

RESEARCH

Open Access



# Genome-wide identification and expression analysis of the anthocyanin-related genes during seed coat development in six *Brassica* species

Daozong Chen<sup>1</sup>, Haidong Chen<sup>1</sup>, Guoqiang Dai<sup>1</sup>, Haimei Zhang<sup>1</sup>, Yi Liu<sup>1</sup>, Wenjie Shen<sup>1</sup>, Bo Zhu<sup>1\*</sup>, Cheng Cui<sup>2\*</sup> and Chen Tan<sup>1\*</sup>

## Abstract

Yellow seed is one favorite trait for the breeding of *Brassica* oilseed crops, but the performance of seed coat color is very complicated due to the involvement of various pigments. The change of seed coat color of *Brassica* crops is related to the specific synthesis and accumulation of anthocyanin, and the expression level of structural genes in anthocyanin synthesis pathway is specifically regulated by transcription factors. Despite some previous reports on the regulations of seed coat color from linkage marker development, gene fine-mapping and multi-omics association analysis, the trait of *Brassica* crops is affected by the evolutionary events such as genome triploidization, the regulatory mechanism is still largely unknown. In this study, we identified genes related to anthocyanin synthesis in six *Brassica* crops in U-triangle at the genome-wide level and performed collinearity analysis. A total of 1119 anthocyanin-related genes were identified, the collinear relationship of anthocyanin-related genes on subgenomic chromosomes was the best in *B. napus* (AACC) and the worst in *B. carinata* (BBCC). The comparisons of gene expressions for anthocyanin metabolic pathways in seed coats during seed development revealed differences in its metabolism among these species. Interestingly, the R2R3-MYB transcription factors *MYB5* and *TT2* were differentially expressed at all eight stages of seed coat development, indicating that they might be the key genes that caused the variation of the seed coat color. The expression curve and trend analyses of the seed coat development period showed that the main reason for the unexpressed copies of *MYB5* and *TT2* was likely gene silencing caused by gene structural variation. These results were valuable for the genetic improvement of *Brassica* seed coat color, and also provided new insights into gene multicopy evolution in *Brassica* polyploids.

**Keywords** *Brassica*, Seed coat color, Anthocyanin-related genes, *MYB5*, *TT2*

\*Correspondence:

Bo Zhu

nczb615@163.com

Cheng Cui

cuicheng005@163.com

Chen Tan

tanchen2020@gnnu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Introduction

The *Brassica* species in U's triangle include three diploid crops, *B. rapa* (AA, 2n=20), *B. nigra* (BB, 2n=16), *B. oleracea* (CC, 2n=18) and three allotetraploid crops, *B. juncea* (AABB, 2n=36), *B. napus* (AACC, 2n=38), *B. carinata* (BBCC, 2n=34), formed by natural pairwise hybridizations of three diploids (U, 1935). The *Brassica* species comprise the important oil crops and vegetables and are widely cultivated worldwide for vegetable and oil acquisition. Generally, the seeds of these *Brassica* species vary in color from yellow, brown to dark brown or black [45, 69]. Previously, yellow seed was a desirable /+trait with great potential to improve the quality of seed oil in *Brassica* crops, because yellow seeds have the thinner coats, a reduced percentage of pigment and hull, and a greater content of oil and protein than the black seeds [68, 38, 40, 56]. In recent years, with these superior characteristics, the yellow seed coat trait of *B. rapa* [61, 81, 82], *B. napus* [20, 21, 29, 30, 79, 86] and *B. juncea* [50, 64] has become the focus of global rapeseed breeding.

In nature, varieties with different seed colors have different seed coat pigment components. In *Arabidopsis* and *Brassica* crops, kaempferol derivatives, quercetin derivatives, isorhamnetin derivatives and epicatechin glycosides are the main pigments of seeds [31, 61, 62, 64, 79]. Importantly, the differences in the content of flavonoids, especially proanthocyanidins (PAs), are the main substances that cause color variation of seed coat color [27, 46, 55, 79, 82]. In the model plant *Arabidopsis thaliana*, the biosynthetic pathway of anthocyanins has been clearly reported. From phenylpropane synthesis pathway (*PAL*, *CAH*, *4CL*), flavonoid synthesis pathway (*CHS*, *CHI*, *F3H/F3'H/F3'5'H*) to anthocyanin synthesis pathway (*DFR*, *ANS*, *UFGT*, etc.), these series of synthases are regulated by three types of genes: MYB transcription factor, bHLH transcription factor and WD-40 protein [53, 76]. In addition, the MBW transcriptional regulatory complexes TT2-TT8-TTG1, MYB5-TT8-TTG1, TT2-EGL3-TTG1, and TT2-GL3-TTG1, composed of three types of transcription factors, were reported to be involved in the synthesis of anthocyanins in the seed coat [76].

For decades, genes that control seed coat color have been studied in different *Brassica* species cultivars collected around the world, by exploiting QTL mapping, resequencing analysis, transcriptome analysis, transcriptome combined metabolome analysis, etc. Numerous studies have revealed that the seed color of *Arabidopsis* and *Brassica* crops was controlled by few major QTLs [50, 57, 74, 76]. Among them, several homologs to *Arabidopsis* transcriptional regulators affecting seed color, such as *MYB11*, *MYB12*, *MYB111*, *TT2*, *TT8*, *GL3*, *EGL3* have been identified, and some of them have been functionally validated in *B. rapa*, *B. juncea* and *B. napus* [64,

79, 82]. Nevertheless, due to differences in genetic background of materials and lack of systematic data, the exact underlying regulatory mechanism of seed coat color remains unclear [28, 50, 64, 72, 75, 79, 80, 82].

In this study, we identified genes related to anthocyanin synthesis in *Brassica* crops at the genome-wide level and performed collinearity analysis. In addition, we collected the transcriptome sequencing data of the seed coats at eight stages of seed development in six *Brassica* species, and we analyzed the anthocyanin-related differentially expressed genes and their expression levels in different developmental stages of the seed coat. Additionally, *MYB5* and *TT2* were differentially expressed at all eight stages of seed coats in the six *Brassica* species. Our findings lay a foundation for elucidating the regulatory mechanisms of seed coat color in *Brassica* species and provide important resources for a comparative and integrated analysis of seed coat color trait in these interrelated *Brassica* species.

## Materials and methods

### Identification of anthocyanin-related genes in *Brassica* species

In this study, the genome and protein sequences of the *B. rapa* (Chiifu-401-42 v3.0), *B. oleracea* (HDEM), *B. nigra* (Ni100-LR), *B. napus* (Darmor-bzh v10) were downloaded from the BRAD database [7]; <http://brassicadb.cn>), *B. juncea* (SCYZ) genome sequence from NCBI PRJNA615316 [22] (<https://www.ncbi.nlm.nih.gov/>), *B. carinata* (zd-1) genome sequence from GenBank JAAMPC000000000 [66] (<https://www.ncbi.nlm.nih.gov/>), and the anthocyanin-related genes genome and protein sequences were downloaded from the *Arabidopsis* database (TAIR; <http://www.arabidopsis.org/index.jsp>). In order to accurately identify anthocyanin-related genes, we mainly divide it into the following steps: Firstly, local BLASTP has been used to search anthocyanin-related genes with E-value < 1e-20, 55 anthocyanin-related genes protein sequences were derived from *Arabidopsis*. Secondly, the candidate anthocyanin-related genes in the six *Brassica* species were identified by a local BLASTN search with 55 anthocyanin-related genes coding sequence from *Arabidopsis* to identify candidates with E-value < 1e-20, identity > 70%, coverage > 60%. Thirdly, SynOrths software [8] has been used to determine the collinear orthologous of two genes based on their own sequence similarity and the homology of their flanking genes, and then extracting the collinear genes of anthocyanin-related genes. Finally, the BLASTP, BLASTN and SynOrths software identified results were pooled and deduplicated, and determined in conjunction with PFAM protein family database (<https://pfam.xfam.org/>).

### 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 anthocyanin-related genes in *Arabidopsis* and six *Brassica* species [4]. 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) [71], SynOrths software (E-value < 1e-20, Query gene = 20, Reference gene = 100) has been used to determine the collinear orthologous [8], TBtools software (v0.67) was used to drop the collinearity genes on each chromosome [4].

