Theoretical and Applied Genetics (2022) 135:2083–2099
https://doi.org/10.1007/s00122-022-04099-8

ORIGINAL ARTICLE

Multi-omics analysis reveals the mechanism of seed coat color formation in *Brassica rapa* L.

Huiyan Zhao[1] · Guoxia Shang[2] · Nengwen Yin[1] · Si Chen[1] · Shulin Shen[1] · Haiyan Jiang[1] · Yunshan Tang[1] · Fujun Sun[1] · Yuhan Zhao[1] · Yongchao Niu[3] · Zhi Zhao[2] · Liang Xu[2] · Kun Lu[1,4,5] · Dezhi Du[2] · Cunmin Qu[1,4,5] · Jiana Li[1,4,5]

Received: 16 December 2021 / Accepted: 8 April 2022 / Published online: 23 May 2022

Abstract

Keymessage **Multi-omics analysis of the transcriptome, metabolome and genome identified major and minor loci and candidate genes for seed coat color and explored the mechanism of flavonoid metabolites biosynthesis in *Brassica rapa*.**

Abstract Yellow seed trait is considered an agronomically desirable trait with great potential for improving seed quality of *Brassica* crops. Mechanisms of the yellow seed trait are complex and not well understood. In this study, we performed an integrated metabolome, transcriptome and genome-wide association study (GWAS) on different *B. rapa* varieties to explore the mechanisms underlying the seed coat color formation. A total of 2,499 differentially expressed genes and 116 differential metabolites between yellow and black seeds with strong association with the flavonoid biosynthesis pathway was identified. In addition, 330 hub genes involved in the seed coat color formation, and the most significantly differential flavonoids biosynthesis were detected based on weighted gene co-expression network analysis. Metabolite GWAS analysis using the contents of 42 flavonoids in developing seeds of 159 *B. rapa* lines resulted in the identification of 1,626 quantitative trait nucleotides (QTNs) and 37 chromosomal intervals, including one major locus on chromosome A09. A combination of QTNs detection, transcriptome and functional analyses led to the identification of 241 candidate genes that were associated with different flavonoid metabolites. The flavonoid biosynthesis pathway in *B. rapa* was assembled based on the identified flavonoid metabolites and candidate genes. Furthermore, *BrMYB111* members (*BraA09g004490.3C* and *BraA06g034790.3C*) involved in the biosynthesis of taxifolin were functionally analyzed in vitro. Our findings lay a foundation and provide a reference for systematically investigating the mechanism of seed coat color in *B. rapa* and in the other plants.

Communicated by Maria Laura Federico.

Huiyan Zhao, Guoxia Shang contributed equally to this paper.

✉ Dezhi Du
qhurape@126.com

✉ Cunmin Qu
drqucunmin@swu.edu.cn

✉ Jiana Li
ljn1950@swu.edu.cn

[1] Chongqing Engineering Research Center for Rapeseed, College of Agronomy and Biotechnology, Southwest University, Chongqing 400716, China

[2] Laboratory for Research and Utilization of Qinghai Tibet Plateau Germplasm Resources, Key Laboratory of Spring Rapeseed Genetic Improvement, Academy of Agricultural and Forestry Sciences, Qinghai University, Xining 810016, Qinghai, China

[3] Biozeron Shenzhen, Inc, Shenzhen, China

[4] Academy of Agricultural Sciences, Germplasm Creation Special Program of Southwest University, Southwest University, Chongqing 400715, China

[5] Engineering Research Center of South Upland Agriculture, Ministry of Education, Chongqing 400715, China

Introduction

Field mustard (*Brassica rapa* L.) is one of the most important oil crops in the world. The yellow seed trait is a desirable quality trait in the breeding of *B. rapa*, as compared to black-seeded germplasm within the similar genetic background, yellow seeds not only contain more oil for human consumption, but also possess more protein and less husk and secondary metabolites, as more edible meal for livestock (Daun 1988; Jiang et al. 2015; Jönsson 1977). Germplasm with yellow seeds has been collected or created in various *Brassicaceae* varieties, including *B. rapa*, rapeseed (*Brassica napus* L.), brown mustard (*Brassica juncea* L.) and Ethiopian mustard (*Brassica carinata* L.). Numerous studies, including quantitative trait loci (QTL) mapping and candidate genes cloning, have revealed that the seed color was controlled by few major QTLs in *Brassica* crops (Marles et al. 2003; Padmaja et al. 2014; Rahman et al. 2010; Schwetka 1982; Xiao et al. 2012). Among them, several homologs to Arabidopsis *transparent testa* (*TT*) genes affecting seed color, such as *TT1*, *TT8*, *TT2* and *TTG1* (*TRANSPARENT TESTA GLABRA1*), have been identified, and some of them have been functionally validated in *B. rapa*, *B. juncea* and *B. napus*. Nevertheless, the exact underlying regulatory mechanism of seed coat color is still obscure (Li et al. 2012; Padmaja et al. 2014; Wang et al. 2017; Xie et al. 2020; Zhai et al. 2020; Zhang et al. 2009).

