

RESEARCH

Open Access



# Molecular evolution analysis of *MYB5* in Brassicaceae with specific focus on seed coat color of *Brassica napus*

Guoqiang Dai<sup>1</sup>, Yi Liu<sup>1</sup>, Wenjie Shen<sup>1</sup>, Bo Zhu<sup>1</sup>, Lunlin Chen<sup>2\*</sup>, Daozong Chen<sup>1\*</sup> and Chen Tan<sup>1\*</sup>

## Abstract

**Background** MYB transcription factors play a vital role in plant biology, with previous research highlighting the significant impact of the R2R3-MYB-like transcription factor *MYB5* on seed mucilage biosynthesis, trichome branching, and seed coat development. However, there is a dearth of studies investigating its role in the regulation of proanthocyanidin (PA) biosynthesis.

**Results** In this study, a total of 51 *MYB5* homologous genes were identified across 31 species belonging to the Brassicaceae family, with particular emphasis on *Brassica napus* for subsequent investigation. Through phylogenetic analysis, these genes were categorized into four distinct subclasses. Protein sequence similarity and identity analysis demonstrated a high degree of conservation of *MYB5* among species within the Brassicaceae family. Additionally, the examination of selection pressure revealed that *MYB5* predominantly underwent purifying selection during its evolutionary history, as indicated by the Ka/Ks values of all *MYB5* homologous gene pairs being less than one. Notably, we observed a higher rate of non-synonymous mutations in orthologous genes compared to paralogous genes, and the Ka/Ks value displayed a stronger correlation with Ka. In *B. napus*, an examination of expression patterns in five tissues revealed that *MYB5* exhibited particularly high expression in the black seed coat. The findings from the WGCNA demonstrated a robust correlation between *MYB5* and *BAN(ANR)* associated with PA biosynthesis in the black seed coat, providing further evidence of their close association and co-expression. Furthermore, the results obtained from the analysis of protein interaction networks offer supplementary support for the proposition that *MYB5* possesses the capability to interact with transcriptional regulatory proteins, specifically *TT8* and *TT2*, alongside catalytic enzymes implicated in the synthesis of PAs, thereby making a contribution to the biosynthesis of PAs. These findings imply a plausible and significant correlation between the unique expression pattern of *MYB5* and the pigmentation of rapeseed coats. Nevertheless, additional research endeavors are imperative to authenticate and substantiate these findings.

**Conclusions** This study offers valuable insights into the genetic evolution of Brassicaceae plants, thereby serving as a significant reference for the genetic enhancement of Brassicaceae seed coat color.

**Keywords** Brassicaceae, *MYB5*, Homologous gene, Seed coat, *Brassica napus*

\*Correspondence:

Lunlin Chen  
 lunlinchen@163.com  
 Daozong Chen  
 chendaizong@gnnu.edu.cn  
 Chen Tan  
 tanchen2020@gnnu.edu.cn

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



© The Author(s) 2024. **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.

## Background

The MYB transcription factor family in plants is known for its large size and is characterized by the presence of 1–4 R repeat sequences, with each repeat sequence consisting of approximately 52 amino acids. Based on the number of R repeats, MYB transcription factors can be classified into four types: 1R-MYB, R2R3-MYB, 3R-MYB and 4R-MYB [1]. The discovery of the first MYB transcription factor, *COLORED1* (*C1*), which regulates anthocyanin synthesis in the maize (*Zea mays*) grain aleurone layer dates back more than 30 years [2]. Since then, extensive research has been conducted on the involvement of MYB transcription factors in various plant species, including *Arabidopsis thaliana* [3], rice (*Oryza sativa*) [4], grape (*Vitis vinifera*) [5], apple (*Malus domestica*) [6], and petunia (*Petunia hybrida*) [7], utilizing sequencing, genetic, and molecular analysis techniques. Consequently, novel insights into the regulatory mechanisms governing MYB protein activity have emerged, along with the identification of the gene expression profiles and certain target genes. Among the four types discussed, the R2R3-MYB transcription factor emerges as the most common prevalent, characterized by the presence of two consecutive R repeats [8]. Extensive research has demonstrated that the R2R3-MYB transcription factor primarily exerts its influence on the structural genes involved in the plant anthocyanin biosynthesis pathway. Additionally, it has been observed to form a transcriptional regulatory complex with bHLH transcription factors and WD40 proteins, collectively known as the MBW complex. This complex plays a crucial role in modulating the synthesis and accumulation of anthocyanin [8, 9].

Throughout the extensive process of evolution, R2R3-MYB transcription factors have displayed functional diversity and can be found in all land plants. Throughout the extensive process of evolution, the functions of R2R3-MYB transcription factor have undergone divergence, leading to the characterization of numerous R2R3-MYB proteins through genetic methods. These proteins have been found to play a role in regulating plant-specific processes, such as responses to biotic and abiotic stresses [10, 11], cell fate and identity, primary and secondary metabolism [8, 12–16] and developmental processes [17]. The R2R3-MYB transcription factor, known as MYB5, has been extensively studied and identified in various plant species. Its primary functions include the regulation of tannin biosynthesis [16], trichome elongation [12, 13], outer seed coat differentiation [13, 16] and procyanidin synthesis [18]. In *A. thaliana*, MYB5 typically forms a complex with *TT8* and *TTG1*, collectively referred to as the MBW complex, to directly target *DFR*, *ANS*, or *BAN* genes and regulate flavonoid biosynthesis in seeds [18].

The Brassicaceae family encompasses a total of 351 genera and approximately 4,000 species, establishing it as a prominent constituent within the plant kingdom [19]. Including the model plant *A. thaliana*, oil crops *Brassica napus*, *B. juncea*, vegetable crops *B. rapa*, *B. oleracea*, *Raphanus sativus*, medicinal plants *Isatis indigotica* and *Lepidium apetalum*, which has important research and utilization value. Of these, *B. napus* is the third oil crop cultivated globally, contributes approximately 15% to the overall production of vegetable oil, thereby playing a significant role in the provision of edible oil, biofuel, and animal feed [20]. Enhancing oil content, refining oil quality, and developing high-yielding rapeseed varieties have been the primary objectives of breeding efforts in recent decades [21–24]. Previously, breeders engaged in cultivar selection observed that rapeseed with yellow seeds exhibited superior quality attributes, including higher oil and protein content, as well as lower levels of antinutrients such as lignin, fiber, pigments, and polyphenols [25]. Similarly, investigations conducted on *B. napus* revealed that silencing *BnTT10* gene expression resulted in elevated soluble PAs within the seeds, decreased extractable lignin, and delayed pigmentation of the seed coat [26]. The implementation of RNA interference to silence *BnTT1* family genes in rapeseed resulted in various phenotypic changes, including yellowing of the plant, thinning of the seed coat, reduction in seed weight, and alterations in the fatty acid composition of the seeds [27]. Additionally, targeted knockout of the *BnTT2* homologs in *B. napus* led to a decrease in flavonoids and an improvement in the fatty acid composition of the seeds [28]. Similarly, targeted mutation of the *BnTT8* homologs hindered the expression of phenylpropanoid and flavonoid biosynthesis genes, consequently inhibiting the accumulation of proanthocyanidins in the seed coats [29]. In our previous investigation, it was observed that *MYB5* and *TT2* exhibited differential expression patterns throughout the 8 stages of seed coat development in 6 distinct *Brassica* species. This finding implies that these genes may play a pivotal role in the modulation of seed coat color alteration [30]. Nonetheless, the precise mechanism by which *MYB5* specifically participates in the regulation of seed coat color in *B. napus* remains unreported and necessitates further investigation.

In this study, we employed the published reference genome information of 31 species belonging to the Brassicaceae family to identify homologous genes of *MYB5*. This was achieved through the utilization of homologous sequence alignment and conserved domain analysis techniques. Additionally, we conducted an analysis of the molecular evolution characteristics of *MYB5* homologous genes in Brassicaceae, which enabled us to construct an evolutionary map. Subsequently, we investigated the

transcriptional regulation mechanism of the *MYB5* gene in *B. napus*, focusing on its role in seed coat color formation. Our analysis involved examining tissue-specific expression patterns and co-expression patterns. The objective of this research was to enhance our comprehension of the gene function of *MYB5*. Furthermore, this study aims to establish a basis for future investigations into the molecular mechanism by which *MYB5* specifically regulates seed coat color formation, and to offer insights for the improving the quality of rapeseed oil.

## Methods

### Retrieval and identification of *MYB5* in Brassicaceae

Here, all the protein sequences and CDS sequences of 31 Brassicaceae species were obtained from the TBGR database (<http://www.tbgr.org.cn/>), respectively [31]. The *MYB5* gene (AT3G13540) from *A. thaliana* was selected as the reference for identifying homologous genes in the Brassicaceae family. To ensure accurate identification of *MYB5* homologous genes in the 31 Brassicaceae species, the following steps were primarily employed. Firstly, initial searches were conducted using local BLASTP and local BLASTN algorithms, with a significance threshold of  $E < 1e - 20$ . Secondly, the screening process involved identifying candidate genes that exhibited a consistency rate exceeding 65% and a coverage rate surpassing 60%. Thirdly, hmmsearch [32] was employed to search for potential family genes containing the MYB domain in 31 species. This search utilized the Hidden Markov model PF00249 (HMM) of the MYB domain, which was obtained from the Pfam database (<http://pfam-legacy.xfam.org/>) [33]. Subsequently, the outcomes of BLAST and hmmsearch were combined, and the CDD database (<https://www.ncbi.nlm.nih.gov/cdd/>) was employed to forecast the conserved domain, thereby eliminating genes lacking the PLN03212 superfamily. The identified *MYB5* homologous genes mentioned earlier were utilized for subsequent phylogenetic analysis. For a comprehensive understanding of the identification procedures, refer to Pucker [34].

