plant traits.

Abbreviations

PAL	Phenylalanine ammonia-lyase
C4H	Cinnamate-4-hydroxylase
4CL	4-Coumarate CoA ligase 4

CHS	Chalcone synthase
CHI	Chalcone isomerase
F3H	Flavanone 3-hydroxylase
FLS	Flavonol synthase
DFR	Dihydroflavonol 4-reductase
ANS	Anthocyanidin synthase
BAN	Anthocyanidin oxidoreductase
AT	Acyltransferase
GT	Glucosyltransferase
UFGT	UDP-flavonoid glucosyltransferase
TTG1	TRANSPARENT TESTA GLABRA 1
TT8	TRANSPARENT TESTA 8
GL3	GLABRA 3
EGL3	ENHANCER OF GLABRA 3
PAP1/2	PRODUCTION OF ANTHOCYANIN PIGMENT ½
MYB75/90/113/114	MYB DOMAIN PROTEIN 75/90/113/114

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06296-1.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Supplementary Material 5.

Supplementary Material 6.

Supplementary Material 7.

Acknowledgements
Thanks to Professor Xianhong Ge and Cheng Cui for his help in the writing process of the thesis, and thanks to Bioyi Biotechnology Co., Ltd., Wuhan, China, for the sequencing service provided.

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

Authors' contributions
DC and CT conceived and designed the experiments. YL, DZ and LC performed the experiments. CW, QC and WS analyzed the data. DC, BZ and CT wrote the manuscript. All authors approved the manuscript and consent to publication this manuscript.

Funding
This research was funded by the Ganpo Talent Support Program of Jiangxi Province, grant number: 20232BCJ23089; the Key Research and Development Program of Jiangxi Province, grant number: 20223BBF61002; Jiangxi Province's training program for academic and technical leaders in major disciplines, grant number: 20204BCJL22046.

Data availability
All RNA-seq data in this study were downloaded from the NCBI (https://www.ncbi.nlm.nih.gov/), with biological projects PRJNA554517 (leaves), PRJAN855492 (stems), PRJNA1054216 and PRJAN855492 (flower petals), PRJNA734925 (siliques), and PRJNA597958 (seeds).

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 no competing interests.

Author details
[1]College of Life Sciences, Ganzhou Key Laboratory of Greenhouse Vegetable, Gannan Normal University, Ganzhou 341000, China. [2]Nanchang Branch of National Center of Oilcrops Improvement, Jiangxi Province Key Laboratory of Oil Crops Biology, Crops Research Institute of Jiangxi Academy of Agricultural Sciences, Nanchang 330200, China. [3]Hunan Key Laboratory of Economic Crops Genetic Improvement and Integrated Utilization, School of Life and Health Science, Hunan University of Science and Technology, Xiangtan 411201, China.

Received: 11 January 2024 Accepted: 24 February 2025
Published online: 06 March 2025