Histological and histochemical analyses show that the lack of pigment deposition in the seed endothelium is the main reason for the yellow seed trait in *Brassica* crops and *tt* mutants in Arabidopsis (Auger et al. 2010; Debeaujon et al. 2003; Li et al. 2012; Qu et al. 2013). Furthermore, proanthocyanidins and anthocyanidins derived from the flavonoid biosynthesis pathway play critical roles as pigments in the seed coat color formation (Auger et al. 2010; Debeaujon et al. 2003; Jiang et al. 2013; Qu et al. 2013). Indeed, analyses of differences in flavonoid contents of yellow and black seeds for several *Brassica* crops and *Arabidopsis* measured by liquid chromatography-mass spectrometry (LC/MS), support the involvement of these phenolic compounds in the establishment of seed coat color (Auger et al. 2010; Qu et al. 2013, 2020; Ren et al. 2021; Routaboul et al. 2006; Shao et al. 2014). These differences in flavonoid contents were also consistent with the difference in the transcript levels of some flavonoid biosynthesis genes (Jiang et al. 2013; Lian et al. 2017; Qu et al. 2013; Wang et al. 2017), and recent transcriptome data generated by transcriptome deep sequencing (RNA-seq) of several *Brassica* crops with yellow and black seeds (Hong et al. 2017; Liu et al. 2013; Niu et al. 2020; Qu et al. 2020; Ren et al. 2021; Zhao et al. 2019). To date, however, there have been few joint analyses of the transcriptome and the metabolome for the seed coat color trait in *Brassica* crops.

Genes controlling the seed coat color have been mapped in different *B. rapa* varieties collected around the world using segregating populations derived from crosses between species with yellow or black seeds. Seed coat color has traditionally been determined by visual inspection, near-infrared spectroscopy (NIRS) or light reflectometry, with limited accuracy (Kebede et al. 2012; Lou et al. 2007; Somers et al. 2001; Stein et al. 2013; Teutonico and Osborn 1994). As the seed coat color represents a continuum in some *B. rapa* mapping populations, coarse phenotyping may affect the accuracy of mapping results. Development of metabolomics offers a solution to turn metabolites into highly quantifiable traits and decompose complex traits into related metabolites (Chen et al. 2014; Shirai et al. 2017; Zeng et al. 2020). Thus, a combination of transcriptomic and metabolomics analyses may increase the sensitivity and accuracy of the genetic mapping and provide a more comprehensive picture of the underlying regulatory processes.

In this study, we profiled the transcriptome and metabolome of several *B. rapa* varieties producing extreme yellow or black seeds. In addition, we detected the relative content of flavonoids in a large population of *B. rapa* accessions with different seed coat colors and implemented GWAS analysis to identify candidate genes for the flavonoid biosynthesis pathway in *B. rapa*. Furthermore, *BrMYB111s* were firstly verified involved in the biosynthesis of taxifolin. Our findings lay a foundation for elucidating the regulatory mechanisms of seed coat color in *B. rapa* and provide important resources for a comparative and integrated analysis of seed coat color trait in related *Brassica* species.

Materials and methods

Plant materials and sampling

For this study, 159 spring *B. rapa* accessions were collected from China and other countries (Supplementary Table S1), including two varieties with extremely yellow seeds (BrY1, Dahuang; BrY2, Yellow Sarson) and two varieties with extremely black seeds (BrB1, Luhua; BrB2, Mayu), which represent typical yellow-and black-seeded genotypes (Fig. 1). The *B. rapa* materials were grown in 2018 (April to August) and 2020 (April to August) in Xining (N36°43′, E101°45′, Qinghai province, China), and in 2018(November)–2019(April) in Beibei (N29°76′, E106°38′, Chongqing, China) under normal field conditions, respectively. Developing seeds of the four extreme yellow- and black-seeded varieties were sampled by combining