### Construction of a phylogenetic tree

In order to validate the findings of the IQ-TREE2 phylogenetic analysis, we utilized the *MYB5* (MD07G0126400) orthologous gene of apple (*Malus domestica*, golden, ASM211411v1, <http://plants.ensembl.org/index.html>) as the outgroup. The protein sequences of the identified 51 *MYB5* homologous genes were subjected to comparison using MAFFT (v7.520) [35] software with default parameters. For the construction of the phylogenetic tree, we employed IQ-TREE2 software (v1.6.12) [36], 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) [37] for constructing a phylogenetic tree. The LG model was chosen for tree construction, with the --bs-trees parameter set to 1000 and the -all parameter selected to facilitate multiple search process aimed at identifying the optimal tree topology and model parameters. Subsequently, the phylogenetic tree was visualized using the iTOL online website (v6.8) [38] accessible at <https://itol.embl.de/>.

### *Cis*-acting elements and conservative domain analysis

All 51 identified *MYB5* homologues genes upstream of the start codon 2Kb genome sequences were extracted and used for the prediction of *cis*-acting elements in the promoter region. PlantCARE [39] website (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) was used to predict the *cis*-acting elements of *MYB5* homologues genes promoters. Then, all predicted *cis*-acting elements were classified and analyzed. The identified 51 *MYB5* 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 software (v1.133) [40] using phylogenetic files (from FastTree2), conserved motifs (from CDD), and *cis*-acting elements prediction files (from PlantCARE).

### Selective pressure analysis

The CDS sequences of all identified genes were compared using the BLAST software, with a threshold value of  $E < 1e-20$ , to identify tandem repeat pairs of *MYB5* homologous genes. Subsequently, the KaKs\_Calculator software (v2.0), as described by Wang et al. [41], was utilized to estimate the synonymous substitution rate (Ks) and non-synonymous substitution rate (Ka). Duplicated genes were excluded if the Ka value is close to 0 and the Ks value exceeded 3 or was below 0.01. This exclusion was necessary due to the potential saturation risk associated with high Ks values and the possibility of unknown outcomes resulting from low sequence divergence. The Ka/Ks  $< 1$ ,  $> 1$ , and  $= 1$  indicated purifying selection, forward selection, and neutral selection, respectively [42]. The visualization and plotting of the pressure analysis data were performed using the R package. For a more comprehensive understanding of the identification methods employed, please refer to Cao et al [43].

### *MYB5* expression analysis in *B. napus*

In order to further investigate the expression patterns of *MYB5* homologous genes in various tissues of *B. napus*, transcriptome data from *B. napus* leaves (green, purple), stems (green, purple), flower petals (orange-red, yellow),

siliques (green, purple), and seed coat (yellow, 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 for stems and flower petals, PRJNA734925 for siliques, and PRJNA597958 for seeds. The raw data obtained from the NCBI was initially converted into fastq format using the SRA toolkit, following which it was processed with default parameters using trimmomatic (v0.39) [44]. Subsequently, the RNA-seq clean reads were aligned to the reference genome of *B. napus* (Darmor-v10, <https://yanglab.hzau.edu.cn/BnIR/download?module=genomics>) using the HISAT2 software (v2.1.0) [45]. The obtained millions of fragments per thousand bases (FPKM) values were then calculated using StringTie (v2.1.1) [45]. The expression histogram was generated using Excel, while the heat map was generated using TBtools software (v1.133) [40].

#### Construction of co-expression network modules using association analysis modules and phenotypes

To explore the metabolite fluxes of the PA biosynthetic pathway regulated by MYB5, we employed the weighted gene co-expression network analysis (WGCNA) to establish a gene co-expression network for the transcriptional regulation of PA in various tissues of *B. napus*. The construction of co-expression networks was carried out using the WGCNA software package, specifically version R-4.3.1. Subsequently, correlation analysis was performed to determine the correlation between each co-expression module and the collected tissue data of *B. napus*, enabling the identification of differentially expressed genes (DEGs) associated with PA. These DEGs were then utilized to construct a co-expression network module. For the detailed method, refer to the articles of Cheng et al. [40]. STRING software (<https://version-11-5.string-db.org/>) was used to reveal a protein-protein interaction network of ‘turquoise’ module genes.

## Results

#### Identification and phylogenetic analysis of MYB5 homologous genes in Brassicaceae

To comprehensively and accurately ascertain the MYB5 homologous genes in Brassicaceae species, the *AT3G13540* sequence from *A. thaliana* was obtained and employed as a reference to explore potential candidate genes across the 31 Brassicaceae species with available genomic data. A total of 51 MYB5-related genes were discovered in these 31 Brassicaceae species, with the number of MYB5 homologous gene copies varying from one to five (Table 1, Supplementary Table S1). A total of 21 species of Brassicaceae plants were identified, with only

one homologous gene being found. Among these species, allotetraploid *B. juncea* exhibited the highest number of homologous genes, totaling five, whereas the diploid crop *B. oleracea* also possessed four homologous genes (Table 1).

To study the evolutionary connections among MYB5 homologous genes, we assigned new names to the MYB5 homologous genes from 31 species (Supplementary Table S2) and proceeded to construct a maximum likelihood phylogenetic tree based on their protein sequences. The 51 MYB5 homologous genes were categorized into four distinct subcategories (S1–S4) and visually represented using different colors (Fig. 1). Notably, the branches identified in *A. thaliana* were marked with green color on the evolutionary tree, indicating their placement within the S4 subcategory (Fig. 1). The S4 subcategory exhibits the highest number of MYB5 homologous genes, totaling 24, followed by the S3 subclass which contains 14 MYB5 homologous genes. Interestingly, within the S8 subgroup, 8 out of the 14 MYB5 orthologs were identified in *Brassica* species, while 8 out of the 10 MYB5 orthologs within the S2 subgroup were also found in *Brassica* species. This observation suggests that the MYB5 gene in *Brassica* species underwent duplication and diverged at an early stage during an early developmental phase. Furthermore, the MYB5 genes derived from *R. sativus*, *S. arvensis*, *S. parvula*, *S. alba*, *S. irio*, *O. violaceus*, and *I. indigotica* exhibit a closer clustering with the MYB5 genes of *Brassica* crops, aligning with the evolutionary analysis findings of Brassicaceae species as reported by Yang et al. [46]. In general, the evolutionary relationship observed among the MYB5 homologous genes across the 31 species concurs with the phylogenetic outcomes reported for Brassicaceae by Nikolov et al. [47].

#### Conserved motifs and cis-acting element analysis of MYB5 homologous genes

To identify the functional differences of MYB5 homologous genes in 31 Brassicaceae species, we analyzed the cis-acting elements in the promoters of 51 MYB5 homologous genes with PlantCARE software. There are 11 kinds of cis-acting elements were found in the promoter region of the MYB5 homologous genes, and most of them are involved in plant growth and metabolism, environmental and hormonal signal responses response, and MYB binding site (Fig. 2, Supplementary Table S3). Among the 11 types of cis-acting elements, light responsiveness, auxin responsiveness, gibberellin-responsiveness, abscisic acid responsiveness, MeJA-responsiveness, defense and stress responsiveness, salicylic acid responsiveness, and auxin responsiveness are involved in environmental and hormonal signal responses, indicating that the MYB5 homologous genes could be induced by