### Expression profiles analysis based on RNA-seq data

The raw data of 144 seed coats RNA-seq data of six *Brassica* species, *B. rapa* (Parkland-R), *B. oleracea* (Chinese Kale-O), *B. nigra* (CR2748-N), *B. napus* (DH12075-P), *B. juncea* (AC Vulcan-I), *B. carinata* (C901163-C) with eight developmental stages (Unfertilized ovule integuments (UO; no embryo), 1- to 2-cell zygote stage (S1), 4- to 8-cell stage (S2, 8-cell stage shown), 16- to 64-cell stage (S3, globular stage shown), heart stage (S4), torpedo stage (S5), bent stage (S6), and mature (S7) stage of seed formation) were collected from Gene Expression Omnibus under accession no. GSE153257. Low-quality reads were removed from the raw reads using Cutadapt and Trimmomatic software to get clean reads [39, 2]. Clean reads were mapped to the corresponding reference genome using HISAT2 software [51]. Gene expression levels of each gene were calculated using StringTie and Ballgown software [51]. The read counts of each gene were calculated using the htseq-count function in htseq software [1]. 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:  $\text{padj} < 0.05$  &  $\log_2\text{FoldChange} > 2$  [5].

### Co-differential expression analysis of anthocyanin-related genes

The RNA-seq data were used to perform co-expression network analysis using R language (v4.2.1). 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 anthocyanin-related genes in the matrix. STRING software (<https://version-11-5.string-db.org/>) was used to reveal a co-expression plot [33, 70].

### Phylogenetic, promoter characteristics, gene structure, conserved motifs analysis of TT2 and MYB5

The *TT2* and *MYB5* protein sequences of the six *Brassica* species and *Arabidopsis* were used to generate phylogenetic trees via ClustalX [26] 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 [52]. The *TT2* and *MYB5* 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 *TT2* and *MYB5* were analyzed according to the GFF annotation file of the gene position information in the six *Brassica* crops and *Arabidopsis* database. 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 SWISS-MODEL online tool (<https://swissmodel.expasy.org/>) was used to draw spatial structure. TBtools software (v0.67) was used to draw the *TT2* and *MYB5* to the different copies of each *Brassica* species, including phylogenetic, promoter characteristics, gene structure, conserved motifs [4].

### Construction of expression profiles of TT2 and MYB5 during seed coat development based on RNA-seq data

Transcriptome data of all eight periods of seed coat development of six *Brassica* species were used to analyze the expression patterns and trends of *TT2* and *MYB5*. The R package ggplot2 (v3.3.6) was used to draw the expression pattern of *TT2* and *MYB5*, and R package ggridges (v0.5.3) was used to draw the expression trends of *TT2* and *MYB5*.

## Results

### Identification and characterization of anthocyanin-related genes

In order to comprehensively and accurately identify anthocyanin-related genes, the 55 anthocyanin-related genes protein sequences of *Arabidopsis* were downloaded and used as seed sequences to search the protein sequences database of *Arabidopsis* and six *Brassica* species to identify homologous anthocyanin-related genes in six *Brassica* crops. The BLASTP (protein to protein), BLASTN (CDS to CDS) and SynOrths software (protein to protein, orthologous gene identification) identified results were pooled and deduplicated, and then the

orthologous and paralogous genes determined by PFAM. A total of 1119 anthocyanin-related genes were identified, these included 120 in *B. rapa*, 131 in *B. oleracea*, 135 in *B. nigra*, 258 in *B. napus*, 229 in *B. juncea* and 246 in *B. carinata* (Table 1, Supplementary Table S1). Among the three subgenomes of A, B, and C, the C subgenome had the most distribution and the A subgenome the least (Table 1).

#### Chromosomal distribution and collinearity analysis of anthocyanin-related genes in six *Brassica* species

To better understand the copy number variation and collinearity of anthocyanins related genes, we mapped all 1174 genes to the genome chromosomes of 7 corresponding species. Chromosome mapping showed that they were located on all chromosomes of each species: 55 genes on 5 chromosomes of *Arabidopsis*, 120 genes on 10 chromosomes of *B. rapa*, 131 genes of *B. oleracea* on 9 chromosomes, 135 genes of *B. nigra* on 8 chromosomes, 258 genes of *B. napus* on 19 chromosomes, 229 genes of *B. juncea* on 18 chromosomes, and 246 genes of *B. carinata* on 17 chromosomes (Fig. 1). Chromosomal distributions showed that these genes were evenly distributed in the three subgenomes of the six *Brassica* crops, and the copy number of the tetraploids was about the sum of the corresponding subgenomes (Table 1).

The interspecific collinearity between *Arabidopsis thaliana* and six *Brassica* species was also analyzed to further explore the evolution of anthocyanin-related genes (Fig. 1). At the genome-wide level, there was a one-to-one collinear relationship between the copies of these genes in the A subgenomes from *B. rapa* (BraA) and *B. napus* (BnaA), *B. oleracea* (BolC) and *B. napus* (BnaC). Interestingly, in the subgenomes between *B. rapa* (BraA) and *B. juncea* (BjuA), *B. nigra* (BniB) and *B. juncea* (BjuB), in spite of the better collinearity, there were similar inversions between most chromosome arms. In addition, in the subgenomes between *B. nigra* (BniB) and *B. carinata* (BcaB), *B. oleracea* (BolC) and *B. carinata* (BcaC),

the collinearity between different copies of the genes was poor, indicating that there might be a large number of rearrangement events at the chromosomal level after the formation of *B. carinata* (Fig. 1).

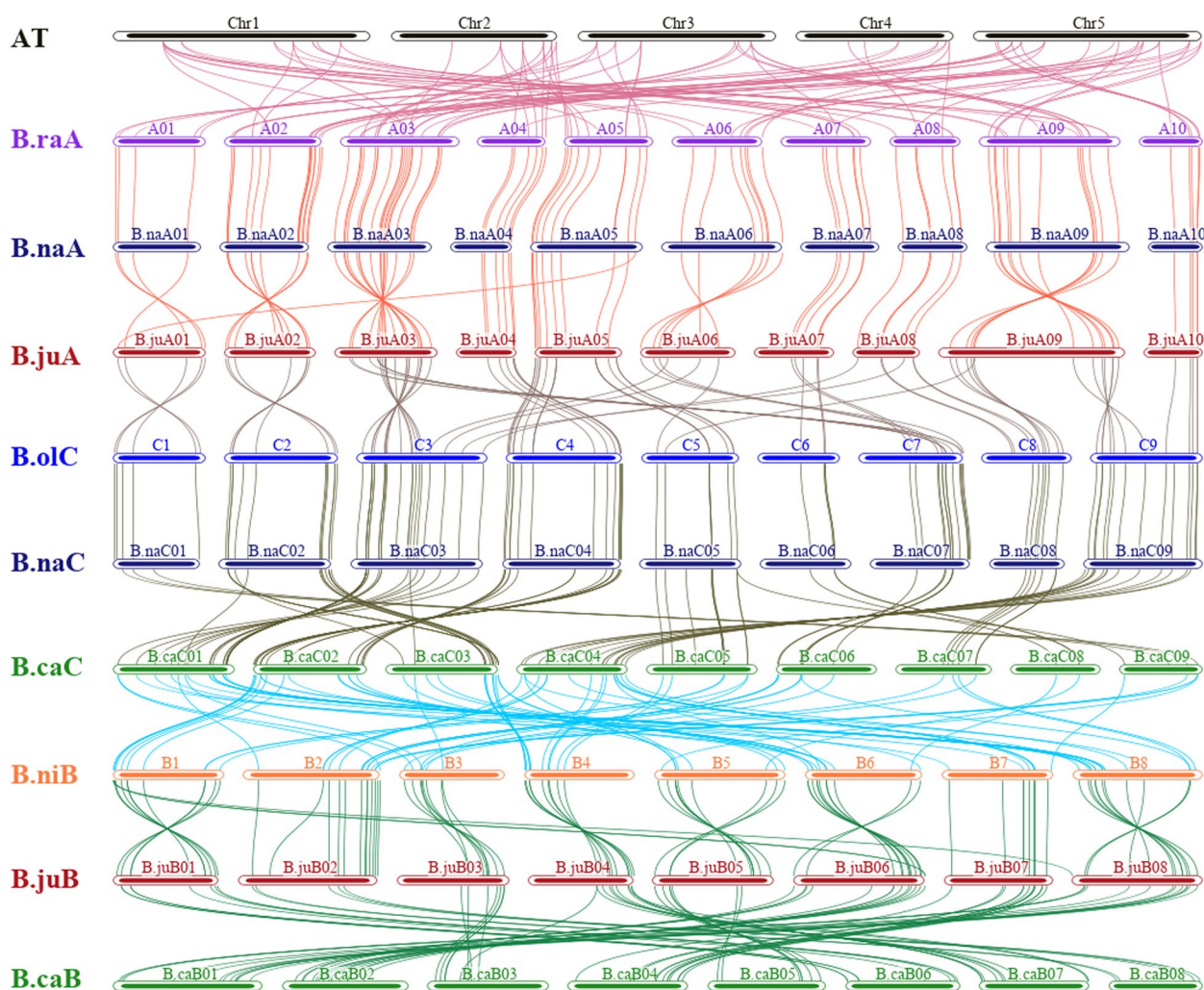
#### Analysis of the transcriptome and anthocyanin-related gene expression patterns