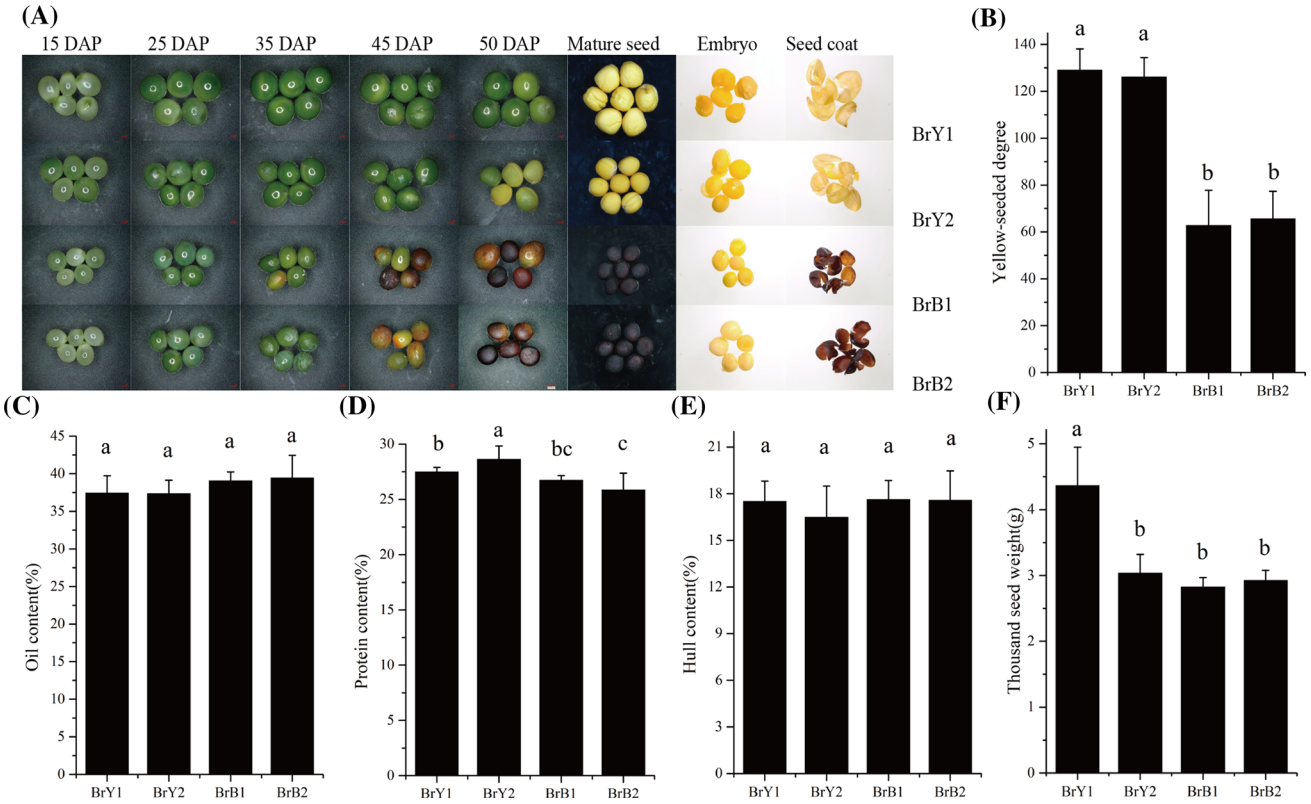


Fig. 1 Characterizations analysis of seeds between the yellow-and black-seeded *B. rapa*. **A** Phenotype analysis of seeds in the yellow- and black-seeded *B. rapa* during seed development. DAP, days after pollination. BrY1 and BrY2, Yellow-seeded *B. rapa* lines Dahuang and Sarson; BrB1 and BrB2, Black-seeded *B. rapa* lines Luhua and Mayu; **B** Yellow-seeded degree; **C** oil content (%); **D** protein content (%); **E** hull content (%); **F** thousand-seed weight (gram)

developing seeds (15, 25, 35, 45 and 50 days after pollination [DAP]) from five individual plants in 2018, respectively. Developing seeds for all varieties were sampled at 35 DAP for 3 years, as well as leaves at the seedling stage (30 days after germination) in 2018. All samples were rapidly frozen and stored at −80 °C until extraction of RNA, DNA and metabolites. Developing seeds were observed and photographed on a three-dimensional microscopy system with super depth of field (VHX-6000, Japan). The yellow-seeded degree, oil and protein contents of mature seeds of four varieties with extremely yellow and black seeds were measured by a NIRS DS2500 analyzer (Foss Analytical A/S) as previously described methods (Fu et al. 2007). The thousand-seed weight and hull content were measured using mature seeds with at least five replicates.

RNA extraction, library construction and sequencing

Total RNA was extracted using the EASYspin RNA Rapid Plant Kit (Bio-med, Beijing China) with two replicates per sample. High-quality RNA samples were used for library construction with the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, USA) and sequenced on an Illumina HiSeq 2000 platform as 150 bp paired-end reads. Clean reads were obtained by removing adapter sequences, reads with N stretches and low-quality reads from the raw sequencing data. Q20 (Base error rate < 0.01), Q30 (Base error rate < 0.001) and GC content of clean reads were calculated as diagnostics (Supplementary Table S2). Clean reads were mapped to the *B. rapa* reference genome (Chiifu-401–42) (http://39.100.233.196:82/download_genome/Brassica_Genome_data/Brapa30/) using Hisat2 (Kim et al. 2019). Novel transcripts were predicted by StringTie (Pertea et al. 2015), and transcript levels were calculated as fragments per kilobase of transcript sequence per millions base pairs (FPKM) with the featureCounts tool in Subread (Liao et al. 2013). Pearson correlation analysis between pairs of samples and principal component analysis (PCA) on all samples were performed using FPKM values. Differential expression analysis between yellow-and black-seeded varieties was conducted with DESeq2 ($|\log_2(\text{fold-change})| > 0$ and $p\text{-adj} < 0.05$) (Love et al. 2014). Gene Ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analyses of DEGs (differentially expressed genes) were implemented with the *clusterProfiler* package in R ($p\text{-adj} < 0.05$) (Yu et al. 2012).

qRT-PCR analysis

Total RNA extracted above was used for first-strand cDNA synthesis with the PrimeScript RT reagent kit with gDNA Eraser (Takara, Japan). The final volume of each qRT-PCR amplification was 20 µL, consisting of 10 µL of 2 × T5 Fast qPCR Mix (Tsingke, China), 1 µL of 20 µM forward primer, 1 µL of 20 µM reverse primer, 100 ng cDNA template. The reaction was performed on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad, USA), following the PCR procedures: 94 °C, 2 min; 94 °C, 3 s, 60 °C, 30 s, 40 cycles; melting curve from 60 to 95 °C. The $2^{-\Delta\Delta Ct}$ method was used to calculate relative transcript levels, with *ACTIN7* as the reference gene (Livak and Schmittgen 2001).