**Table 1** Genome version information and *MYB5* homologous genes in 31 Brassicaceae species

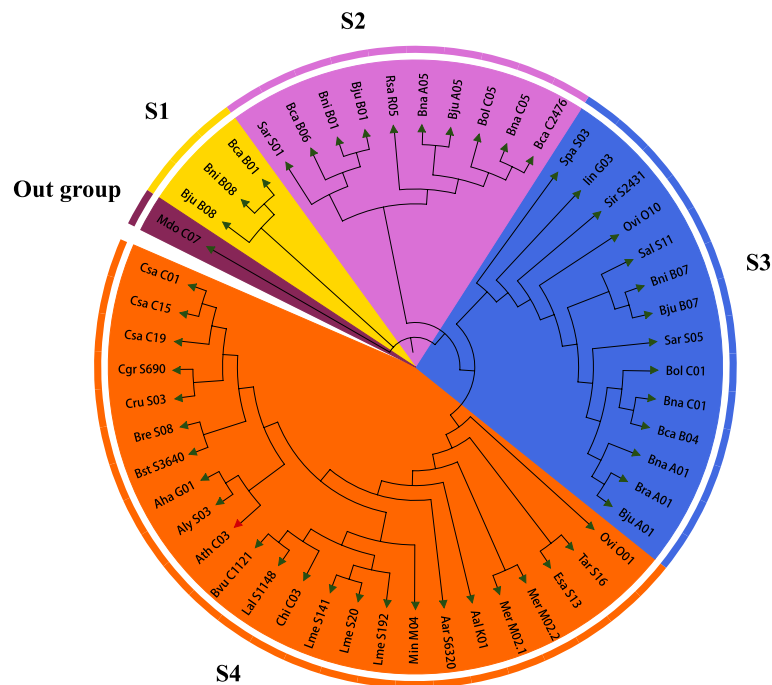
Species	Version	Blast+Hmsearch	CDD	Result
<i>Aethionema_arabicum</i>	A.arabicum-v1.0	1	1	1
<i>Arabidopsis_halleri</i>	A.halleri-v2.2	1	1	1
<i>Arabidopsis_lyrata</i>	Lyrata -v2.1	1	1	1
<i>Arabidopsis_thaliana</i>	Thale -Araport11	1	1	1
<i>Arabis_alpina</i>	Gray -v4.0	2	1	1
<i>Barbarea_vulgaris</i>	Bittercress-v1.0	1	1	1
<i>Brassica_junceae</i>	SCHZ	5	5	5
<i>Boechera_retrofracta</i>	Holboell-v1.0	1	1	1
<i>Boechera_stricta</i>	Drummond-v1.2	1	1	1
<i>Brassica_carinata</i>	Zd-1-v1.0	4	4	4
<i>Brassica_napus</i>	Darmor-v10.0	4	4	4
<i>Brassica_nigra</i>	Ni100_LR-v2.0	3	3	3
<i>Brassica_oleracea</i>	HDEM-v0.0	4	2	2
<i>Brassica_rapa</i>	Chiifu-v3.5	1	1	1
<i>Camelina_sativa</i>	Camelina-v2.0	3	3	3
<i>Capsella_grandiflora</i>	C.grandiflora-v1.0	2	1	1
<i>Capsella_rubella</i>	Red_shepherd -v1.1	1	1	1
<i>Cardamine_hirsuta</i>	Hairy -v1.0	1	1	1
<i>Eutrema_salsugineum</i>	173_v1.0	1	1	1
<i>Isatis_indigotica</i>	Woad-v1.0	1	1	1
<i>Leavenworthia_alabamica</i>	Alabama -v1.0	1	1	1
<i>Lepidium_meyenii</i>	Peruvian -v1.0	3	3	3
<i>Matthiola_incana</i>	M.incana-v1.0	1	1	1
<i>Microthlaspi_erraticum</i>	M.erraticum-v1.0	2	2	2
<i>Orychophragmus_violaceus</i>	O.violaceus-v1.0	2	2	2
<i>Raphanus_sativus</i>	Radish-v1.0	1	1	1
<i>Schrenkiella_parvula</i>	Saltwater -v1.0	1	1	1
<i>Sinapis_alba</i>	S.alba-v1.0	1	1	1
<i>Sinapis_arvensis</i>	S.arvensis-v1.0	2	2	2
<i>Sisymbrium_iriio</i>	London -v1.0	1	1	1
<i>Thlaspi_arvense</i>	Field -v1.1	1	1	1
Total	31	55	51	51

environmental and hormonal to regulate plant growth and development. Anaerobic induction, abscisic acid responsiveness, defense and stress responsiveness are the main members of the regulation of biological and abiotic stress. At the same time, a *cis*-element related to meristem expression was found in all 51 *MYB5* homologous genes promoter, indicating that *MYB5* may be involved in the regulation of plant growth and development by regulating the expression of meristem genes. In addition, we also found that the MYB binding site is distributed in the promoters of most *MYB5* homologous genes, indicating that *MYB5* may combine with other MYB transcription factors to participate in the regulation of metabolic pathways. Our results show that during the duplication of *MYB5* gene, the *cis*-element in the promoter region

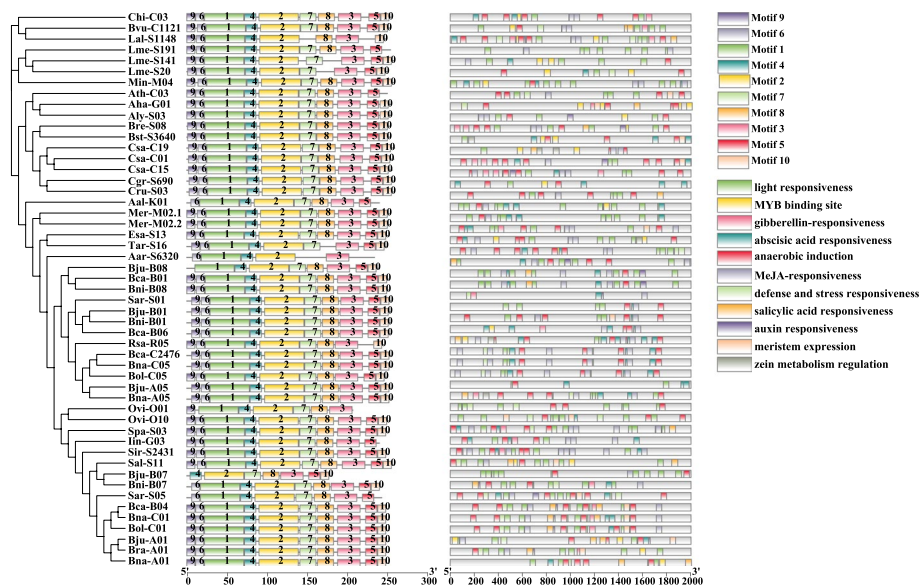
between homologous genes may be mutated, resulting in the change of *cis*-acting element, and the change of *cis*-element may lead to different expression levels of homologous genes, resulting in differentiation of gene function. However, these speculations need to be further verified by follow-up experiments.

#### Positive selection in *MYB5* homologous genes

To explore the impact of selection pressure on the evolutionary process of *MYB5* homologous genes, the Ks and Ka values of ortholog/paralog pairs within the Brassicaceae family were computed (Fig. 3, Supplementary Table S4). The analysis revealed that out of the total 2467 gene pairs, all exhibited Ka/Ks ratios below 1, with 30 gene pairs falling within the range of 0.5 to 0.99, and



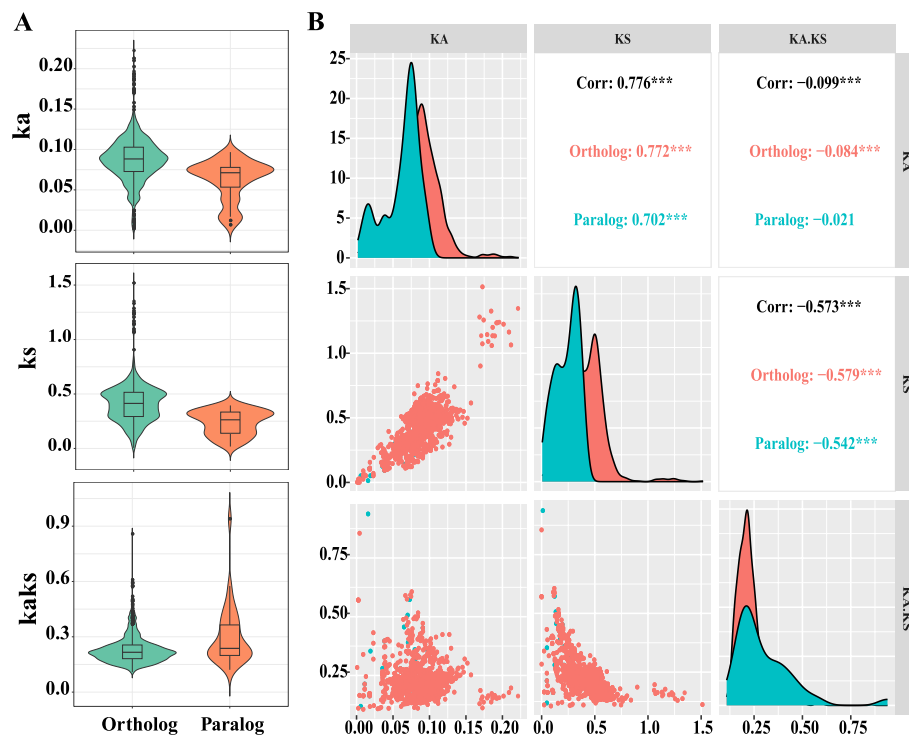
**Fig. 1** Phylogenetic relationships of MYB5 proteins between 31 Brassicaceae species. All 4 subfamilies of MYB5 proteins were well separated in different clades and represented by different colors. IQ-TREE software (v1.6.12) to build a phylogenetic tree, where the -m parameter selects MFP, automatically detects the best model and builds a tree



**Fig. 2** Phylogenetic, conserved motifs and promoter characteristics analysis of MYB5. Promoter sequences (-2,000 bp) of 51 MYB5 homologous genes 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 have been drawn by TBtools

the remaining 2437 gene pairs displaying Ka/Ks ratios  $< 0.5$ . Notably, all 2467 gene pairs had Ka values  $< 1$ , with the highest value recorded as 0.22 (Supplementary Table S5), indicating a remarkably low occurrence

of non-synonymous mutations in the *MYB5* homologous genes throughout their evolutionary history. Concurrently, within the set of 2467 gene pairs, 35 pairs (1.42%) exhibited a Ks value > 1, while 705 pairs (28.58