In order to explore the molecular mechanism of anthocyanin biosynthetic pathway-related genes regulating seed coat color formation during seed development, transcriptome data at eight periods (U0, S1, S2, S3, S4, S5, S6, S7) of seed coat development of six *Brassica* species were used to construct gene expression profiles (Fig. 2). In order to better show the expression patterns, the expression data of U0-S7 in a total of 8 developmental stages were divided into 4 stages, namely U0-S1, S2, S3-S5 and S6-S7 (Fig. 2). The results showed that the genes related to the anthocyanin synthesis pathway had high expression levels in all two stages, S3-S5 and S6-S7, which indicated that the synthesis of anthocyanins in the seed coat might mainly start from the S3 period (Fig. 2). In general, the expression levels of structural genes were higher than those of transcription factors *MYB5* and *TT2* in each developmental stage, especially in the S3-S7 developmental stage. Similar to previous research, glycosyltransferases (*GST*) were expressed at high levels in all developmental stages, indicating that glycosylation might be the main modification mode of anthocyanin color development. Interestingly, *DFR* and *ANS* (*LDOX*) genes were scarcely expressed in all developmental stages of *B. juncea*, which was quite different from the other five *Brassica* species.

Tetraploidy origin of *Brassica* crops brought abundant homologous gene copies. In order to understand how subgenome homoeologs were coordinately expressed during anthocyanin biosynthesis, we identified scenarios by comparisons among three tetraploid species. In general, the expression dominance of BnaA subgenome was significant over the BnaC subgenome

**Table 1** Anthocyanin-related genes distribution in U's triangle *Brassica* species

Species	Referenc_genome	Gene number				
		AA	BB	CC	unknow	Total
<i>B.rapa</i>	Chiifu3.0	120			0	120
<i>B.nigra</i>	Ni100-LR		135		0	135
<i>B.oleracea</i>	HDEM			131	0	131
<i>B.juncea</i>	SCYZ	105	123		1	229
<i>B.napus</i>	Darmor-bzh v10	116		141	1	258
<i>B.carinata</i>	Zd-1		99	122	24	246
total		341	357	394	26	1119



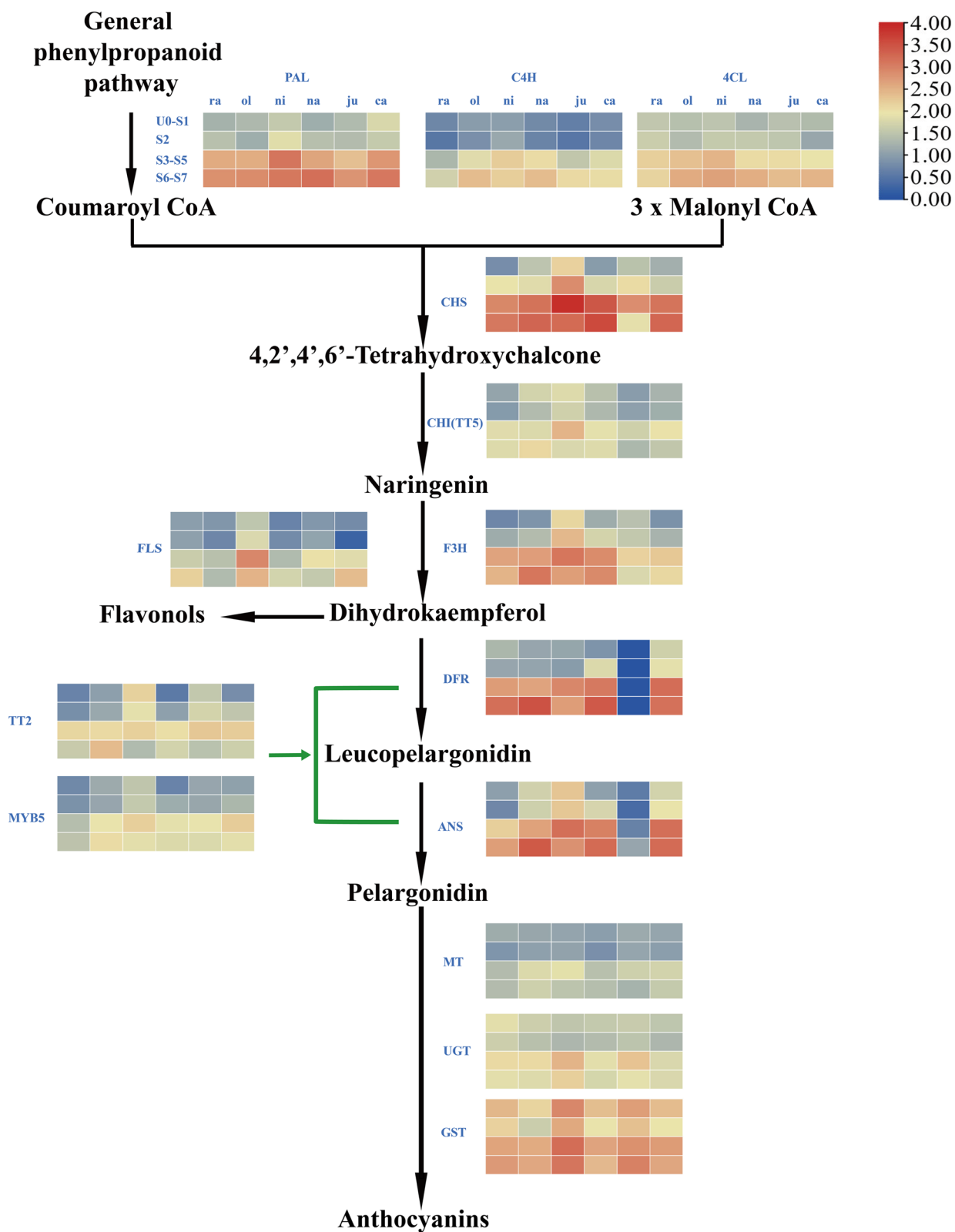
**Fig. 1** Interspecies synteny of anthocyanin-related genes between the six species in U's triangle and *Arabidopsis*. anthocyanin-related genes colinearity of the genomes of the six *Brassica* species in U's triangle and *Arabidopsis*, including three diploid species, *B. rapa* (A genome, BraA), *B. 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)

in *B. napus*, the expression dominance of BjuB subgenome was obvious over the BjuA subgenome in *B. juncea*, but the expression dominance of BcaC subgenome was over the BcaB subgenome in *B. carinata* (Fig. 3). Unlike the pattern of expression bias at the subgenomic level (C > B > A) reported by Tan et al. [67], we did not find a significantly elevated number of dominant features in any subgenome in the three allotetraploid species (Fig. 3), suggesting a different transcriptional regulation of biased and dominant expression, which was consistent with the report by Gao et al. [13].

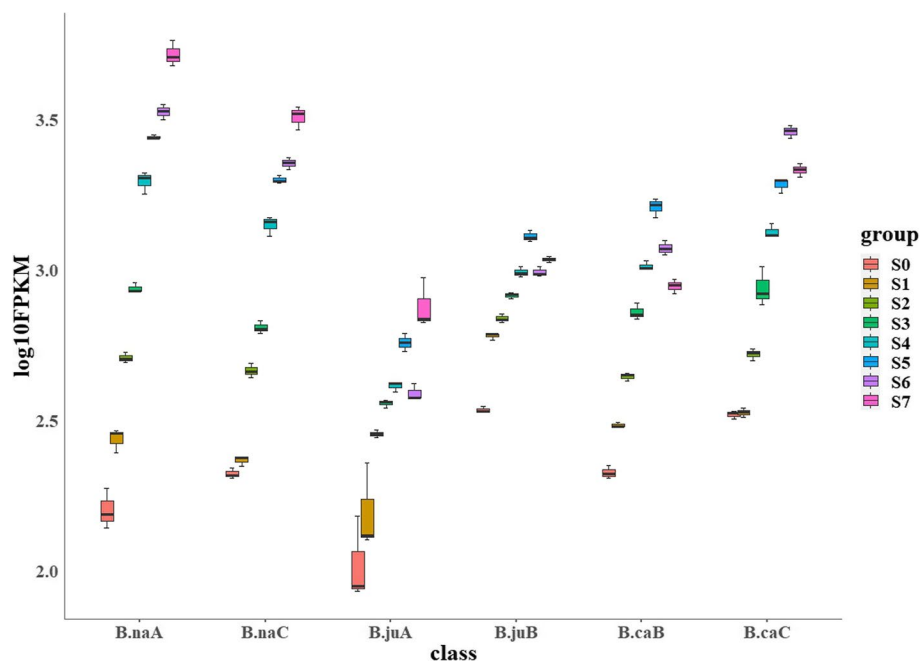
#### Co-expression analysis of anthocyanin-related genes