Broad spectrum metabolome detection

The extraction of metabolites from developing seeds was performed as previously described, with slight modifications (Qu et al. 2020). Frozen seeds were ground into a fine powder in liquid nitrogen with a mortar and pestle. Approximately, 0.2 g of powder was accurately and rapidly weighed with a precision balance and placed in pre-chilled 2-mL centrifuge tubes (Axygen, USA), and then 800 µL of 80% (v/v) methanol was added and mixed well with a shaker. The homogenates were extracted using an ultrasound water bath at 4 ℃. The volume of the homogenates was adjusted to 1 mL with 80% methanol before centrifugation at 4 ℃ and 10,000 *g* for 10 min. Two hundred microliters of the supernatant were filtered through a 0.22-µm-filter and used for liquid chromatography on a Dionex Ultimate 3000 HPLC system (Thermo Fisher Scientific, USA) with an Acclaim 120 C18 chromatographic column (5 µm, 2.1 mm × 150 mm, Waters, USA) and mass spectrometric detection with the Thermo Scientific Q-Exactive system (Thermo Fisher Scientific, USA). Mobile phase A consisted of 0.1% (w/v) formic acid (Sigma-Aldrich, USA). Mobile phase B was 0.1% (v/v) formic acid-acetonitrile (Sigma-Aldrich, USA). The elution program was 5% phase A, 2 min; 5% A ~ 95% B, 11 min; and 95% B, 2 min. The detection was carried out in anion mode, and the scanning range of mass spectrometry was 100 ~ 1500 m/z. The Thermo Xcalibur 3.0.63 (Thermo Fisher Scientific, USA) was used for data collection and analysis.

Liquid chromatography–mass spectrometry (LC/MS) data processing

The chromatographic peaks and secondary mass spectra were extracted by Thermo Xcalibur 3.0.63 (USA) Qual Browser, and the peak areas were calculated by Quan Browser. The PubChem (https://pubchem.ncbi.nlm.nih.gov/) and Massbank (https://massbank.eu/MassBank/) databases were used to identify metabolites. Quantitative data were processed with MetaboAnalyst 5.0 (https://www.metaboanalyst.ca/) for statistical analysis and metabolic pathway enrichment analysis (Pang et al. 2021). Fold change (FC) analysis, T-tests, PLSDA (Partial Least Squares–Discriminant Analysis) and OrthoPLSDA (Orthogonal Partial Least Squares–Discriminant Analysis) were used to screen differential metabolites between yellow-and black-seeded groups.

Weighted gene co-expression network analysis (WGCNA)

The WGCNA R package (version 1.67, https://cran.r-project.org/web/packages/WGCNA/) (Langfelder and Horvath 2008) was used to generate co-expression networks and detect co-expression modules from the 49,734 genes detected by RNA-seq. The contents for 295 metabolites were defined as trait data to identify modules whose eigengene correlated with various metabolites. The gene co-expression networks were exported to Cytoscape v 3.6.1 for visualization (Shannon et al. 2003).

DNA extraction, restriction site-associated DNA sequencing (RAD-Seq) and metabolite genome-wide association study (mGWAS)

Genomic DNA for 159 *B. rapa* varieties were extracted with the Plant Genomic DNA Extraction kit (Tiangen, Beijing) and used for RAD-seq analysis. A set of 57, 589 single nucleotide polymorphism (SNP) markers, evenly distributed over the chromosomes, were identified and used for genome-wide association mapping. The Q matrix of population structure was estimated with STRUCTURE v2.3.4 (Pritchard et al. 2000) (Supplementary Fig. S1). The peak areas of 42 metabolites from the flavonoid biosynthesis pathway in developing seeds (35 DAP) from all 159 *B. rapa* varieties were used for metabolite GWAS (mGWAS), which were grown in 2018, 2019 and 2020 seasons, respectively. To detect significant quantitative trait nucleotides (QTNs), multi-locus GWAS were implemented in mrMLM v3.1 (https://cran.r-project.org/web/packages/mrMLM/index.html) with mrMLM, FASTmrEMMA, ISIS EM-BLASSO and FASTmrMLM algorithms (Zhang et al. 2020). Genes located 100 kb upstream and downstream of significant QTNs were selected as putative candidate genes for flavonoid metabolism, which were eventually confirmed by the RNA-seq data to identify genes affecting seed coat color. The genomic position of detected QTNs and genes were visualized with Circos plots produced by the TBtools software (Chen et al. 2020) and on maps drawn with the MapChart software (Voorrips 2002).

Protein–protein interaction analysis and chromosome collinearity

Protein interaction partners were predicted for all proteins encoded by candidate genes in the STRING database (https://string-db.org/; Szklarczyk et al. 2011). Reference genomes for several *B. rapa* varieties were obtained from the National Center for Biotechnological Information (NCBI) database, including ZYCX (PRJNA576336), CT001 (PRJNA546028), Z1 (PRJEB26620) and R-o-18 (PRJNA649364). The sequence of chromosome A09 was aligned across different *B. rapa* varieties with the Mauve software (Darling et al. 2004).