References

1. Town CD, Cheung F, Maiti R, Crabtree J, Haas BJ, Wortman JR, Hine EE, Althoff R, Arbogast TS, Tallon LJ, et al. Comparative genomics of Brassica oleracea and Arabidopsis thaliana reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell. 2006;18(6):1348–59.
2. Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S. Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. Proc Natl Acad Sci U S A. 2010;107(43):18724–8.
3. Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, et al. The genome of the mesopolyploid crop species Brassica rapa. Nat Genet. 2011;43(10):1035–9.
4. Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al. Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science. 2014;345(6199):950–3.
5. Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IA, Zhao M, Ma J, Yu J, Huang S, et al. The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun. 2014;5:3930.
6. Perumal S, Koh CS, Jin L, Buchwaldt M, Higgins EE, Zheng C, Sankoff D, Robinson SJ, Kagale S, Navabi ZK, et al. A high-contiguity Brassica nigra genome localizes active centromeres and defines the ancestral Brassica genome. Nat Plants. 2020;6(8):929–41.
7. Kang L, Qian LW, Zheng M, Chen LY, Chen H, Yang L, You L, Yang B, Yan ML, Gu YG, et al. Genomic insights into the origin, domestication and diversification of. Nat Genet. 2021;53(9):1392.
8. Song X, Wei Y, Xiao D, Gong K, Sun P, Ren Y, Yuan J, Wu T, Yang Q, Li X, et al. Brassica carinata genome characterization clarifies U's triangle model of evolution and polyploidy in Brassica. Plant Physiol. 2021;186(1):388–406.
9. Stracke R, Werber M, Weisshaar B. The R2R3-MYB gene family in Arabidopsis thaliana. Curr Opin Plant Biol. 2001;4(5):447–56.
10. Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L. MYB transcription factors in Arabidopsis. Trends Plant Sci. 2010;15(10):573–81.
11. Feller A, Machemer K, Braun EL, Grotewold E. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. Plant J. 2011;66(1):94–116.
12. Liu J, Osbourn A, Ma P. MYB transcription factors as regulators of phenylpropanoid metabolism in plants. Mol Plant. 2015;8(5):689–708.
13. Wang X, Niu Y, Zheng Y: Multiple Functions of MYB Transcription Factors in Abiotic Stress Responses. Int J Mol Sci 2021, 22(11).
14. Li C, Yu W, Xu J, Lu X, Liu Y: Anthocyanin Biosynthesis Induced by MYB Transcription Factors in Plants. Int J Mol Sci 2022, 23(19).
15. Wang Y, Zhou H, He Y, Shen X, Lin S, Huang L. MYB transcription factors and their roles in the male reproductive development of flowering plants. Plant Sci. 2023;335:111811.
16. Wu Y, Wen J, Xia Y, Zhang L, Du H. Evolution and functional diversification of R2R3-MYB transcription factors in plants. Hortic Res. 2022;9:uhac058.
17. Zhao L, Gao L, Wang H, Chen X, Wang Y, Yang H, Wei C, Wan X, Xia T. The R2R3-MYB, bHLH, WD40, and related transcription factors in flavonoid biosynthesis. Funct Integr Genomics. 2013;13(1):75–98.