**Fig. 3** Comparison of substitution rates between orthologous and paralogous MYB5 genes among 31 Brassicaceae species. **A** Comparative analysis of *Ka*, *Ks* and *Ka/Ks* of paralogous and orthologous MYB5 genes. **B** Correlation analysis between *Ka*, *Ks* and *Ka/Ks* of paralogous and orthologous MYB5 genes

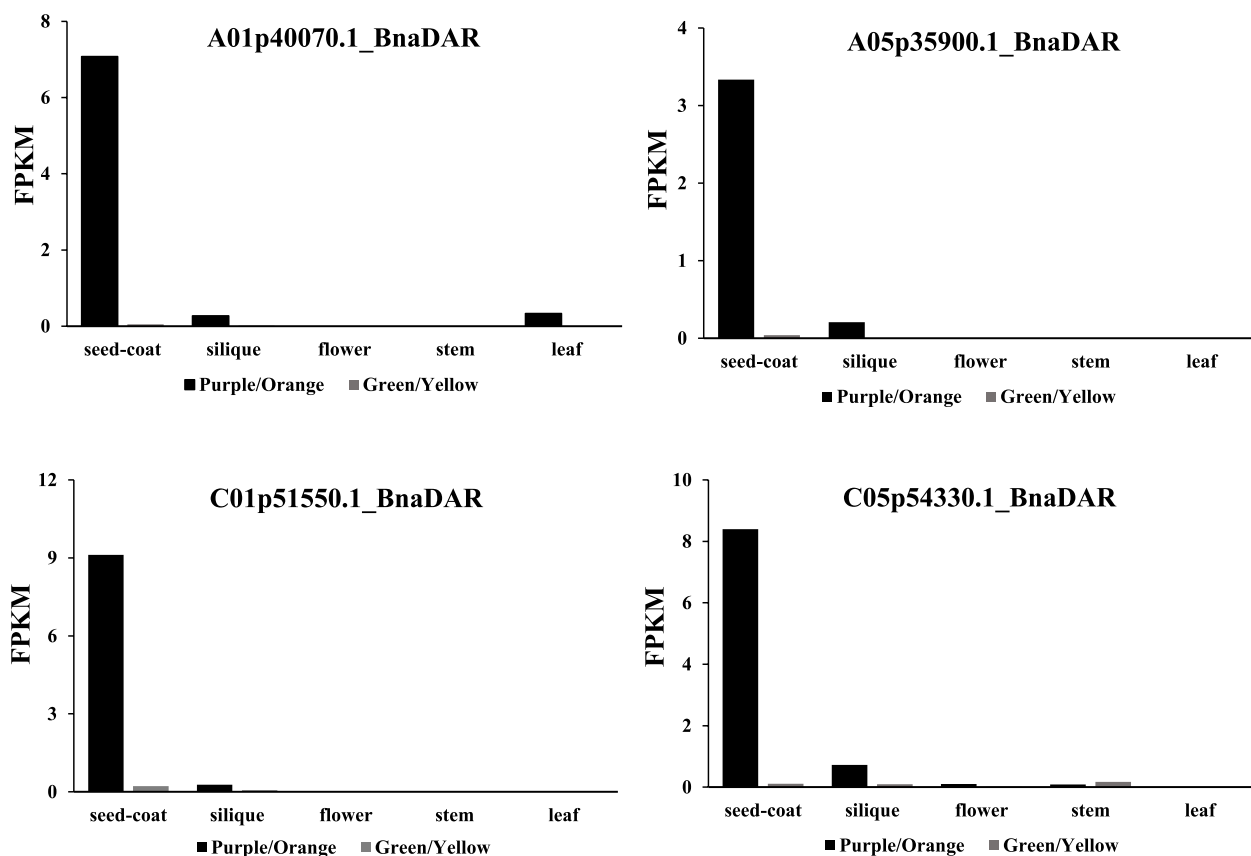
%) displayed a *Ks* value ranging from 0.5 to 0.99 (Supplementary Table S4). These findings suggest a substantial occurrence of synonymous mutations in *MYB5* homologous genes, indicating their remarkable conservation and the influence of robust purifying selection throughout the course of evolution.

Subsequently, an analysis was conducted on the selection pressure exerted on *MYB5* orthologs and paralogous genes (Fig. 3). The results revealed that the average values of *MYB5* orthologous genes were 0.08753, 0.41296, and 0.2308, whereas the average *Ka*, *Ks*, and *Ka/Ks* values of paralogous genes were 0.06259, 0.23772, and 0.30224 (Fig. 3A). These findings suggest that the synonymous mutation rate and non-synonymous mutation rate of orthologous genes are higher than those of paralogous genes, indicating a greater accumulation of variation during the evolutionary process. However, it is observed that the *Ka/Ks* value of paralogous genes surpasses that of orthologous genes, suggesting that paralogous genes have experienced more pronounced selection pressure and are subject to restricted rates of evolution. Subsequently, an examination of the correlation between *Ks*, *Ka*, and *Ka/Ks* values of orthologous and paralogous genes of *MYB5* was conducted (Fig. 3B). The findings indicate a positive correlation between the *Ka* value and *Ks* value in both

orthologous and paralogous genes. Conversely, a negative correlation is observed between the *Ka* value and *Ks* value with respect to the *Ka/Ks* values of both orthologous and paralogous genes. Furthermore, the strength of this correlation is primarily influenced by the *Ks* value. Overall, these results provide additional evidence supporting the notion that purifying selection primarily drives interspecies evolutionary pressure in *MYB5* homologous genes, with the orthologous genes being more prevalent abundant in the evolutionary process.

#### Tissue-specific expression profile of *MYB5* homologous genes in *B. napus*

In order to comprehensively investigate the expression profile of *MYB5* homologous genes, we conducted an analysis on the expression patterns of four specific *MYB5* homologous genes (*A01p40070.1\_BnaDAR*, *A05p35900.1\_BnaDAR*, *C01p51550.1\_BnaDAR*, and *C05p54330.1\_BnaDAR*) in five distinct tissues of *B. napus*. Based on their expression levels (Fig. 4), the expression patterns of *MYB5* homologous genes can be broadly categorized into three groups: high expression, low expression, and minimal expression. The expression levels of the four homologous genes of *MYB5* in *B. napus* varied across different tissues. They exhibited high



**Fig. 4** Expression analysis of four *MYB5* homologous genes in *Brassica napus* five different tissues. The FPKM (Fragments Per Kilobase Million) value obtained from transcriptome sequencing data analysis was used to represent the expression pattern of the four copies of the *MYB5* gene in different tissues, both five different tissues with three biological replicates

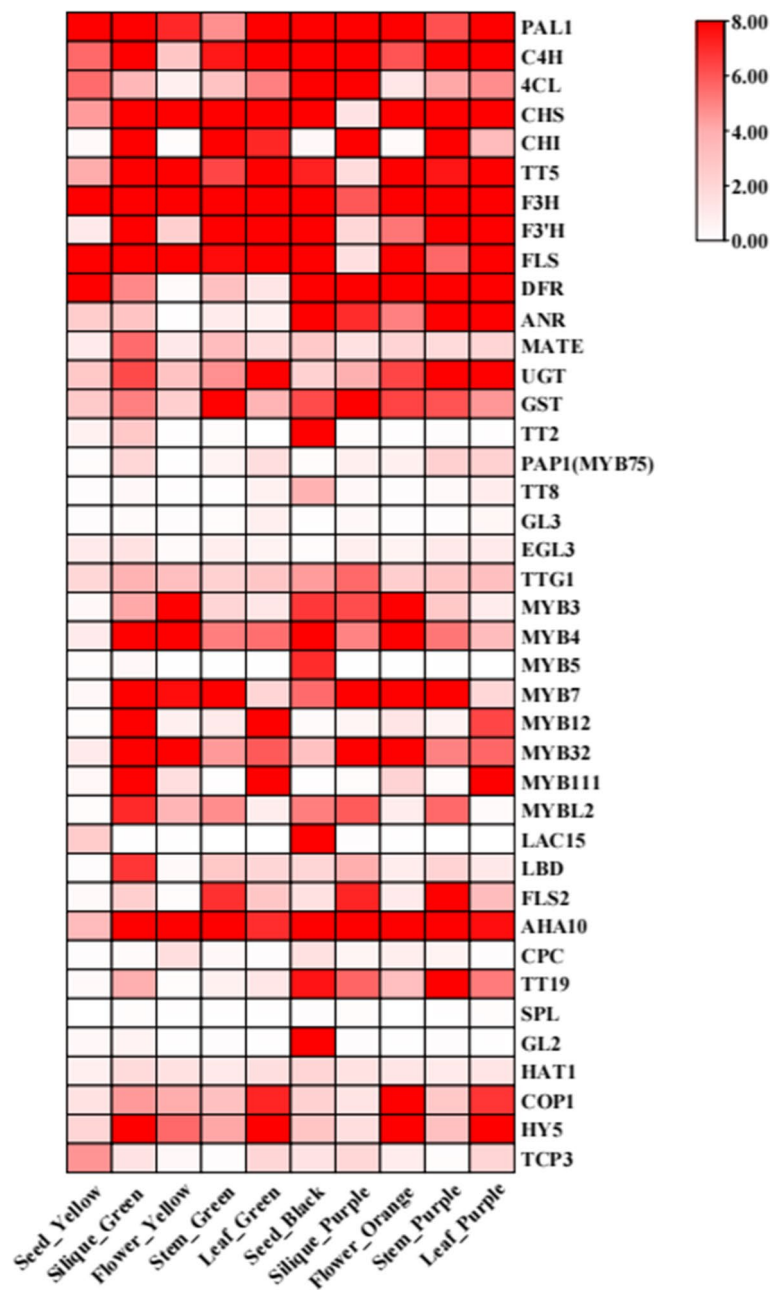
expression in purple seed coat, low expression in purple siliques, and minimal expression in orange-red petals and purple stems. Conversely, these four *MYB5* homologous genes showed negligible expression in yellow seed coats and petals, green leaves, stems, and siliques. Notably, *A01p40070.1\_BnaDAR* displayed low expression in purple leaves, whereas the remaining *MYB5* homologous copies exhibited minimal expression. The findings of this study suggest that the four *MYB5* homologous genes in *B. napus* exhibit predominant expression in the seed coat and likely play a role in regulating of PA biosynthesis in seeds. However, it is worth noting that the expression levels of these genes vary across different tissues.