Transient expression of *BrMYB111* in *Nicotiana benthamiana* leaves

Based on the above-mentioned multi-omics analysis, the gene *BrMYB111* was considered an important transcription factor regulating the early stage of the flavonoid biosynthesis pathway in *B. rapa*. Subsequently, the coding sequences of two *BrMYB111s*, *BraA09g004490.3C* and *BraA06g034790.3C*, were cloned from BrB1 (Luhua) and were assembled into pEarleyGate101 vector under the control of the CaMV35S promoter. The recombinant vectors were transferred into *Agrobacterium tumefaciens* strain GV3101. The cells containing recombinant vectors were collected and resuspended in buffer solution (10 mM $MgCl_2$, 10 mM MES monohydrate, 100 µM Acetosyringone, pH = 5.6, OD_{600} = 0.5). Fully expanded leaves of 5-week-old *N. benthamiana* plants were pressure-infiltrated by suspended *Agrobacterium* and water using a needleless syringe. After dark (24 h) and light (48 h) treatments, the infiltrated and untreated leaves were sampled and quickly frozen in liquid nitrogen. The methods for extraction and detection of metabolites were the same as those for the developing seeds mentioned above.

Results

Quality traits of extreme yellow and black seeds

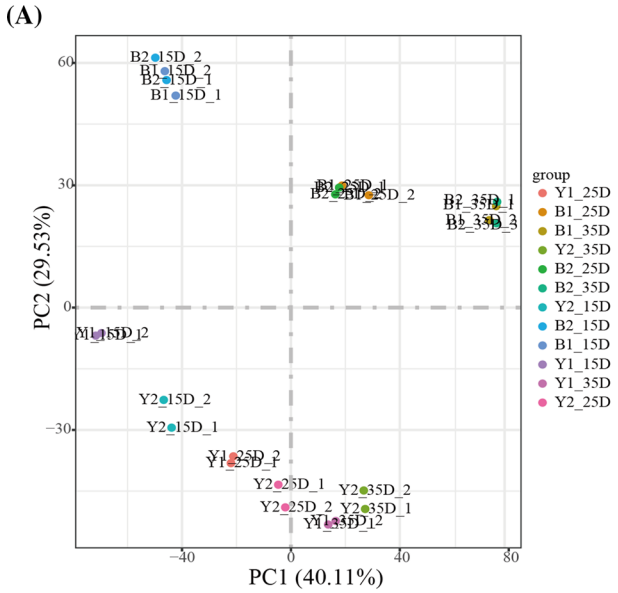
In this study, analyses of seed coat color were performed in four extreme *B. rapa* varieties, starting from 15 DAP to seed maturity. We found that the seed coat color started to change in black-seeded varieties around 35 DAP, and large-area coloring occurred around 45 DAP, whereas yellow-seeded varieties remained green (Fig. 1A). At maturity, the yellow-seeded degree was much higher in yellow seeds relative to varieties with black seeds (Fig. 1B). The oil content (Fig. 1C), protein content (Fig. 1D) and hull content (Fig. 1E), as well as the thousand-seed weight (Fig. 1F) did not clearly correlate with the seed coat color trait.

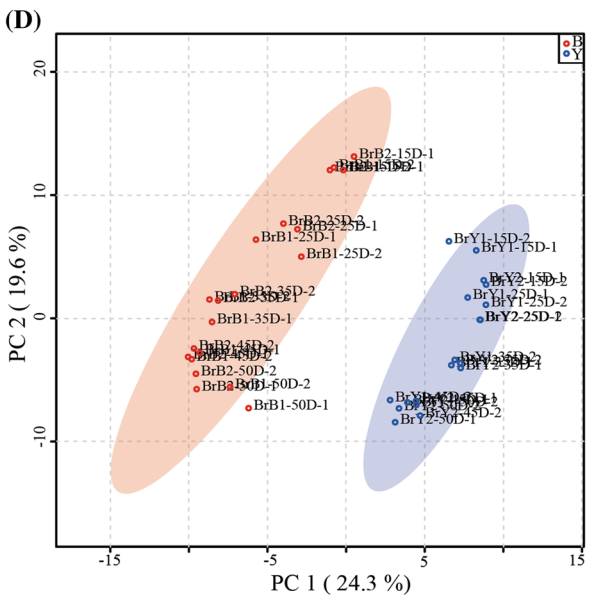
Transcriptome deep sequencing and identification of DEGs

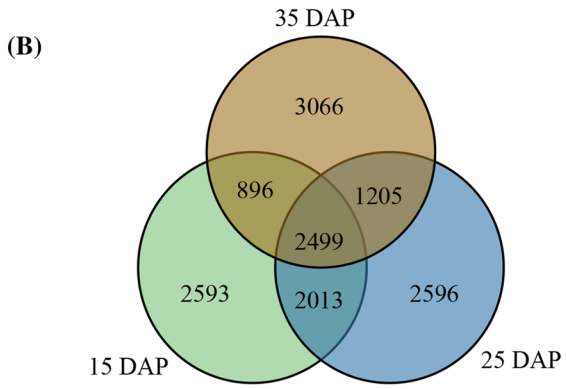
We collected seeds for each of the four extreme varieties at 15, 25 and 35 DAP for transcriptome deep sequencing (RNA-seq). 1,182,593,810 clean reads were obtained across all samples, which we then mapped to the *B. rapa* reference genome. Mean error rate was less than 3%, while mean Q20 and Q30 values were 97.2% and 92.2%, respectively, and the GC content was between 46.4% and 49.1% (Supplementary Table S2). We performed a principal component analysis (PCA) and correlation analysis with expression estimates (Fig. 2A), revealing good reproducibility between the two replicates and clear separation of the samples according to seed color. In addition, black-seeded varieties showed strong clustering as a function of seed developmental stage, with both varieties behaving identically, while yellow-seeded varieties tended to inhabit a broader continuum between early and later seed developmental stages (Fig. 2A), indicating that they showed the apparently differences among these varieties. A comparison of transcript levels between yellow- and black-seeded varieties identified 14,868 DEGs across all varieties, including 8,001 DEGs at 15 DAP, 8,313 DEGs at 25 DAP, and 7,666 DEGs at 35 DAP. Of all DEGs, 2,499 were consistently different between yellow-and black-seeded varieties for all three developmental stages, with 1,231 genes always expressed at lower levels in yellow-seeded varieties, 1,160 genes always expressed at higher levels in yellow-seeded varieties, and 108 genes with variable expression in 15 to 35 DAP seeds (Fig. 2B). We then used the 14,868 DEGs above for clustering analysis, of which 10,717 DEGs showed consistent differences between yellow-and black-seeded varieties (Supplementary Fig. S2). Meanwhile, linear regression analysis indicated that the fold changes for gene transcriptional levels investigated by qRT-PCR and RNA-seq data were significantly positively correlated (R^2 = 0.8596, Supplementary Fig. S3), indicating that these results were reliable.