In order to identify the key genes that regulated the formation of seed coat color during seed coat

development, the transcriptome data of seed coats at 8 stages of six *Brassica* species were used for differentially expressed gene (DEG) analysis. Among the anthocyanin-related DEGs, a total of 95 genes were differentially expressed in the transcriptome data of all eight developmental stages, with 9, 23, 13, 19, 12, 19 DEGs from *B. rapa*, *B. nigra*, *B. oleracea*, *B. napus*, *B. juncea* and *B. carinata*, respectively (Supplementary Table S2). Interestingly, only *MYB5* (*AT3G13540*) and *TT2* (*AT5G35550*) homologous copies were differentially expressed in all eight stages of six *Brassica* crops (Fig. 4, Supplementary Table S2). This was different from the results we discovered that only *PAP1/2* homologous copies were differentially expressed in green and purple leaves of five *Brassica* crops [5], indicating the



**Fig. 2** Anthocyanin-related gene expression patterns of six U's triangle *Brassica* species in eight periods of seed coat development



**Fig. 3** Phylogenetic, gene structural analysis and conserved motif identification of *MYB5* and *TT2*. *MYB5* and *TT2* mainly contain nine protein conserved domains, which are marked with different colors. Predicted cis-elements in *MYB5* and *TT2* promoters and gene structure. Promoter sequences (~2,000 bp) of *MYB5* and *TT2* were analyzed using PlantCARE, different shapes and colors represent different elements. Gene regions are marked with black lines and green modules, respectively, with black lines representing introns and green modules representing exons

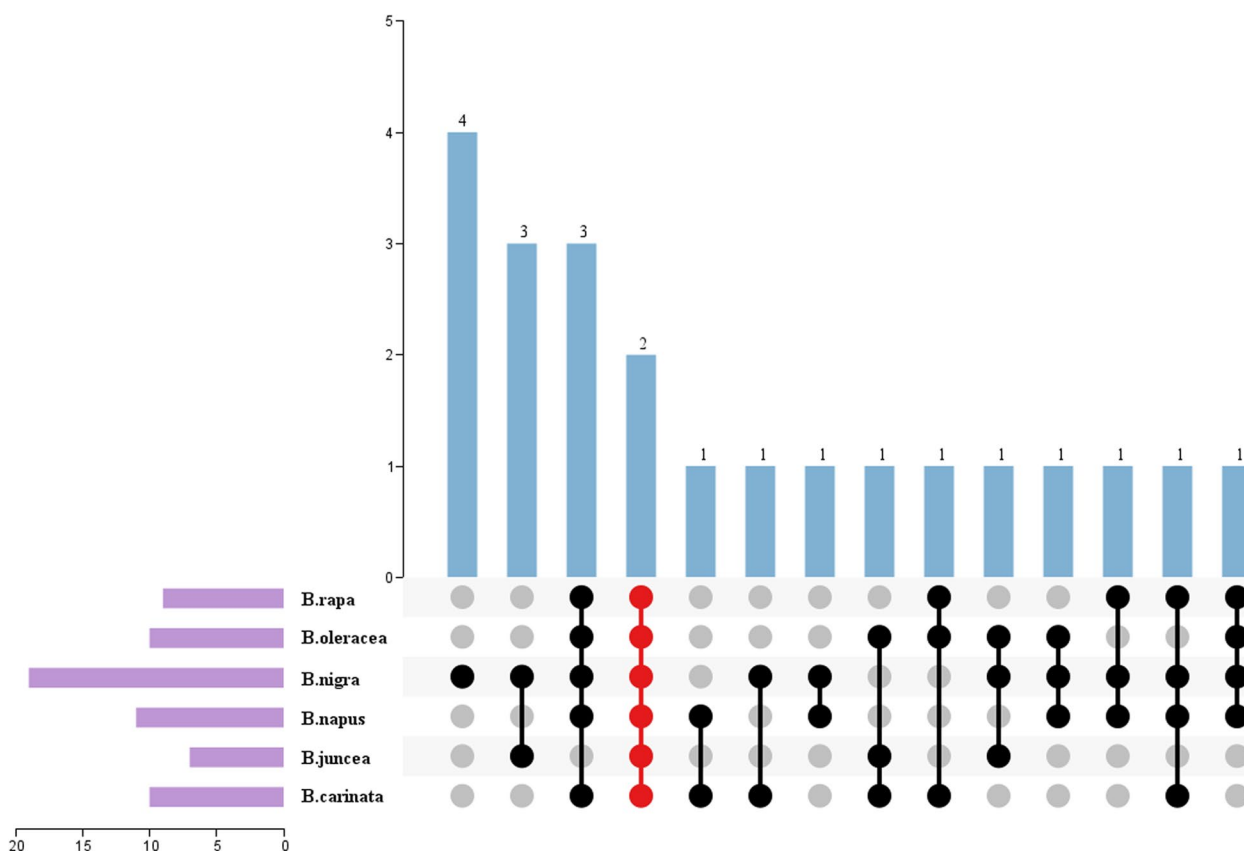
different transcriptional regulations of anthocyanins between vegetative and reproductive organs in *Brassica* crops. At the same time, *DFR(AT5G42800)*, *MYB5(AT3G13540)* and *AHA10(AT1G17260)* were differentially expressed in eight developmental stages of five *Brassica* crops, especially *DFR(AT5G42800)*, which is generally considered to be the target gene of the MBW transcriptional regulatory complex, was not differentially expressed only in *B. juncea* (Fig. 4, Supplementary Table S2).

#### Phylogenetics, gene structure and conserved motif of *MYB5* and *TT2*

Phylogenetic analysis, gene structure analysis, and identification of conserved motifs contribute to a better understanding of gene functional variation and evolutionary pathways. To further explore the functions of *TT2* and *MYB5*, we analyzed their copy variation and chromosomal distribution, a total of 20 *MYB5* homologous copies and 11 *TT2* homologous copies were identified (Fig. 5). The 2,000 base pairs (bp) upstream of *TT2* and *MYB5* homologous of the coding region were used to predict *cis* regulatory elements via the PlantCARE online tool, the sequences of gene region were used to draw the gene structure, and protein sequences were used for phylogenetic and

conserved domain analysis. (Fig. 5). Chromosomal mapping revealed two homologous copies of *MYB5* between the A and C subgenomes (except for *B. carinata*\_(BcaC), and three between the B subgenomes, and with good correspondence between the subgenomes and chromosomes (Fig. 6). Similarly, *TT2* had one homologous copy between A and C subgenomes, and two copies between B subgenomes, and the correspondence between A, B and C subgenomes and chromosomes was good (Fig. 6).

Phylogenetic analysis of *TT2* showed that its distribution among A, B (B1, B7 chromosomes) and C subgenomes could be well clustered. The gene structure and conserved domains of different copies of *TT2* were basically the same, and only *BniB01g032210.2N.1* and *BjuB01g23610S* had large variation (Fig. 6). Since the number of copies of *MYB5* reached 25, its orthologous genes could be clustered well. Except for *BcaB04g18801* and *BjuB07g46240S*, the gene structure and protein conserved domain had less variation. At the same time, *BraA05g033570.3C* was found to have an insertion of about 4 Kb in the intron region. The number and types of regulatory elements in each anthocyanin-related gene were different, and the distribution of copies of regulatory elements with similar clusters was similar (Fig. 6).