#### Analysis of *MYB5* homologous genes expression patterns related to PA biosynthesis in *B. napus*

In order to gain a deeper understanding of the expression patterns of *MYB5* homologous genes associated with PA biosynthesis in *B. napus*, we conducted RNA-seq analysis on five different tissues: seed coat, pod, flower, stem, and leaf. Subsequently, we obtained

relevant FPKM expression data. By utilizing the anthocyanin biosynthesis genes previously reported in *A. thaliana* as a reference, we identified the PA-related genes in *B. napus* through a two-way BLAST analysis. Based on the expression data of the PA-related genes across the five tissues, we employed TBtools software to generate an expression heat map (Fig. 5, Supplementary Table S6). In nearly all tissues, there was notable upregulation of genes associated with the phenylpropane synthesis pathway and flavonol synthesis pathway, including *PAL1*, *C4H*, *4CL*, *CHS*, *CHI*, *F3H*, *F3H*, and *FLS*. Additionally, genes related to the biosynthesis of PA namely *DFR*, *ANS*, *ANR*, *UGT*, and *GST*, exhibited significant upregulation specifically in purple or orange organs where PA synthesis and accumulation occurred. Particularly, *ANR* (*BAN*) is highly expressed specifically in the black seed coat (Fig. 5). Furthermore, certain PA transporters, such as *TT19*, displayed high expression levels purple or orange-red tissues, while being either lowly expressed or virtually absent in yellow or green tissues. The transcriptional



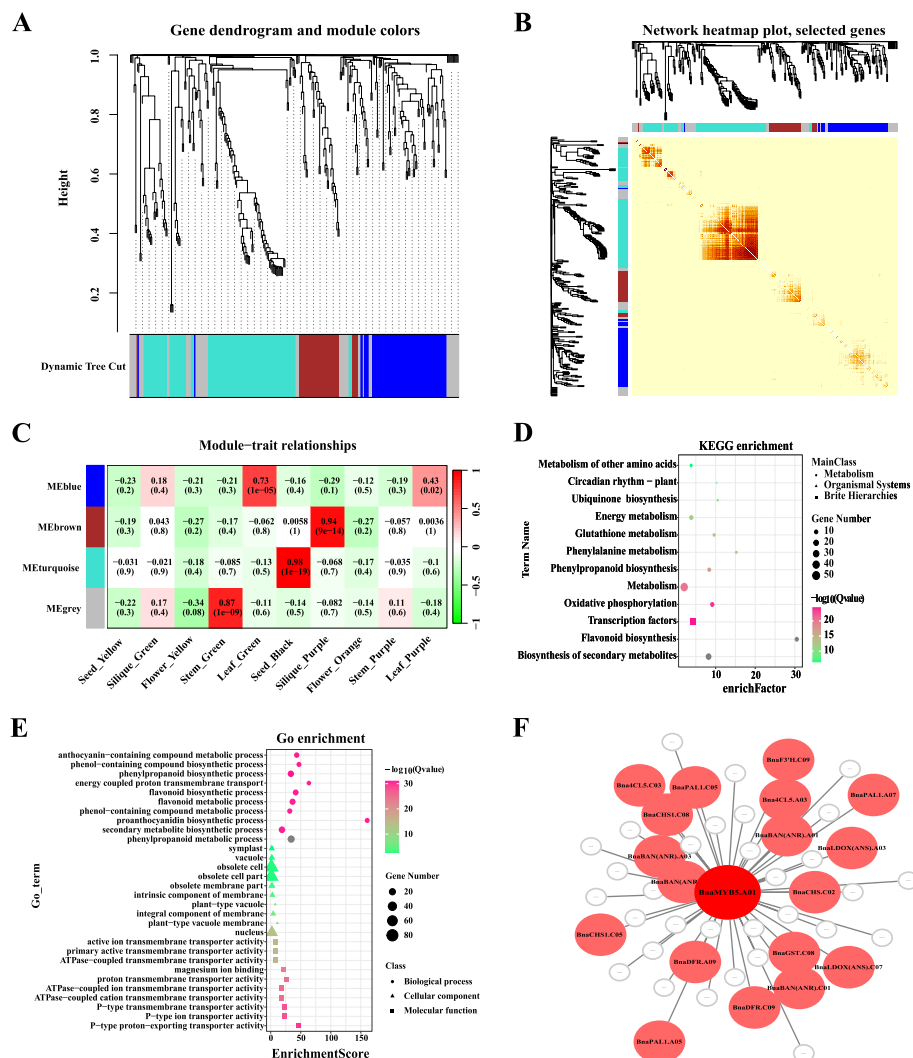


**Fig. 5** The expression patterns of anthocyanin-related genes in five tissues of *B. napus* were analyzed using a heat map that represents the FPKM values of these genes. The colors on the heat map, ranging from red to pink and white, indicate the expression levels from high to low

regulators *MYB5*, *TT2*, *TT8*, *LAC15*, and *GL2* associated with PA biosynthesis exhibited significant up-regulation in the purple-black seed coat, whereas their expression levels were comparatively low or negligible in other tissues (Fig. 5). These findings provide additional evidence supporting the specific expression in the seed coat, where it plays a regulatory role in PA biosynthesis and accumulation.

**WGCNA analysis of genes related to PA biosynthesis in different tissues of *B. napus***

In order to investigate the key genes associated with PA biosynthesis in various tissues of *B. napus*, we conducted a weighted gene co-expression network analysis (WGCNA) to examine the expression modules of PA biosynthesis-related genes across five tissues (Fig. 6, Supplementary Table S6). Specifically, we examined the



**Fig. 6** Co-expression analysis of genes related to anthocyanin biosynthesis was conducted in five tissues of *B. napus*. **A** Modular hierarchical clustering: The co-expression modules are depicted in different colors, while gray modules indicate no correlation between genes. **B** Module gene clustering heatmap: The gene expression network of anthocyanin biosynthesis-related genes in different tissues was analyzed using WGCNA, leading to the clustering of genes into distinct co-expression modules. **C** Module-to-sample correlation heatmap: Correlation analysis was performed between the co-expression modules of various genes associated with anthocyanin biosynthesis in different tissues. The numbers above the heat map indicate the Pearson correlation coefficient ( $r$ ) values. **D** KEGG enrichment analysis of 'turquoise' module. **E** GO enrichment analysis of 'turquoise' module. **F** Cytoscape representation of co-expression network of the hub gene in the 'turquoise' module

expression levels of PA synthesis-related genes in five tissues of *B. napus* and generated a topological overlapping heat map. The results of the co-expression network analysis revealed that the genes involved in PA biosynthesis in *B. napus* were categorized into four distinct co-expression modules (Fig. 6A). Following this, we conducted an analysis of the correlation heatmap between modules and samples. The results indicate that the 'turquoise' module exhibits the strongest connectivity (Fig. 6B). Additionally, when examining the correlation between modules and

PA biosynthesis-related genes across 5 tissues, we identified that 4 sample-specific modules (module-sample correlations  $> 0.65$  or  $< -0.65$ ,  $p$ -values  $< 0.05$ ). Notably, the 'blue' module (module-trait correlations = 0.73,  $p$ -values =  $1e-05$ ) in green leaves the 'brown' module (module-trait correlations = 0.94,  $p$ -values =  $9e-14$ ) in purple siliques, the 'turquoise' module (module-trait correlations = 0.98,  $p$ -values =  $1e-19$ ) in black seed coats and the 'grey' module (module-trait correlations = 0.87,  $p$ -values =  $1e-09$ ) in green stems display a significant positive