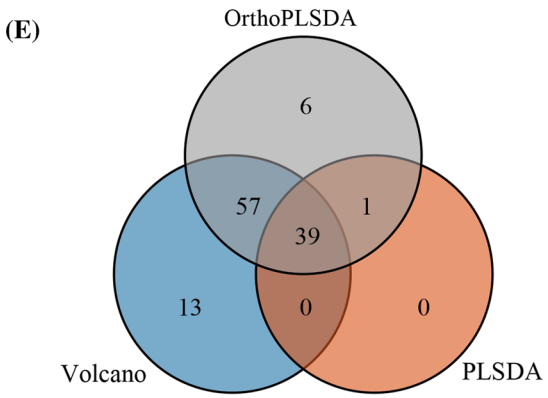
GO and KEGG enrichment analysis on DEGs

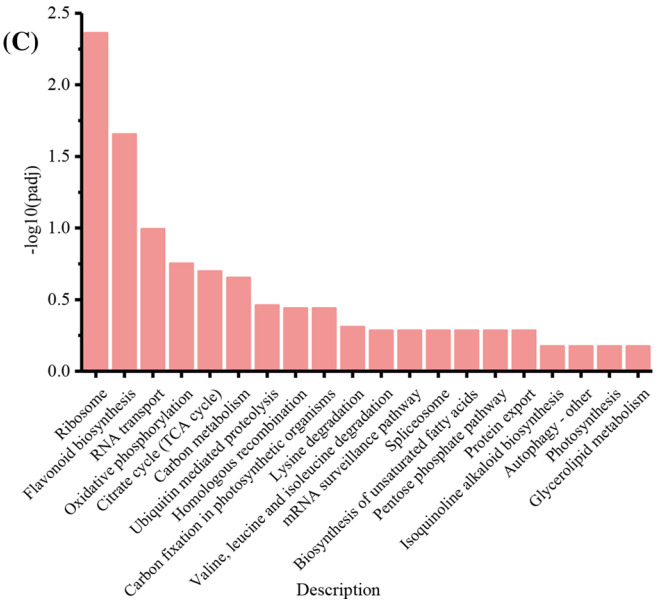
To explore the function of identified DEGs, we performed a GO and KEGG enrichment analysis on the core set of 2,499 DEGs between yellow-and black-seeded varieties. Based on KEGG enrichment analysis, flavonoid biosynthesis (brp00941) was the most significantly enriched and down-regulated pathway in yellow-seeded varieties (Fig. 2C). Among them, these down-regulated genes mainly encoded enzymes along the flavonoid biosynthesis pathway,

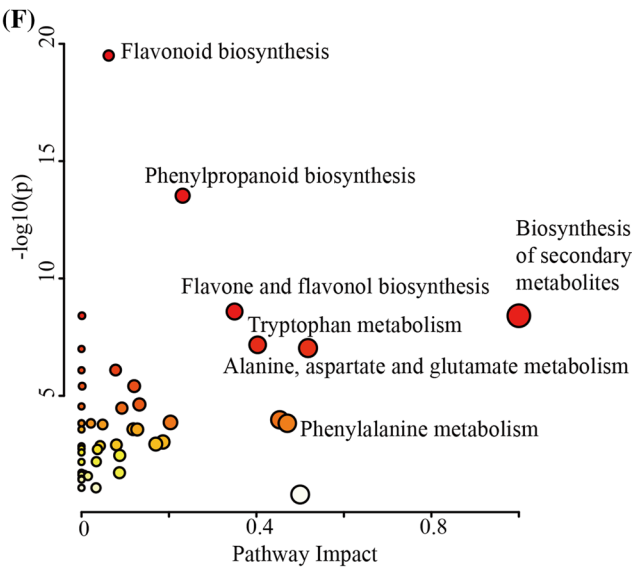

(A)
PC2 (29.53%)
PC1 (40.11%)
group
Y1_25D
B1_25D
B1_35D
Y2_35D
B2_25D
B2_35D
Y2_15D
B2_15D
B1_15D
Y1_15D
Y1_35D
Y2_25D


(D)
PC 2 (19.6 %)
PC 1 (24.3 %)