**Fig. 4** The differentially expressed anthocyanin-related genes between six *Brassica* crops seed coats of eight development RNA-seq data. The red color indicates the co-differentially expressed anthocyanin-related genes in six *Brassica* crops, the purple column indicated that anthocyanin related genes were differentially expressed in 8 developmental stages of the same material, the blue column shows the number of differentially expressed genes in common across all conditions

### Expression pattern analysis of MYB5 and TT2 during seed coat development

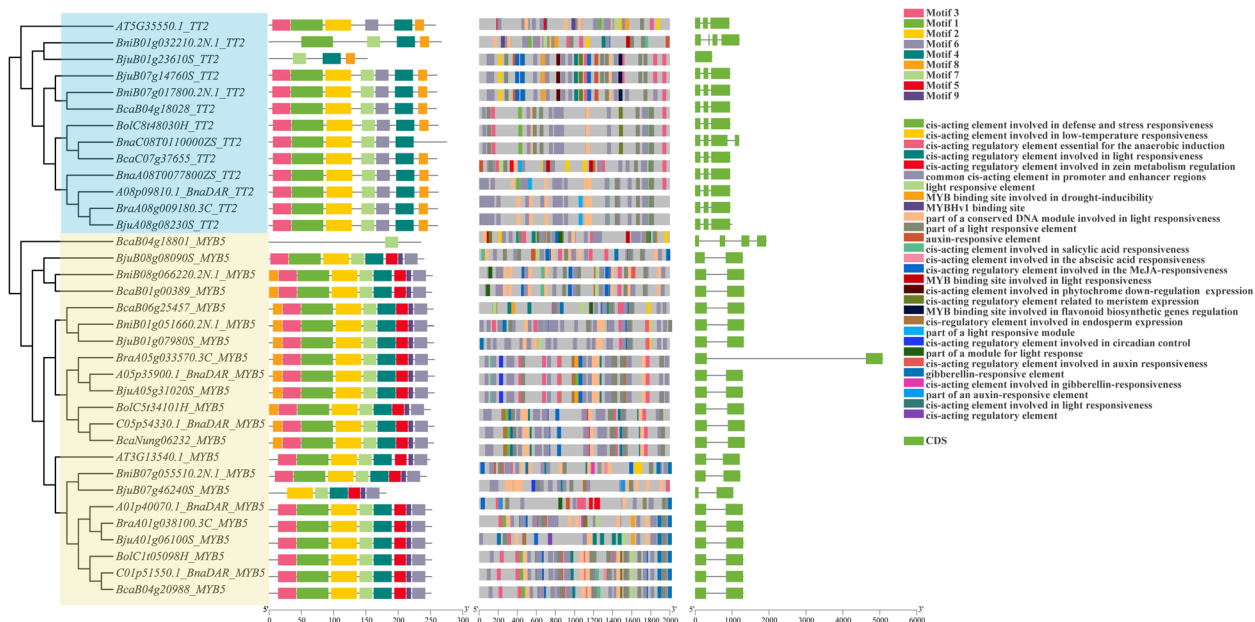
In order to gain insight into the expression patterns of *MYB5* and *TT2* during seed coat development of *Brassica* species, transcriptome data from eight developmental stages were selected for expression curve and trend analysis, respectively (Fig. 7). Most of the *MYB5* homologous gene copies initiated to be up-regulated in the S2 period, and the difference between the highest expression and the down-expression period was large. The most prominent one of *MYB5* homologous was *BniB01g051660.2N*, the expression level almost showed a parabolic trend from S1-S7 period, and the expression level was the highest in S3 period (Fig. 7). The homologous copy of *TT2* gene was mainly expressed at a high level in the S3-S4 period, *BniB07g017800.2N* had the highest expression level in the S1 period, and then decreased rapidly, while the *BolC8t48030H* expression was slowly up-regulated in the S0-S4 period, and expressed highest in the S6-S7 period (Fig. 7). In addition, six *MYB5* homologous copies (*BraA05g033570.3C*,

*BniB08g066220.2N*, *BjuA05g31020S*, *A05p35900.1\_Bna-DAR*, *BcaB01g00389*, *BcaB04g18801*) and two *TT2* homologous copies (*BniB01g032210.2N*, *BjuB01g23610S*) were not expressed in eight seed coat development stages of six *Brassica* crops, and these genes might have been silenced during evolution (Supplementary Table S3).

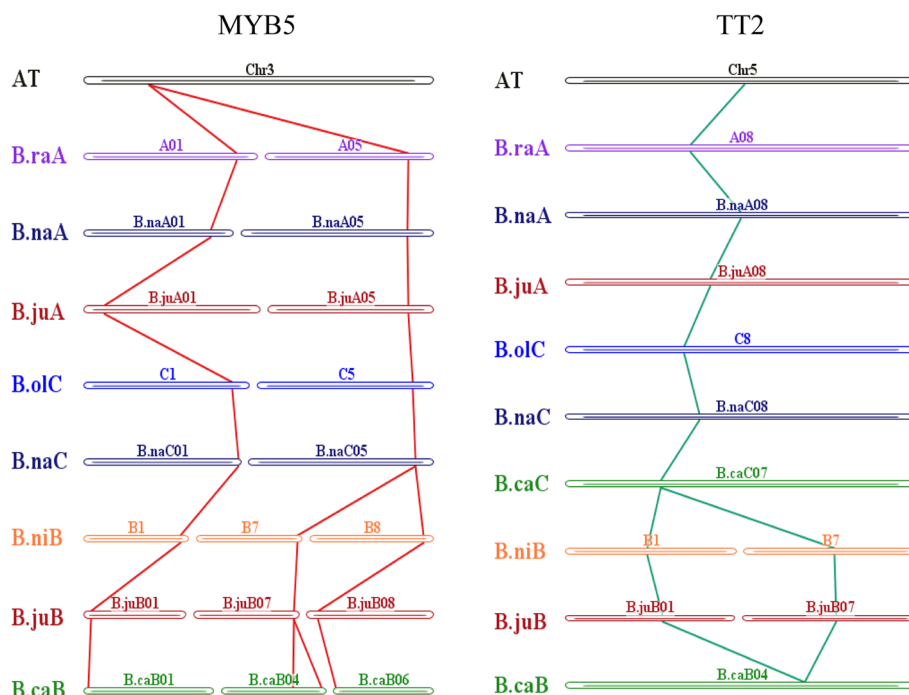
### Discussion

The diploid *Brassica* genome underwent whole-genome triple duplication, followed by gene loss and reshuffling of genome fragments, which generated the currently grown tetraploid *Brassica* plants, followed by crossing, polyploidization, and breeding selection [43]. The diversity of allelic combinations among *Brassica* species may affect the expression and regulation of anthocyanin biosynthesis-related genes during seed coat development, resulting in variations in *Brassica* species seed coat color [13, 85]. Here, we finely identified the orthologous and paralogous genes of anthocyanin biosynthesis-related genes in six *Brassica* species, and determined their copy number and chromosomal distribution. Gene expression





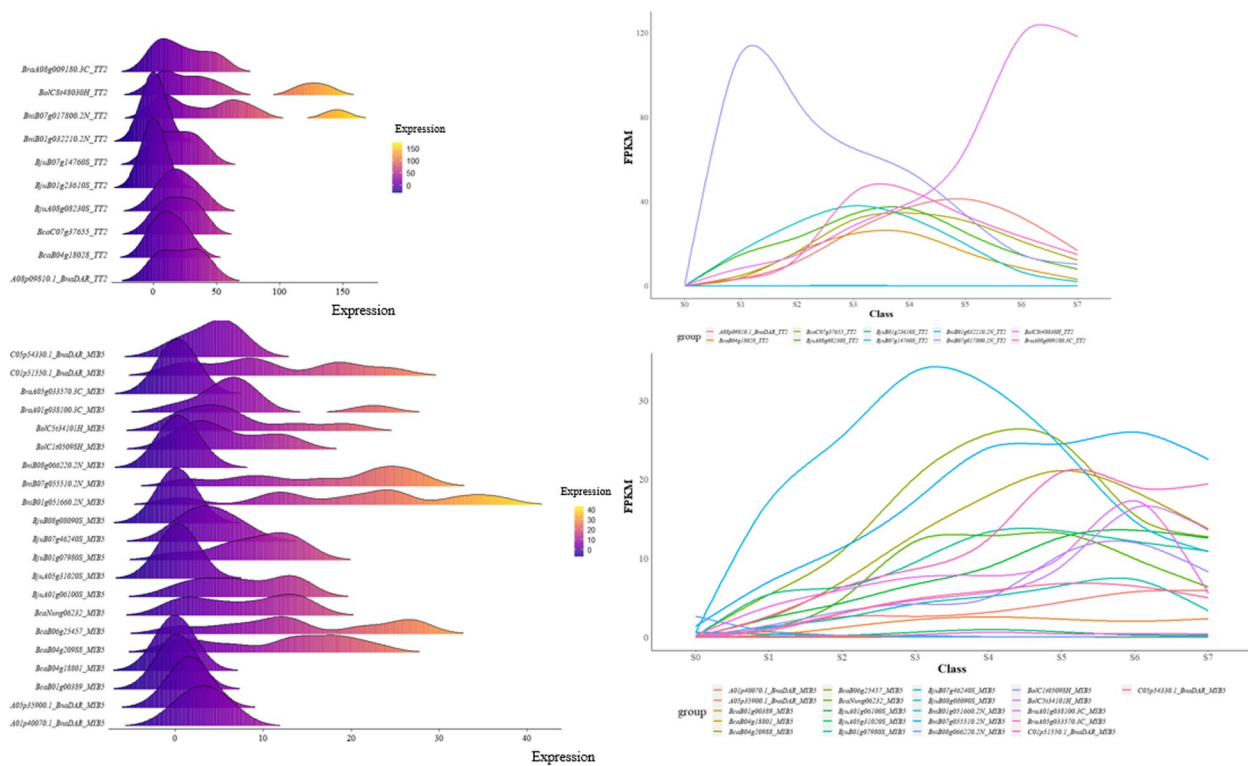
**Fig. 5** Phylogenetic, promoter characteristics, gene structure, conserved motifs analysis of *TT2* and *MYB5*. Predicted cis-elements in 7 co-differentially expressed R2R3-MYBs promoters and gene structure. Promoter sequences (-2,000 bp) of 7 co-differentially expressed R2R3-MYBs were analyzed using PlantCARE. Different shapes and colors represent different elements, PAFM was used for conserved domain prediction, and phylogenetic, promoter characteristics, gene structure, conserved motifs has been drawn by TBtools