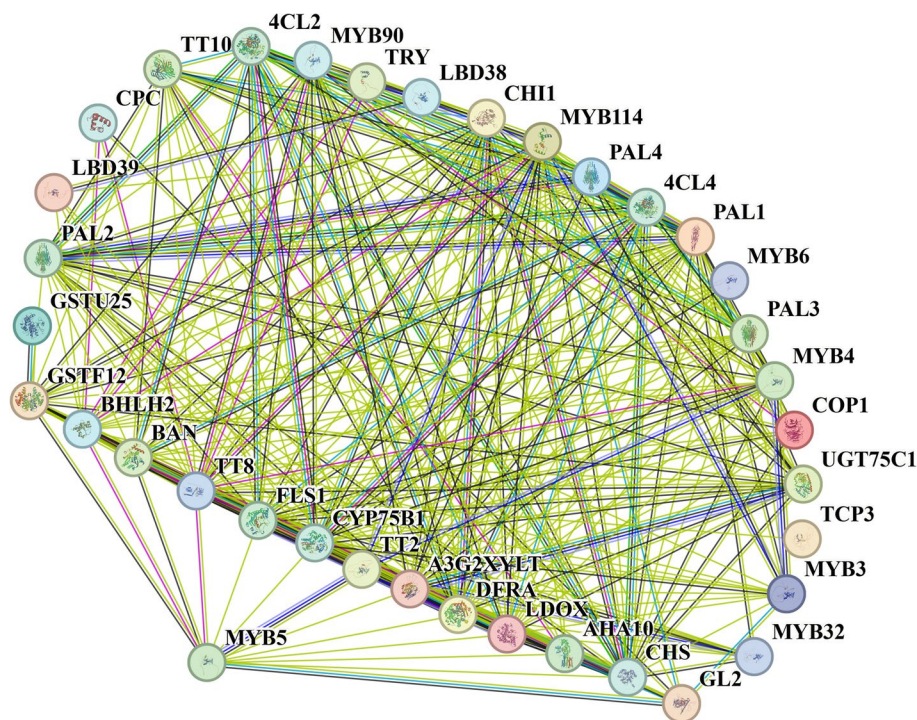
correlation (Fig. 6C). Significantly, the 'turquoise' module characterized by black seed coats exhibits the strongest correlation relationship suggesting a high degree of similarity in the expression patterns of genes involved in PA biosynthesis within the seed coat.

In addition, an analysis of GO and KEGG enrichment was conducted utilizing the 114 genes belonging to the 'turquoise' module (Supplementary Table S7). The KEGG enrichment outcomes indicate that metabolism and transcription factors exhibit the highest number of enriched genes, with the flavonoid biosynthesis entry displaying the highest enrich factor value (Fig. 6D). Similarly, the GO enrichment results reveal that the PA biosynthetic process entry attains the highest enrichment score value (Fig. 6E). These findings provide additional evidence supporting the close association between the 'turquoise' module and the metabolic regulation of PA biosynthesis. This suggests that the 'turquoise' module may exhibit the highest correlation with the development of black seed coat in rapeseed. In order to gain a deeper understanding of the relationship between genes within the module and the identification of hub genes (genes with high connectivity), a correlation network was constructed for the 'turquoise' module. In the 'turquoise' module, a total of 50 genes (copies) that are associated with the biosynthesis of proanthocyanidins (PAs) exhibited robust regulatory

relationships. These genes include *PLA1*, *4CL5*, *CHS1*, *F3'H*, *DFR*, *ANS*, *ANR* and *GST*, which are structural genes involved in the PA biosynthesis pathway. Additionally, there were 32 transcriptional regulators identified that are linked to PAs biosynthesis (Fig. 6F, Supplementary Table S8). Notably, MYB5 and BAN (ANR) demonstrated the highest degree of correlation (Supplementary Table S8). These highly correlated genes are referred to as hub genes within the gene co-expression network and play a critical role in the analysis of PAs biosynthesis regulation in the seed coat of *B. napus*.

#### Protein-protein interaction network analysis of 'turquoise' module genes

Protein-protein interaction networks serve as valuable and practical resources for systematically investigating the assembly of protein complexes or signaling cascades. To gain deeper insights into the association between the genes in the core module and the screen hub genes, the interaction network of the 'turquoise' module was predicted using the *A. thaliana* homologous gene map and the online STRING website (Fig. 7). The findings revealed direct interactions and connections between MYB5 and GSTF12, BHLH2, BAN(ANR), TT8, FLS1, CYP75B1, TT2, A3G2XYLT, DFRA, LDOX(ANS), AHA10, CHS(TT4) and GL2, with distinct interaction



**Fig. 7** Protein-protein interaction network analysis of 'turquoise' module genes. Interactions between the proteins are represented by connecting lines, with the line thickness reflecting the strength of the interaction

modes observed for GSTF12, GL2, and TT8 (Fig. 7). Specifically, MYB5 exhibits a robust and direct interaction with CHS, CHI, DFR, and ANS. Notably, CHS is early catalytic proteins involved in the biosynthesis pathway of PAs, facilitating the production of flavonoids. Conversely, LDOX(DFR) and BAN(ANR) serve as the late catalytic proteins in the PA biosynthesis pathway, enabling the conversion of flavonoids into PAs. This further substantiates the pivotal role of MYB5 as a key transcription factor governing PA biosynthesis in the seed coats of *B. napus*. Furthermore, MYB5 has the ability to directly engage with the bHLH-like transcription factor TT8, exhibiting two distinct interaction modes. This suggests that MYB5 potentially collaborates with TT8 to recruit WD40 protein, thereby facilitating the formation of a prototypical MBW transcriptional regulatory complex. This intricate network is believed to play a pivotal role in the involvement of transcriptional regulation of PAs.

## Discussion

Brassicaceae plants exhibit range of diversity and significant variation, with certain species having undergone processes such as hybridization, polyploidization, and breeding selection [48]. Consequently, these plants serve as excellent subjects for investigating alterations in the quantity of homologous genes, evolutionary patterns, and functional differentiation [48]. Currently, owing to advancements in sequencing technology and the decrease in sequencing expenses, a comprehensive compilation of reference genome data for 31 distinct species of Brassicaceae has been published [31]. These species encompass diverse evolutionary lineages and ploidy levels within the Brassicaceae family [31], thereby furnishing valuable data for our investigation into the evolutionary patterns of the MYB5 gene. The primary objective of this study is to perform a genome-wide identification and phylogenetic analysis of the MYB5 gene across Brassicaceae species possessing available genome information. The transcriptome data of various tissues of *B. napus* were utilized to conduct a comprehensive analysis of the tissue-specific expression of MYB5 and its role in the regulation of PA transcription. To achieve this objective, a systematic examination of MYB5 was performed, encompassing duplication events, genetic evolution, motif distribution, promoter cis-acting elements, tissue-specific expression, and expression patterns. The findings presented in this study reveal the identification of a total of 51 homologous genes of MYB5 from 31 within the Brassicaceae species family, with *B. napus* possesses 4 homologous copies. The analysis of expression patterns revealed a significant association MYB5 and genes related to PA biosynthesis in the seed coat of *B. napus*. All 4 homologous genes exhibited specific and high expression levels, suggesting that MYB5 may serve as a crucial

transcription factor in regulating PA biosynthesis in the seed coat. These findings will be further discussed.

## Evolution of MYB5 in Brassicaceae species

Over the course of their evolutionary trajectory, Brassicaceae species have experienced diploidization, triploidization, and natural hybridization occurrences, leading to a wide array of variations, particularly within the six *Brassica* crops situated in the U's triangle [48–52]. The evolution of MYB5 in the Brassicaceae was found to be synchronized with the overall genome evolution. This synchronization was observed in conjunction with polyploidization and hybridization events, which led to changes in both the copy number and gene function of homologous genes (Fig. 1). Our findings indicate that among the 31 species included in our study, the distribution of MYB5 homologous genes varied. Specifically, 21 species had a single copy, 4 species had two copies, 3 species in 3 copies, 2 species had 1 copy, and 1 species had 5 copies (Table 1). These variations in copy number of MYB5 homologous genes are closely associated with polyploidization and hybridization events. In addition, an analysis was conducted on the alterations in copy number and evolutionary relationship of MYB5 homologous genes across six species of *Brassica*. The diploid species *B. rapa* (AA genome), *B. oleracea* (CC genome) and *B. nigra* (BB genome) exhibited copy numbers of 1, 2 and 3, respectively. Conversely, the allotetraploid species *B. juncea* (AABB genome), *B. napus* (AACC genome) and *B. carinata* (BBCC genome) displayed copy numbers of 5, 4, and 4, respectively (Table 1). The findings indicate that the allotetraploid species' genome undergoes intricate chromosomal rearrangements rather than a straightforward combination of two diploid species. Furthermore, the constructed phylogenetic tree, based on MYB5 homologous proteins, exhibits a strong alignment with the evolutionary relationship of Brassicaceae as reported by Nikolov et al. [47].

Furthermore, an analysis was conducted on the conserved domains of 51 homologous proteins of MYB5, a transcription factor belonging to the R2R3-MYB type. Subsequently, an examination was performed to assess the selection pressure on both orthologs and paralogs of MYB5 (Fig. 3). The findings revealed that MYB5 orthologous genes were present in 31 species of the Brassicaceae family, while paralogous genes were identified in 10 species of the same family (Supplementary Table S2). The Ks value of both orthologous and paralogous genes exhibited a significantly higher magnitude compared to the Ka value, suggesting a pronounced prevalence of synonymous substitutions in the MYB5 homologous gene. This observation implies a relatively conservative evolutionary pattern. Furthermore, the Ka/Ks values for all 2467



gene pairs were consistently less than 1, indicating that the interspecies evolutionary pressure on *MYB5* homologous genes primarily stemmed from purifying selection. Additionally, the variation in orthologous genes appeared to be more abundant throughout the course of evolution. Furthermore, the results of the correlation analysis for *Ka*, *Ks*, and *Ka/Ks* values provided additional evidence that the correlation coefficient of orthologous genes was significantly higher than that of paralogous genes. This finding suggests that the synonymous substitution of orthologous genes played a more substantial role in the evolutionary process of *MYB5*.