(B)
35 DAP
3066
896
1205
2499
2593
2013
2596
15 DAP
25 DAP


(E)
OrthoPLSDA
6
57
1
39
13
0
0
Volcano
PLSDA


(C)
-log10(padj)
Ribosome
Flavonoid biosynthesis
RNA transport
Oxidative phosphorylation
Citrate cycle (TCA cycle)
Carbon metabolism
Ubiquitin mediated proteolysis
Homologous recombination
Carbon fixation in photosynthetic organisms
Lysine degradation
Valine, leucine and isoleucine degradation
mRNA surveillance pathway
Biosynthesis of unsaturated fatty acids
Spliceosome
Pentose phosphate pathway
Protein export
Isoquinoline alkaloid biosynthesis
Autophagy - other
Photosynthesis
Glycerolipid metabolism
Description


(F)
Flavonoid biosynthesis
Phenylpropanoid biosynthesis
Biosynthesis of secondary metabolites
Flavone and flavonol biosynthesis
Tryptophan metabolism
Alanine, aspartate and glutamate metabolism
Phenylalanine metabolism
-log10(p)
Pathway Impact

◂**Fig. 2** Combined transcriptome and metabolome analysis of developing seeds in *B. rapa*. **A** Principle component analysis (PCA) of gene expression between the yellow-and black-seeded *B. rapa*; **B** Differentially expressed gene (DEG) analysis in different developmental seeds of the yellow-and black-seeded *B. rapa*; **C** KEGG enrichment analysis of the DEGs between the yellow-and black-seeded *B. rapa*; **D** Principle component analysis (PCA) of metabolite content between the yellow-and black-seeded *B. rapa*; **E** Screening of differentially accumulated metabolites between yellow-and black-seeded *B. rapa*; **F** KEGG enrichment analysis of the differential metabolites between the yellow-and black-seeded *B. rapa*

including *BraA02g005190.3C* and *BraA03g005990.3C* (*TT4*), *BraA10g028200.3C* (*TT5*), *BraA09g019440.3C* (*TT3*), *BraA01g013470.3C* and *BraA03g050560.3C* (*TT18*), *BraA03g064730.3C* (*BAN*), and so on. Up-regulated KEGG pathways among yellow-seeded varieties included the ribosome (brp03008), photosynthesis (brp00195) and oxidative phosphorylation (brp00190). GO enrichment analysis showed that protein acetylation (GO:0,006,473), internal protein amino acid acetylation (GO:0,006,475) and histone acetylation (GO:0,016,573) are down-regulated in yellow seeds, while up-regulated GO terms were largely congruent with the KEGG pathway enrichment analysis, with over-representation of photosynthesis (GO:0,015,979) and the ribosome (GO:0,005,840) among genes (Supplementary Table S3). These results provide the crucial clues for elucidating the mechanisms of seed coat color formation in *B. rapa*.

Identification of metabolites in *B. rapa* seeds

In parallel to the transcriptome, the dynamic accumulation of metabolites were also analyzed between yellow and black seeds, resulting in 295 metabolites involved in different pathways, such as secondary metabolism, amino acid metabolism and fatty acid metabolism (Supplementary Fig. S4). The original peak intensities data were log-transformed and normalized by median for statistical analysis (one factor). Furthermore, the PCA on metabolites sorted black-seeded and yellow-seeded varieties in separate groups (Fig. 2D), indicating that the biological replicates of the same sample types had a good stability and high reliability. Correlation analysis across all 295 metabolites identified 5 clusters, 18 metabolites from cluster I mainly explained the differences in profiles between black-and yellow-seeded varieties, which included 11 flavonoids (Supplementary Fig. S5). The contents of 116 metabolites were significantly different between black-and yellow-seeded samples (fold-change > 1.5, T-tests q-value < 0.05, variable importance in projection (VIP) > 1), 39 and 57 differential metabolites were detected by four and three methods (Fig. 2E). A pathway enrichment analysis using 116 differentially accumulated metabolites between yellow and black seeds showed that flavonoid biosynthesis, phenylpropanoid biosynthesis and flavone and flavonol biosynthesis pathways were significantly enriched (Fig. 2F), which was consistent with the RNA-seq data. Furthermore, 42 flavonoids were detected and divided into three groups according to their structures and the order of flavonoids biosynthesis pathway (Supplementary Table S4). Of which, the flavonoids from the early biosynthesis stage included flavanones (naringenin, luteolin) and dihydroflavonols (taxifolin); the middle biosynthesis stage mainly contained three different flavonols and their derivatives (kaempferol, quercetin and isorhamnetin), and the late biosynthesis stage mainly includes the colored compositions, such as anthocyanins (cyanidin, delphinidin, petunidin, malvidin) and proanthocyanidins (epicatechin, procyanidin B1, procyanidin C1, [DP3], [DP4]). The contents of 20 flavonoids were significantly different between yellow and black seeds, including 4, 8 and 8 flavonoids from the early, middle and late stage, respectively (Table 1). Eleven of them were with extremely higher content in black seeds, i.e., C145 (Procyanidin C1), C146 (Procyanidin C1), C140 (3′ 4′ 5 7-tetrahydroxyflavanone), C137 ((–)-Catechin), C122 (Procyanidin B1), C138 (Epicatechin), C165 (Isorhamnetin-3-O-sinapoyldiglucoside-7-O-glucoside), C174 ([DP3]-2), C150 ([DP4] Procyanidin D), C185 (Taxifolin) and C213 (Quercetin) (Table 1). The 11 flavonoids were clustered in the cluster I by correlation analysis (Supplementary Fig. S5), and they were also the most related variables with seed coat color trait by PLSDA and OrthoPLSDA (Supplementary Fig. S6), we believed that they played a key role during the seed coat color formation.