**Fig. 6** The distribution and collinearity of *MYB5* and *TT2* in A, B, C subgenome chromosomes

analysis revealed many important differences in anthocyanin metabolic pathways during seed coat development in *Brassica* species, especially the transcription

factors *MYB5* and *TT2* were differentially expressed in eight developmental stages of six *Brassica* species, which provided new insights into the prediction and



**Fig. 7** Expression pattern analysis of *MYB5* and *TT2* during seed coat development. **A** Expression trend of *TT2* homologous genes in six *Brassica* species; **B** Expression curve of *TT2* homologous genes in six *Brassica* species; **C** Expression trends of *MYB5* homologous genes in six *Brassica* species; **D** Expression curve of *MYB5* homologous genes in six *Brassica* species

characterization of anthocyanin regulatory mechanisms and the evolutionary divergence that drove seed coat phenotypic differentiation in *Brassica* species.

**The anthocyanin-related genes in *Brassica* species**

The genes related to the biosynthetic pathway of anthocyanins have been well identified, and the related gene functions have been reported in a variety of plants [10, 11, 41, 19]. In *Arabidopsis*, rice and maize, the biosynthesis, regulation and transport of anthocyanins, especially most of the catalytic enzymes and regulatory genes involved in anthocyanin synthesis, have been identified and functionally validated over the past few decades [9, 24, 25, 33, 35, 36, 42, 60, 76]. These studies have played an important role in comprehensively understanding plant anthocyanin biosynthetic pathways and regulatory mechanisms, and revealing the accumulation and metabolic characteristics of plant anthocyanins. In order to further promote the research on the transcriptional regulations of anthocyanins in *Brassica* crops, we have finely identified the anthocyanin-related genes of six species based on the deciphered genomic information. In this study, 1119 anthocyanin-related genes were identified in six *Brassica* species (Table 1), which gave the first report

for *B. nigra*, *B. juncea* and *B. carinata*. In addition, the gene numbers of *B. rapa* (120), *B. oleracea* (131) and *B. napus* (258) were higher than those reported previously, 73 genes in *B. rapa* [14], 88 genes in *B. oleracea* [15], and 152 [16] and 157 [12] genes in *B. napus*. The main reasons were as follows: firstly, the reference genomes of *B. rapa*, *B. oleracea* and *B. napus* selected in this study were the latest published and had the best completeness after genome-wide comparison, secondly, the recent v10 and v11 genome information of *Arabidopsis* was used and 55 anthocyanin-related genes were identified in *A. thaliana*, finally, different methods were used to identify orthologous and paralogous genes, which made the identification results more complete and accurate.

After a long period of natural evolution and human selection, *Brassica* crops in U’s triangle have a lot of variation, including abundant types of leaf, stem, flower, silique and seed coat color variation. In the past two decades, from the mapping of the first gene related to anthocyanin synthesis in *B. rapa* [3] to the identification and characterization of a large number of genes related to anthocyanin synthesis, the transcriptional regulations of anthocyanin in *Brassica* species have been largely reported [3, 4, 17, 18, 64, 72, 73, 75, 77, 79, 82, 83]).

However, because the three *Brassica* diploids underwent a whole genome triploidization (WGT) event and then crossed to form three allotetraploids, the copy number of anthocyanin-related genes varied, and the functional analysis of anthocyanin-related genes lagged due to the existence of multiple copies in *Brassica* species. This study found the copies of different anthocyanin-related genes varied, some copies lost in *Brassica* crops (e.g. *MYB11*), and some were massively expanded (e.g. *PAL*, *GST*, *AHA10*) (Table 1). Some copies of *MYB5* and *TT2* led to gene silencing due to structural variation during genome rearrangement (Figs. 5 and 7). Collectively, our findings provided an in-depth genome-wide view of genes involved in the anthocyanin biosynthesis pathways in six interconnected *Brassica* species, which further disclosed the formation mechanism of color variation in different tissues or organs of *Brassica* crops.

#### Molecular mechanism of seed coat color variation in *Brassica* species

Yellow seed varieties of *B. rapa*, *B. juncea*, *B. carinata* and *B. napus* have attracted the attentions of researchers due to their reduced thickness, reduced lignin content, and defects in PAs biosynthesis. The excellent characteristics of yellow seeds result in low seed coat pigmentation, high proportion of seed oil, and synthesis of crude protein [59, 38]. In this study, through genome-wide identification and combining transcriptome data from eight seed coat developmental stages of six *Brassica* species, we constructed anthocyanin-related expression profiles during seed coat development (Fig. 2). Structural genes related to anthocyanin synthesis pathway were significantly up-regulated during seed coat development, and only *MYB5* and *TT2* were differentially expressed in eight developmental stages of six *Brassica* species. In *Arabidopsis*, the *TT2-TT8-TTG1* complex played a major role in seed development, but three other MBW complexes (*MYB5-TT8-TTG1*, *TT2-EGL3-TTG1*, and *TT2-GL3-TTG1*) also functioning as tissue-specific ways were shown to be involved [76]. However, in *Brassica* species, Sharma and Mohapatra [63] first mapped the yellow seed coat color locus in *B. juncea* by constructing an RFLP linkage map. Subsequently, a number of markers for seed coat color of *B. rapa* [58], *B. juncea* [47], *B. carinata* [37] and *B. napus* [34] were developed, and the yellow seed coat trait was partially dominant and under monogenic control. In recent years, with the development of sequencing technology, the genome information of six *Brassica* crops has been deciphered successively, and the research on genes related to seed coat color has entered the stage of fine mapping, cloning and functional verification. *BjuA.TT8* in *B. juncea* [50], *BnaC.TT2.a* and *BnTT8* in *B. napus* [84, 79]), *qSC9.1*, *qSCb9.1* and *qSC3.1* were successfully

validated in *B. rapa* [81]. In *B. napus*, *BnaC.TT2.a* was reported to be associated with seed coat coloration, and its InDel at position 738 on exon 3 indicated changes in protein function significantly associated with seed coat color [84]. At the same time, transcriptome sequencing combined with metabolome analysis revealed new progress in the regulation mechanism of seed coat color and color substances. Structural genes and transcription factors such as *DFR*, *LDOX(ANS)*, *BAN*, *TT3*, *TT8*, *TT18*, *TT10*, *TT12* and *TT19* were involved in the formation of seed coat coloration, and metabolites such as phenolic acids, epicatechin, flavonoids and PAs were important factors affecting seed coat formation in *Brassica* crops [55, 61, 64, 79].

Although the previous reports have provided a lot of information for revealing the variation of the seed coat color of *Brassica* species, the exact regulatory mechanism of the gene is less analyzed. In addition, there were few reports on the use of forward genetics to locate key genes of seed coat color and to analyze its molecular mechanism, which limited the use of molecular markers to assist in the breeding of yellow seed coat oilseed *Brassica*. Previous studies suggested that a single R2R3-MYB transcription factor combined with bHLH transcription factor and TTG1 protein to form the MBW transcriptional regulatory complex targeting *DFR/ANS* to drive anthocyanin synthesis and accumulation [76]. Our results further verified the important roles of *MYB5* and *TT2* and their homologous genes in the development of the seed coat of the six *Brassica* species, and from the analysis of the expression patterns and trends of different copies of *MYB5* and *TT2* at eight different developmental stages (Fig. 4). We speculate that during the formation of the MBW transcriptional regulatory complex, it is possible that multiple R2R3-MYBs (e.g. *MYB5* and *TT2*) associate simultaneously to form an MBW complex, which still needs to be verified by subsequent experiments. In summary, our results could broaden the scope of research on seed coat color and provided a reference for the subsequent improvement for yellow seed coat oilseed in *Brassica*.

#### 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
TT2	TRANSPARENT TESTA 2

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09170-2>.

**Additional file 1.** Anthocyanin-related genes in *B. carinata*.

**Additional file 2.** Co-differentially expressed genes in six Brassica species seed coats eight development.

**Additional file 3.** Expression of TT2 in all eight seed coat development stages of six Brassica species.

### Acknowledgements

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

### 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. In this study, all data were download from Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) under accession no. GSE153257, the author did not select any other material.

### Authors' contributions

CT, BZ, CC and ZL 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.