#### ***MYB5* may specifically regulate the formation of seed coat color**

In Brassicaceae species, the seed coat colors predominantly consist of yellow and black, with some intermediate shades. Yellow seeds are generally regarded as superior in quality due to their thinner shells, higher oil and protein content, and lower levels of fiber, pigments, and polyphenols [53]. Breeding yellow-seeded rapeseed is a significant objective within the realm of rapeseed breeding. Over the course of several decades, researchers have conducted extensive investigations into the formation and regulatory mechanisms underlying the yellow seed coat [24, 28–30, 54–56]. The transition from yellow to black seed coat primarily occurs as a result of pigment deposition, particularly proPAs [57]. Consequently, a thorough analysis of the transcriptional regulation mechanism governing seed coat PAs is crucial for the successful breeding of yellow rapeseed. Xu et al. [18] conducted a comprehensive investigation on the seed coat color formation mechanism in the model plant *A. thaliana*. Within the PA synthesis pathway, only *DFR*, *ANS(LDOX)*, *BAN*, *TT12*, *TT19* and *AHA10* were identified as direct targets of the MBW complex. Notably, the TT2-TT8-TTG1 complex emerged as a key player in seed development, while MYB5-TT8-TTG1, TT2-EGL3-TTG1 and TT2-GL3-TTG1 also exhibited tissue-specific involvement [18]. In *Brassica* crops, *TT1* [27], *TT2* [28, 58], *TT8* [21, 29, 56, 59, 60] and *TT10* [26, 61] were reported to be involved in the regulation of seed coat color formation in multiple species. The findings of this study demonstrate that certain R2R3-MYB-type transcription factors (such as *TT1*, *TT2*, *TT8*, *TT10*) facilitate the recruitment of bHLH-type transcription factors (*TT8*, *GL3*, *EGL3*) and the TTG1 protein. This collaborative assembly forms a MBW complex that specifically targets the late structural genes involved in the synthesis pathway of PAs. Consequently, this regulatory mechanism plays a crucial role in the biosynthesis and accumulation of PAs, which ultimately determines the formation of the black seed coat.

In the Brassicaceae family, *MYB5* was initially identified as being specifically expressed in *A. thaliana* seeds and playing a role in the regulation of seed coat color [18]. In *Medicago truncatula*, the interaction between *MtMYB5* and *MtMYB14* (a homolog of *TT2*) was found to synergistically activate anthocyanin reductase, thereby regulating procyanidin biosynthesis [62]. Furthermore, *MYB5* was also implicated in the regulation of the heat stress response in *A. thaliana* [63]. Interestingly, our previous study revealed differential expression of the R2R3-MYB transcription factor *MYB90* in the leaves of six *Brassica* species. This factor potentially interacts with *TT8* to regulate the expression of *F3H*, *LDOX*, *ANS*, and *UFGT*, thereby facilitating the bioavailability of anthocyanins biosynthesis [64]. Moreover, in the seed coats of these six *Brassica* crops, the R2R3-MYB transcription factors *MYB5* and *TT2* exhibited differential expression, suggesting that they may serve as pivotal genes in the regulation of seed coat color in *Brassica* crops [30]. In this study, it was observed that four homologous copies of *MYB5* exhibited specific expression in the seed coat of rapeseed, as determined through tissue-specific analysis (Fig. 4). Both expression pattern analysis and WGCNA revealed a close association with and strong co-expression between *MYB5* and *TT8*, as well as late structural genes *ANR* involved in anthocyanin synthesis (Figs. 6 and 7). The findings from protein interaction network analysis further supported the direct targeting of transcription factors *TT8* and *GL2* by *MYB5*, along with structural genes such as *CHS*, *CHI*, *DFR*, *ANS* (*LDOX*) and *BAN(ANR)*. Consequently, *MYB5* may specifically regulate *BAN(ANR)* expression and thus actively participate in the biological regulation of PAs. The findings of this study suggest that within *Brassica* crops, the production and buildup of anthocyanins in various tissues may exhibit tissue-specific patterns, with *MYB5* potentially exerting a significant influence on the control of seed coat color development. These findings necessitate additional validation through subsequent experimental investigations. In conclusion, the outcomes of this research contribute novel insights into the exploration of the mechanism underlying the variation in rape seed coat color mediated by *MYB5*.

#### **Conclusions**

In this study, a total of 51 *MYB5* homologous genes were identified from 31 species within the Brassicaceae family, and through phylogenetic analysis, they were categorized into 4 distinct subfamilies. The analysis of *Ka/Ks* ratios revealed that the interspecific evolution of the *MYB5* gene is primarily driven by purifying selection acting on homologous genes, with *Ks* exerting a greater influence on the selection pressure. By investigating the expression



of the *MYB5* gene expression in various tissues of *B. napus*, including stems, leaves, petals, siliques and seed coats, we hypothesized that *MYB5* plays a specific role in regulating PAs biosynthesis in the seed coats. The findings from expression pattern analysis and WGCNA further supported the proposition that *MYB5* exhibits high and specific expression in the seed coat and is closely associated with genes involved in PA biosynthesis. At the same time, by searching for module hub genes, our findings indicate a strong correlation and regulatory relationship between *MYB5* and *BAN(ANR)*. Additionally, the analysis of the protein-protein interaction network revealed significant interactions between *MYB5* and *BAN(ANR)*, *LDOX(ANS)*, *TT2*, *TTT8*, and other structural genes and transcriptional regulatory proteins involved in the PA biosynthetic pathway. These results suggest that *MYB5* may be specifically expressed in the seed coat of *B. napus* to drive the expression of *BAN(ANR)*, facilitate PA biosynthesis, and regulate the color formation of rapeseed seed coat. Nevertheless, additional experimental validation is imperative to confirm the veracity of these findings.

#### 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
UGT	UDP-flavonoid glucosyltransferase
TT2	TRANSPARENT TESTA 2

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-023-04718-6>.

**Additional file 1: Supplementary Table 1.** MYB5 homologous and their sequence in 31 Brassicaceae species.

**Additional file 2: Supplementary Table 2.** MYB5 homologous in 31 Brassicaceae species.

**Additional file 3: Supplementary Table 3.** Cis-element analysis of MYB5.

**Additional file 4: Supplementary Table 4.** Selection pressure analysis of MYB5.

**Additional file 5: Supplementary Table 5.** The expression patterns of anthocyanin-related genes in different tissues of *B. napus*.

**Additional file 6: Supplementary Table 6.** The FPKM values in different tissues of *B. napus*.

**Additional file 7: Supplementary Table 7.** The genes of 'turquoise' module.

**Additional file 8: Supplementary Table 8.** The hub gene in the 'turquoise' module.

#### Acknowledgments

Thanks to Professor Zaiyun Li and Xianhong Ge for his help in the writing process of the thesis. We would also like to thank the database, website and software developers who facilitated the smooth development of this study.

#### 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.

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

#### Authors' contributions

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

#### Funding

This research was funded by the Key Research and Development Program of Jiangxi Province, grant number: 20223BBF61002; the Natural Science Foundation of Jiangxi Province, grant number: 20224BAB215016; Jiangxi Province Key Laboratory of Oilcrops Biology, grant number: YLKFKT202202; Jiangxi Province Postgraduate Innovation Special Fund Project, grant number: YC2022-s929.

#### Availability of data and materials

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, 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

<sup>1</sup>College of Life Sciences, Ganzhou Key Laboratory of Greenhouse Vegetable, Gannan Normal University, Ganzhou 341000, China. <sup>2</sup>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.

Received: 13 December 2023 Accepted: 31 December 2023

Published online: 16 January 2024

#### References

1. 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.
2. Paz-Ares J, Ghosal D, Wienand U, Peterson PA, Saedler H. The regulatory c1 locus of Zea mays encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. Embo J. 1987;6(12):3553–8.
3. Urao T, Noji M, Yamaguchi-Shinozaki K, Shinozaki K. A transcriptional activation domain of ATMYB2, a drought-inducible Arabidopsis Myb-related protein. Plant J. 1996;10(6):1145–8.
4. Gubler F, Watts RJ, Kalla R, Matthews P, Keys M, Jacobsen JV. Cloning of a rice cDNA encoding a transcription factor homologous to barley GAMyb. Plant Cell Physiol. 1997;38(3):362–5.