18. Shelton D, Stranne M, Mikkelsen L, Pakseresht N, Welham T, Hiraka H, Tabata S, Sato S, Paquette S, Wang TL, et al. Transcription factors of Lotus: regulation of isoflavonoid biosynthesis requires coordinated changes in transcription factor activity. Plant Physiol. 2012;159(2):531–47.
19. Tohge T, de Souza LP, Fernie AR. Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. J Exp Bot. 2017;68(15):4013–28.
20. Zhang N, Jing P. Anthocyanins in Brassicaceae: composition, stability, bioavailability, and potential health benefits. Crit Rev Food Sci Nutr. 2022;62(8):2205–20.
21. Shi MZ, Xie DY. Biosynthesis and metabolic engineering of anthocyanins in Arabidopsis thaliana. Recent Pat Biotechnol. 2014;8(1):47–60.
22. Zhang Y, Butelli E, Martin C. Engineering anthocyanin biosynthesis in plants. Curr Opin Plant Biol. 2014;19:81–90.
23. Sunil L, Shetty NP. Biosynthesis and regulation of anthocyanin pathway genes. Appl Microbiol Biotechnol. 2022;106(5–6):1783–98.
24. Xu WJ, Grain D, Bobet S, Le Gourrierec J, Thévenin J, Kelemen Z, Lepiniec L, Dubos C. 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. 2014;202(1):132–44.
25. Quattrocchio F, Wing J, van der Woude K, Souer E, de Vetten N, Mol J, Koes R. Molecular analysis of the anthocyanin2 gene of petunia and its role in the evolution of flower color. Plant Cell. 1999;11(8):1433–44.
26. Chen D, Chen H, Dai G, Zhang H, Liu Y, Shen W, Zhu B, Cui C, Tan C. Genome-wide identification of R2R3-MYB gene family and association with anthocyanin biosynthesis in Brassica species. BMC Genomics. 2022;23(1):441.
27. Pucker B. Automatic identification and annotation of MYB gene family members in plants. BMC Genomics. 2022;23(1):220.
28. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37(5):1530–4.
29. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics. 2019;35(21):4453–5.
30. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293–6.
31. Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y, et al. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. Mol Plant. 2023;16(11):1733–42.
32. Chen D, Liu Y, Yin S, Qiu J, Jin Q, King GJ, Wang J, Ge X, Li Z. Alternatively Spliced BnaPAP2.A7 Isoforms Play Opposing Roles in Anthocyanin Biosynthesis of Brassica napus L. Front Plant Sci. 2020;11:983.
33. He Z, Ji R, Havlickova L, Wang L, Li Y, Lee HT, Song J, Koh C, Yang J, Zhang M, et al. Genome structural evolution in Brassica crops. Nat Plants. 2021;7(6):757–65.
34. Song XM, Wang JP, Sun PC, Ma X, Yang QH, Hu JJ, Sun SR, Li YX, Yu JG, Feng SY, et al. Preferential gene retention increases the robustness of cold regulation in Brassicaceae and other plants after polyploidization. Hortic Res. 2020;7:20.
35. Mandakova T, Hlouskova P, German DA, Lysak MA. Monophyletic Origin and Evolution of the Largest Crucifer Genomes. Plant Physiol. 2017;174(4):2062–71.
36. Nikolov LA, Shushkov P, Nevado B, Gan X, Al-Shehbaz IA, Filatov D, Bailey CD, Tsiantis M. Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. New Phytol. 2019;222(3):1638–51.
37. Arabidopsis Genome I. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature. 2000;408(6814):796–815.
38. Liu Z, Li N, Yu T, Wang Z, Wang J, Ren J, He J, Huang Y, Shi K, Yang Q, et al. The Brassicaceae genome resource (TBGR): A comprehensive genome platform for Brassicaceae plants. Plant Physiol. 2022;190(1):226–37.
39. Tao J, Li S, Wang Q, Yuan Y, Ma J, Xu M, Yang Y, Zhang C, Chen L, Sun Y: Construction of a high-density genetic map based on specific-locus amplified fragment sequencing and identification of loci controlling anthocyanin pigmentation in Yunnan red radish. Hortic Res 2022;9.
40. Jung CS, Griffiths HM, De Jong DM, Cheng S, Bodis M, Kim TS, De Jong WS. The potato developer (D) locus encodes an R2R3 MYB transcription factor that regulates expression of multiple anthocyanin structural genes in tuber skin. Theor Appl Genet. 2009;120(1):45–57.
41. Heng S, Cheng Q, Zhang T, Liu X, Huang H, Yao P, Liu Z, Wan Z, Fu T. Fine-mapping of the BjPur gene for purple leaf color in Brassica juncea. Theor Appl Genet. 2020;133(11):2989–3000.
42. Chen D, Jin Q, Pan J, Liu Y, Tang Y, E Y, Xu L, Yang T, Qiu J, Chen X et al: Fine mapping of genes controlling pigment accumulation in oilseed rape (Brassica napus L.). Mol Breed 2023;43(3):19.
43. Ye SH, Hua SJ, Ma TT, Ma XW, Chen YP, Wu LM, Zhao L, Yi B, Ma CZ, Tu JX, et al. Genetic and multi-omics analyses reveal BnaA07. PAP2 as the key gene conferring anthocyanin-based color in Brassica napus flowers. J Exp Bot. 2022;73(19):6630–45.
44. Zhang S, Wang H, Wang T, Liu W, Zhang J, Fang H, Zhang Z, Peng F, Chen X, Wang N. MdMYB305-MdbHLH33-MdMYB10 regulates sugar and anthocyanin balance in red-fleshed apple fruits. Plant J. 2023;113(5):1062–79.
45. He Q, Wu J, Xue Y, Zhao W, Li R, Zhang L. The novel gene BrMYB2, located on chromosome A07, with a short intron 1 controls the purple-head trait of Chinese cabbage (Brassica rapa L.). Hortic Res. 2020;7:97.
46. Yan C, An G, Zhu T, Zhang W, Zhang L, Peng L, Chen J, Kuang H. Independent activation of the BoMYB2 gene leading to purple traits in Brassica oleracea. Theor Appl Genet. 2019;132(4):895–906.
47. An GH, Chen JJ: Frequent gain- and loss-of-function mutations of the gene accounted for leaf color variation in. Bmc Plant Biol 2021;21(1).
48. Fu H, Chao H, Zhao X, Wang H, Li H, Zhao W, Sun T, Li M, Huang J. Anthocyanins identification and transcriptional regulation of anthocyanin biosynthesis in purple Brassica napus. Plant Mol Biol. 2022;110(1–2):53–68.
49. Chen D, Chen H, Dai G, Zhang H, Liu Y, Shen W, Zhu B, Cui C, Tan C. Genome-wide identification and expression analysis of the anthocyanin-related genes during seed coat development in six Brassica species. BMC Genomics. 2023;24(1):103.

Publisher's Note