### Funding

This research was funded by the National Natural Science Foundation of China, grant number: 32160454; the Natural Science Foundation of Jiangxi Province, grant number 20212BAB215002.

### 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 download from Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) under accession no. GSE153257. If you need these materials and related data, you can contact Daoquan Xiang ([daoquan.xiang@nrc-cnrc.gc.ca](mailto:daoquan.xiang@nrc-cnrc.gc.ca)).

### 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 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.

#### Author details

<sup>1</sup>College of Life Sciences, Ganzhou Key Laboratory of Greenhouse Vegetable, Gannan Normal University, Ganzhou 341000, China. <sup>2</sup>Crop Research Institute, Sichuan Academy of Agricultural Sciences, Chendu 610066, China.

Received: 22 July 2022 Accepted: 7 February 2023

Published online: 09 March 2023

### References

- Anders S, Pyl P, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31:166–9.
- Bolger A, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
- Burdzinski C, Wendell DL. Mapping the anthocyaninless (anl) locus in rapid-cycling *Brassica rapa* (RBr) to linkage group R9. *BMC Genet*. 2007;8:64.
- Chen C, Chen H, Zhang Y, et al. TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol Plant*. 2020a;13:1194–202.
- Chen D, Chen H, Dai G, et al. Genome-wide identification of R2R3-MYB gene family and association with anthocyanin biosynthesis in *Brassica* species. *BMC Genomics*. 2022;23(1):441.
- Chen D, Liu Y, Jin Q, et al. Identification and functional analysis of BnaPAP2.A7 revealed that its alternative splicing isoforms play opposite roles in anthocyanin biosynthesis in *Brassica napus* L. *Front Plant Sci*. 2020b;11:983.
- Chen H, Wang T, He X, et al. BRAD V3.0: an upgraded Brassicaceae database. *Nucleic Acids Res*. 2022;50(D1):D1432–41.
- Cheng F, Wu J, Fang L, et al. Syntenic gene analysis between *Brassica rapa* and other *Brassicaceae* species. *Front Plant Sci*. 2012;3:198.
- Cone KC, Cocciolone SM, Burr FA, et al. Maize anthocyanin regulatory gene *pl* is a duplicate of *c1* that functions in the plant. *Plant Cell*. 1993;5(12):1795–805.
- Dooner HK, Robbins TP, Jorgensen RA. Genetic and developmental control of anthocyanin biosynthesis. *Annu Rev Genet*. 1991;25:173–99.
- Franken P, Niesbach-Klößgen U, Weydemann U, et al. The duplicated chalcone synthase genes *C2* and *Whp* (white pollen) of *Zea mays* are independently regulated; evidence for translational control of *Whp* expression by the anthocyanin intensifying gene in. *EMBO J*. 1991;10(9):2605–12.
- Fu H, Chao H, Zhao X, et al. Anthocyanins identification and transcriptional regulation of anthocyanin biosynthesis in purple *Brassica napus*. *Plant Mol Biol*. 2022;110:53.
- Gao P, Quilichini TD, Yang H, et al. Evolutionary divergence in embryo and seed coat development of U's Triangle *Brassica* species illustrated by a spatiotemporal transcriptome atlas. *New Phytol*. 2022;233(1):30–51.
- Guo N, Cheng F, Wu J, et al. Anthocyanin biosynthetic genes in *Brassica rapa*. *BMC Genomics*. 2014;15(1):426.
- Han F, Zhang X, Yang L, et al. Genome-wide characterization and analysis of the anthocyanin biosynthetic genes in *Brassica oleracea*. *Planta*. 2021;254(5):92.
- He D, Zhang D, Li T, et al. Whole-Genome Identification and Comparative Expression Analysis of Anthocyanin Biosynthetic Genes in *Brassica napus*. *Front Genet*. 2021;12:764835.
- He Q, Wu J, Xue Y, et al. 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.
- Heng S, Cheng Q, Zhang T, et al. Fine-mapping of the *BjPur* gene for purple leaf color in *Brassica juncea*. *Theor Appl Genet*. 2020;133(11):2989–3000.
- Holton TA, Cornish EC. Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell*. 1995;7(7):1071–83.
- Hong M, Hu K, Tian T, et al. Transcriptomic analysis of seed coats in yellow-seeded *Brassica napus* reveals novel genes that influence proanthocyanidin biosynthesis. *Front Plant Sci*. 2017;8:1674.
- Jiang J, Zhu S, Yuan Y, et al. Transcriptomic comparison between developing seeds of yellow-and black seeded *Brassica napus* reveals that genes influence seed quality. *BMC Plant Biol*. 2019;19:203.
- Kang L, Qian L, Zheng M, et al. Genomic insights into the origin, domestication and diversification of *Brassica juncea*. *Nat Genet*. 2021;53(9):1392–402.
- Katoh K, Standley D. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
- Kim CK, Cho MA, Choi YH, et al. Identification and characterization of seed-specific transcription factors regulating anthocyanin biosynthesis in black rice. *J Appl Genet*. 2011;52(2):161–9.
- Kubasek WL, Shirley BW, McKillop A, et al. Regulation of Flavonoid Biosynthetic Genes in Germinating *Arabidopsis* Seedlings. *Plant Cell*. 1992;4(10):1229–36.
- Larkin M, Blackshields G, Brown N, et al. Clustal W and clustal X version 2.0. *Bioinformatics*. 2007;23:2947–8.
- Lepiniec L, Debeaujon I, Routaboul JM, et al. Genetics and biochemistry of seed flavonoids. *Annu Rev Plant Biol*. 2006;57:405–30.