5. Kobayashi S, Ishimaru M, Hiraoka K, Honda C. Myb-related genes of the Kyoho grape (*Vitis labruscana*) regulate anthocyanin biosynthesis. *Planta*. 2002;215(6):924–33.
6. Takos AM, Jaffe FW, Jacob SR, Bogs J, Robinson SP, Walker AR. Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. *Plant Physiol*. 2006;142(3):1216–32.
7. Avila J, Nieto C, Canas L, Benito MJ, Paz-Ares J. *Petunia hybrida* genes related to the maize regulatory C1 gene and to animal myb proto-oncogenes. *Plant J*. 1993;3(4):553–62.
8. Gonzalez A, Zhao M, Leavitt JM, Lloyd AM. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J*. 2008;53(5):814–27.
9. Peng Y, Lin-Wang K, Cooney JM, Wang T, Espley RV, Allan AC. Differential regulation of the anthocyanin profile in purple kiwifruit (*Actinidia* species). *Hortic Res*. 2019;6:3.
10. Cominelli E, Galbiati M, Vavasseur A, Conti L, Sala T, Vuylsteke M, Leonhardt N, Dellaporta SL, Tonelli C. A guard-cell-specific MYB transcription factor regulates stomatal movements and plant drought tolerance. *Curr Biol*. 2005;15(13):1196–200.
11. Seo PJ, Xiang F, Qiao M, Park JY, Lee YN, Kim SG, Lee YH, Park WJ, Park CM. The MYB96 transcription factor mediates abscisic acid signaling during drought stress response in *Arabidopsis*. *Plant Physiol*. 2009;151(1):275–89.
12. Kirik V, Lee MM, Wester K, Herrmann U, Zheng Z, Oppenheimer D, Schiefelbein J, Hulskamp M. Functional diversification of MYB23 and GL1 genes in trichome morphogenesis and initiation. *Development*. 2005;132(7):1477–85.
13. Li SF, Milliken ON, Pham H, Seyit R, Napoli R, Preston J, Koltunow AM, Parish RW. The *Arabidopsis* MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. *Plant Cell*. 2009;21(1):72–89.
14. Kang YH, Kirik V, Hulskamp M, Nam KH, Hagley K, Lee MM, Schiefelbein J. The MYB23 gene provides a positive feedback loop for cell fate specification in the *Arabidopsis* root epidermis. *Plant Cell*. 2009;21(4):1080–94.
15. Lepiniec L, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, Nesi N, Caboche M. Genetics and biochemistry of seed flavonoids. *Annu Rev Plant Biol*. 2006;57:405–30.
16. Gonzalez A, Mendenhall J, Huo Y, Lloyd A. TTG1 complex MYBs, MYB5 and TT2, control outer seed coat differentiation. *Dev Biol*. 2009;325(2):412–21.
17. Mandaokar A, Browse J. MYB108 acts together with MYB24 to regulate jasmonate-mediated stamen maturation in *Arabidopsis*. *Plant Physiol*. 2009;149(2):851–62.
18. Xu WJ, Grain D, Bobet S, Le Gourrierier 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.
19. Koch MA, German DA, Kiefer M, Franzke A. Database Taxonomics as Key to Modern Plant Biology. *Trends Plant Sci*. 2018;23(1):4–6.
20. Wittkop B, Snowdon RJ, Friedt W. New NIRS calibrations for fiber fractions reveal broad genetic variation in *Brassica napus* seed quality. *J Agric Food Chem*. 2012;60(9):2248–56.
21. Patel MK, Chaudhary R, Taak Y, Pardeshi P, Nanjundan J, Vinod KK, Saini N, Vasudev S, Yadava DK. Seed coat colour of Indian mustard [(L.) Czern. and Coss.] is associated with homologs identifiable by targeted functional markers. *Front Plant Sci*. 2022;13:1012368.
22. Liu F, Chen H, Yang L, You L, Ju J, Yang S, Wang X, Liu Z. QTL Mapping and Transcriptome Analysis Reveal Candidate Genes Regulating Seed Color in *Brassica napus*. *Int J Mol Sci*. 2023;24(11):9262.
23. Qu C, Zhu M, Hu R, Niu Y, Chen S, Zhao H, Li C, Wang Z, Yin N, Sun F, et al. Comparative genomic analyses reveal the genetic basis of the yellow-seed trait in *Brassica napus*. *Nat Commun*. 2023;14(1):5194.
24. Zhang Y, Qin Y, Li D, Wang W, Gao X, Hao C, Feng H, Wang Y, Li T. Fine mapping and cloning of a novel BrSCC1 gene for seed coat color in *Brassica rapa* L. *Theor Appl Genet*. 2023;136(1):11.
25. Chen YY, Lu HQ, Jiang KX, Wang YR, Wang YP, Jiang JJ. The Flavonoid Biosynthesis and Regulation in *Brassica napus*: A Review. *Int J Mol Sci*. 2022;24(1):357.
26. Zhang K, Lu K, Qu C, Liang Y, Wang R, Chai Y, Li J. Gene silencing of BnTT10 family genes causes retarded pigmentation and lignin reduction in the seed coat of *Brassica napus*. *Plos One*. 2013;8(4):e61247.
27. Lian J, Lu X, Yin N, Ma L, Lu J, Liu X, Li J, Lu J, Lei B, Wang R, 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.
28. Xie T, Chen X, Guo T, Rong H, Chen Z, Sun Q, Batley J, Jiang J, Wang Y. Targeted Knockout of BnTT2 Homologues for Yellow-Seeded *Brassica napus* with Reduced Flavonoids and Improved Fatty Acid Composition. *J Agric Food Chem*. 2020;68(20):5676–90.
29. Zhai Y, Yu K, Cai S, Hu L, Amoo O, Xu L, Yang Y, Ma B, Jiao Y, Zhang C, 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.
30. 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.
31. 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.
32. Wheeler TJ, Eddy SR. nhmmer: DNA homology search with profile HMMs. *Bioinformatics*. 2013;29(19):2487–9.
33. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44(D1):D279–285.
34. Pucker B. Automatic identification and annotation of MYB gene family members in plants. *Bmc Genomics*. 2022;23(1):220.
35. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80.
36. 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.
37. 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.
38. 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.
39. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res*. 2002;30(1):325–7.
40. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant*. 2020;13(8):1194–202.
41. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs\_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics*. 2010;8(1):77–80.
42. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet*. 2002;18(9):486.
43. Cao Y, Fan T, Wang L, Zhang L, Li Y. Large-scale analysis of putative Euphorbiaceae R2R3-MYB transcription factors identifies a MYB involved in seed oil biosynthesis. *BMC Plant Biol*. 2023;23(1):145.
44. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
45. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT StringTie and Ballgown. *Nat Protoc*. 2016;11(9):1650–67.
46. Yang TH, Cai BW, Jia ZB, Wang Y, Wang J, King GJ, Ge XH, Li ZY. Genomes provide insights into whole-genome triplication and divergence patterns within tribe Brassicaceae. *Plant J*. 2023;113(2):246–61.
47. 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.
48. Chalhou 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.
49. Cai X, Chang L, Zhang T, Chen H, Zhang L, Lin R, Liang J, Wu J, Freeling M, Wang X. Impacts of allopolyploidization and structural variation on intraspecific diversification in *Brassica rapa*. *Genome Biol*. 2021;22(1):166.

50. 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.
51. 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–+.
52. 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.
53. 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.
54. Albert S, Delseny M, Devic M. BANYULS, a novel negative regulator of flavonoid biosynthesis in the Arabidopsis seed coat. *Plant J*. 1997;11(2):289–99.
55. Sabharwal V, Negi MS, Banga SS, Lakshmikumaran M. Mapping of AFLP markers linked to seed coat colour loci in Brassica juncea (L.) Czern. *Theor Appl Genet*. 2004;109(1):160–6.
56. Padmaja LK, Agarwal P, Gupta V, Mukhopadhyay A, Sodhi YS, Pental D, Pradhan AK. 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.
57. Jia L, Wu Q, Ye N, Liu R, Shi L, Xu W, Zhi H, Rahman AN, Xia Y, Zhang J. Proanthocyanidins inhibit seed germination by maintaining a high level of abscisic acid in Arabidopsis thaliana. *J Integr Plant Biol*. 2012;54(9):663–73.
58. Zhou L, Li Y, Hussain N, Li Z, Wu D, Jiang L. 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.
59. Li X, Chen L, Hong M, Zhang Y, Zu F, Wen J, Yi B, Ma C, Shen J, Tu J, et al. A large insertion in bHLH transcription factor BrTT8 resulting in yellow seed coat in Brassica rapa. *Plos One*. 2012;7(9):e44145.
60. Chen M, Xuan L, Wang Z, Zhou L, Li Z, Du X, Ali E, Zhang G, Jiang L. TRANSPARENT TESTA8 inhibits seed fatty acid accumulation by targeting several seed development regulators in arabidopsis. *Plant Physiol*. 2014;165(2):905–16.
61. David LC, Dechorgnat J, Berquin P, Routaboul JM, Debeaujon I, Daniel-Vedele F, Ferrario-Méry S. Proanthocyanidin oxidation of seeds is altered in mutant of the high-affinity nitrate transporter NRT2.7. *J Exp Bot*. 2014;65(3):885–93.
62. Liu C, Jun JH, Dixon RA. MYB5 and MYB14 Play Pivotal Roles in Seed Coat Polymer Biosynthesis in Medicago truncatula. *Plant Physiol*. 2014;165(4):1424–39.
63. Jacob P, Brisou G, Dalmais M, Thevenin J, van der Wal F, Latrasse D, Suresh Devani R, Benhamed M, Dubreucq B, Boualem A, et al. The Seed Development Factors TT2 and MYB5 Regulate Heat Stress Response in Arabidopsis. *Genes (Basel)*. 2021;12(5):746.
64. 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.

## Publisher's Note

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