28. Li X, Chen L, Hong M, et al. A large insertion in bHLH transcription factor *BrTT8* resulting in yellow seed coat in *Brassica rapa*. PLoS ONE. 2012;7:e44145.
29. Lian J, Lu X, Yin N, et al. Silencing of *BnTT1* family genes affects seed flavonoid biosynthesis and alters seed fatty acid composition in *Brassica napus*. Plant Sci. 2017;254:32–47.
30. Lin A, Ma J, Xu F, et al. Differences in alternative splicing between yellow and black-seeded rapeseed. Plants (Basel). 2020;9(8):977.
31. Lin LZ, Harnly JM. Identification of the phenolic components of collard greens, kale, and Chinese broccoli. J Agric Food Chem. 2009;57:7401–8.
32. Liu H, Liu Z, Wu Y, et al. Regulatory mechanisms of anthocyanin biosynthesis in apple and pear. Int J Mol Sci. 2021;22(16):8441.
33. Liu J, Wang J, Wang M, et al. 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. 2021;12: 702160.
34. Liu Z, Fu T, Tu J, et al. Inheritance of seed colour and identification of RAPD and AFLP markers linked to the seed colour gene in rapeseed (*Brassica napus* L.). Theor Appl Genet. 2005;110(2):303–10.
35. Lloyd AM, Walbot V, Davis RW. *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regulators *R* and *C1*. Science. 1992;258(5089):1773–5.
36. Luo M, Lu B, Shi Y, et al. A newly characterized allele of *ZmR1* increases anthocyanin content in whole maize plant and the regulation mechanism of different *ZmR1* alleles. Theor Appl Genet. 2022;135:3039.
37. Marles MA, Gruber MY, Scoles GJ, et al. Pigmentation in the developing seed coat and seedling leaves of *Brassica carinata* is controlled at the dihydroflavonol reductase locus. Phytochemistry. 2003;62(5):663–72.
38. Marles MAS, Gruber MY. Histochemical characterisation of unextractable seed coat pigments and quantification of extractable lignin in the *Brassicaceae*. J Sci Food Agr. 2004;84(3):251–62.
39. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–12.
40. Meng JL, Shi SW, Gan L, et al. The production of yellow-seeded *B. brassica napus* (AACC) through crossing interspecific hybrids of *B. campestris* (AA) and *B. carinata* (BBCC) with *B. napus*. Euphytica. 1998;103:329–33.
41. Moi J, Stuitje A, Gerats A, et al. Saying it with genes: molecular flower breeding. Trends Biotechnol. 1989;7:148–53.
42. Moreno YS, Sánchez GS, Hernández DR, Lobato NR. Characterization of anthocyanin extracts from maize kernels. J Chromatogr Sci. 2005;43(9):483–7.
43. Murat F, Louis A, Maumus F, et al. Understanding *Brassicaceae* evolution through ancestral genome reconstruction. Genome Biol. 2015;16:262.
44. Nagaharu U, Nagaharu N. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Ja J Bot. 1935;7:389–452.
45. Nesi N, Delourme R, Brégeon M, et al. Genetic and molecular approaches to improve nutritional value of *Brassica napus* L. seed. Comptes Rendus Biologies. 2008;331(10):763–71.
46. Nesi N, Lucas MO, Auger B, et al. The promoter of the *Arabidopsis thaliana* BAN gene is active in proanthocyanidin-accumulating cells of the *Brassica napus* seed coat. Plant Cell Rep. 2009;28(4):601–17.
47. Padmaja LK, Arumugam N, Gupta V, et al. Mapping and tagging of seed coat colour and the identification of microsatellite markers for marker-assisted manipulation of the trait in *Brassica juncea*. Theor Appl Genet. 2005;111(1):8–14.
48. Padmaja LK, Agarwal P, Gupta V, et al. Natural mutations in two homoeologous *TT8* genes control yellow seed coat trait in allotetraploid *Brassica juncea* (AABB). Theor Appl Genet. 2014;127(2):339–47.
49. Padmaja LK, Agarwal P, Gupta V, et al. Natural mutations in two homoeologous in different parts of the *Arabidopsis thaliana* seedling. Plant J. 2014;50:660–77.
50. Padmaja LK, Agarwal P, Gupta V, et al. Natural mutations in two homoeologous *TT8* genes control yellow seed coat trait in allotetraploid *Brassica juncea* (AABB). Theor Appl Genet. 2014;127(2):339–47.
51. Pertea M, Kim D, Pertea GM, et al. Transcript-level expression analysis of RNA-seq experiments with HISAT StringTie and Ballgown. Nat Protoc. 2016;11:1650–67.
52. Price M, Dehal P, Arkin A. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS ONE. 2010;5: e9490.
53. Qi T, Song S, Ren Q, et al. The Jasmonate-ZIM-domain proteins interact with the WD-Repeat/bHLH/MYB complexes to regulate Jasmonate-mediated anthocyanin accumulation and trichome initiation in *Arabidopsis thaliana*. Plant Cell. 2011;23(5):1795–814.
54. Qu C, Yin N, Chen S, et al. Comparative analysis of the metabolic profiles of yellow-vs. black-seeded rapeseed using UPLC-HESI-MS/MS and transcriptome analysis. J Agr Food Chem. 2020;68:3033–49.
55. Qu C, Yin N, Chen S, et al. Comparative Analysis of the Metabolic Profiles of Yellow- versus Black-Seeded Rapeseed Using UPLC-HESI-MS/MS and Transcriptome Analysis. J Agric Food Chem. 2020;68(10):3033–49.
56. Rahman M, Joersbo M, Poulsen M. Development of yellow-seeded *Brassica napus* of double low quality. Plant Breed. 2008;120(6):473–8.
57. Rahman M, Mamidi S, McClean P, et al. Quantitative trait loci mapping of seed colour, hairy leaf, seedling anthocyanin, leaf chlorosis and days to flowering in F2 population of *Brassica rapa* L. Plant Breed. 2014;133:381–9.
58. Rahman M, McVetty PB, Li G. Development of SRAP, SNP and multiplexed SCAR molecular markers for the major seed coat color gene in *Brassica rapa* L. Theor Appl Genet. 2007;115(8):1101–7.
59. Rahman MH, Joersbo M, Poulsen MH. Development of yellow-seeded *Brassica napus* of double low quality. Plant Breeding. 2001;120(6):473–8.
60. Reddy VS, Dash S, Reddy AR. Anthocyanin pathway in rice (*Oryza sativa* L.): identification of a mutant showing dominant inhibition of anthocyanins in leaf and accumulation of proanthocyanidins in pericarp. Theor Appl Genet. 1995;91(2):301–12.
61. Ren Y, Zhang N, Li R, et al. Comparative transcriptome and flavonoids components analysis reveal the structural genes responsible for the yellow seed coat color of *Brassica rapa* L. PeerJ. 2021;9: e10770.
62. Romani A, Vignolini P, Isolani L, et al. HPLC-DAD/MS characterization of flavonoids and hydroxycinnamic derivatives in turnip tops (*Brassica rapa* L. Subsp. *sylvestris* L.). J Agric Food Chem. 2006;54(4):1342–6.
63. Sharma RP, Mohapatra T. Molecular mapping and tagging of genes in crop plants. Genetica. 1996;97(3):313–20.
64. Shen S, Tang Y, Zhang C, et al. Metabolite profiling and transcriptome analysis provide insight into seed coat color in *Brassica juncea*. Int J Mol Sci. 2021;22(13):7215.
65. Shi M, Xie D. Biosynthesis and metabolic engineering of anthocyanins in *Arabidopsis thaliana*. Recent Pat Biotechnol. 2014;8(1):47–60.
66. Song X, Wei Y, Xiao D, et al. *Brassica carinata* genome characterization clarifies U's triangle model of evolution and polyploidy in *Brassica*. Plant Physiol. 2021;186(1):388–406.
67. Tan C, Pan Q, Cui C, et al. Genome-wide gene/genome dosage imbalance regulates gene expressions in synthetic *Brassica napus* and derivatives (AC, AAC, CCA, CCAA). Front Plant Sci. 2016;7:1432.
68. Tang Z, Li J, Zhang X, et al. Genetic variation of yellow-seeded rapeseed lines (*Brassica napus* L.) from different genetic sources. Plant Breed. 1997;116:471–4.
69. Theodoridou K, Yu P. Effect of processing conditions on the nutritive value of canola meal and presscake. Comparison of the yellow and brown-seeded canola meal with the brown-seeded canola presscake. J Sci Food Agr. 2013;93(8):1986–95.
70. Wang J, Liu Y, Tang B, et al. Genome-wide Identification and capsacinoid biosynthesis-related expression analysis of the R2R3-MYB gene family in *Capsicum annuum* L. Front Genet. 2020;11: 598183.
71. Wang Y, Tang H, Debarry JD, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene syntenies and collinearity. Nucleic Acids Res. 2012;40: e49.
72. Wang Y, Xiao L, Dun X, et al. Characterization of the *BrTT1* gene responsible for seed coat color formation in Dahuang (*Brassica rapa* L. landrace). Mol Breeding. 2017;37:137.
73. Wang Y, Liu Z, Li Y, et al. Identification of sequence-related amplified polymorphism markers linked to the red leaf trait in ornamental kale (*Brassica oleracea* L. var. *acephala*). Genet Mol Res. 2013;12(2):870–7.
74. Xiao L, Zhao Z, Du D, et al. Genetic characterization and fine mapping of a yellow-seeded gene in Dahuang (a *Brassica rapa* landrace). Theor Appl Genet. 2012;124:903–9.
75. Xie T, Chen X, Guo T, et al. Targeted knockout of *BnTT2* homologues for yellow seeded *Brassica napus* with reduced flavonoids and improved fatty acid composition. J Agric Food Chem. 2020;68:5676–90.
76. Xu W, Grain D, Bobet S, et al. 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.

77. Yan C, An G, Zhu T, et al. Independent activation of the *BoMYB2* gene leading to purple traits in *Brassica oleracea*. *Theor Appl Genet.* 2019;132(4):895–906.
78. Yu CY. Molecular mechanism of manipulating seed coat coloration in oilseed *Brassica* species. *J Appl Genet.* 2013;54(2):135–45.
79. Zhai Y, Yu K, Cai S, et al. Targeted mutagenesis of *BnTT8* homologs controls yellow seed coat development for effective oil production in *Brassica napus* L. *Plant Biotechnol J.* 2020;18(5):1153–68.
80. Zhang J, Lu Y, Yuan Y, et al. Map-based cloning and characterization of a gene controlling hairiness and seed coat color traits in *Brassica rapa*. *Plant Mol Biol.* 2009;69:553–63.
81. Zhang Y, Sun Y, Sun J, Feng H, et al. Identification and validation of major and minor QTLs controlling seed coat color in *Brassica rapa* L. *Breed Sci.* 2019;69(1):47–54.
82. Zhao H, Shang G, Yin N, et al. Multi-omics analysis reveals the mechanism of seed coat color formation in *Brassica rapa* L. *Theor Appl Genet.* 2022;135(6):2083–99.
83. Zhao Z, Xiao L, Xu L, et al. Fine mapping the *BjPl1* gene for purple leaf color in B2 of *Brassica juncea* L. through comparative mapping and whole-genome re-sequencing. *Euphytica.* 2017;213:80.
84. Zhou L, Li Y, Hussain N, et al. Allelic Variation of *BnaC.TT2.a* and Its Association with Seed Coat Color and Fatty Acids in Rapeseed (*Brassica napus* L.). *PLoS One.* 2016;11(1):e0146661.
85. Ziegler DJ, Khan D, Kalichuk JL, et al. Transcriptome landscape of the early *Brassica napus* seed. *J Integr Plant Biol.* 2019;61:639–50.
86. Zhu M, Hu R, Zhao H, et al. Identification of quantitative trait loci and candidate genes controlling seed pigments of rapeseed. *J Integr Agr.* 2021;20(11):2862–79.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