Weighted gene co-expression network analysis (WGCNA)

To facilitate our understanding of the regulatory network of the *B. rapa* metabolites and contributing to seed coat color formation, the peak area of 285 metabolites, the yellow-seeded degree and the expression of 49,733 genes were subjected to WGCNA using the block-wise network construction method. A total of 56 modules were detected (soft threshold power = 12, mergeCutHeight = 0.2). We then calculated the correlation between traits and modules, followed by their hierarchical clustering. As a result, these traits roughly fell into four clusters (Supplementary Table S5). We noticed that seed coat color (yellow-seeded degree, referred to as SC hereafter) clustered into cluster 1, while 10 of the top 11 differential flavonoids grouped closely into cluster 4 (including catechin and epicatechins, 3′ 4′ 5 7-tetrahydroxyflavanone, isorhamnetin-3-O-sinapoyldiglucoside-7-O-glucoside and taxifolin) (Supplementary Table S5). The most highly positively correlated module with SC was the blue module, which was most strongly negatively correlated with epicatechin levels, as higher epicatechin content is associated with

Table 1 Significantly different flavonoids between yellow and black seeds in *B. rapa*

Code	Metabolites	RT (min)	m/z ([M-H]-)	Cluster	FC	*q* value	VIP1	VIP2	Stage
C122	Procyanidin B1	6.21	577.14	I	9110.00	3.19E–15	4.90	2.31	L
C137	(-)-Catechin	6.93	335.08	I	10,426.00	3.91E–25	5.39	2.43	L
C138	Epicatechin	6.97	289.07	I	3190.10	3.44E–19	4.86	2.37	L
C140	3′ 4′ 5 7-tetrahydroxyflavanone	6.98	287.06	I	13,261.00	3.38E–25	5.16	2.43	E
C145	Procyanidin C1	7.43	865.20	I	635.80	3.31E–19	3.37	2.37	L
C146	Procyanidin C1	7.47	865.20	I	635.76	3.31E–19	3.37	2.37	L
C150	[DP4] Procyanidin D	7.82	1153.26	I	33.79	2.72E–07	1.37	1.91	L
C165	Isorhamnetin-3-O-sinapoyldiglucoside-7-O-glucoside	8.73	1007.27	I	1006.80	1.36E–07	3.12	1.94	M
C174	[DP3]-2	9.41	865.20	I	628.32	3.74E–15	3.34	2.30	L
C185	Taxifolin	10.00	303.05	I	4.15	9.73E–05	0.64	1.56	E
C213	Quercetin	13.12	301.04	I	43.02	2.38E–06	1.66	1.78	M
C116	Quercetin-3,4′-o-di-beta-glucopyranoside	5.69	625.14	III	0.45	1.09E–03	0.59	1.42	M
C118	Quercetin-3,4′-O-di-beta-glucopyranoside	5.70	625.14	III	0.45	9.74E–04	0.58	1.43	M
C126	Kaempferol-3-O-sinapoylsophorotrioside-7-O-glucoside	6.33	1139.31	IV	0.02	2.97E–07	2.35	1.90	M
C153	Isorhamnetin-3-O-diglucoside-sulfate	7.94	719.12	IV	0.04	3.66E–04	1.89	1.53	M
C178	Taxifolin	9.64	303.05	IV	0.60	1.56E–04	0.39	1.57	E
C194	Isorhamnetin-3-Glucoside-4′-Glucoside	10.85	639.16	IV	0.34	5.12E–04	0.78	1.48	M
C195	Isorhamnetin-3-o-glucoside	11.05	477.17	IV	0.60	1.56E–04	0.39	1.57	M
C199	Naringenin-7-O-glucoside	11.37	433.12	III	0.20	2.30E–03	1.26	1.37	E
C205	Malvidin-3-o-beta-d-galactoside	12.70	491.12	IV	0.10	7.17E–03	1.46	1.24	L

RT, the retention time of metabolites; m/z, mass-to-charge ratio; Cluster, the detected metabolites could be divided into 5 clusters based on their contents in different samples, details were presented in Supplementary figure S5; FC, the fold change of matabolites' content in black seeds compared with that in yellow seeds; *q* value, adjusted *p* value of *t*-tests; VIP1 and VIP2, the VIP value of PLSDA and OrthoPLSDA; Stage, the early (E), middle (M) and late (L) stage of flavonoid biosynthetic pathway

lower yellowness. Based on our module-trait correlation analysis, we extracted 10 modules exhibiting high correlation with SC and the top 10 differential flavonoids, including the blue, cyan, green, lightcyan, salmon, lightgreen, skyblue, maroon, palevioletred3 and turquoise modules, for further analysis (Supplementary Table S5). KEGG and GO enrichment analysis were conducted on genes from 10 selected modules. The down-regulated pathway flavonoid biosynthesis (brp00941) in yellow seeds was enriched in green module, the up-regulated pathways ribosome (brp03010) and oxidative phosphorylation (brp00190), and the down-regulated pathways spliceosome (brp03040) and RNA transport (brp03013) were enriched in blue module, which indicated that these two modules were the most important modules for the seed coat color formation. Besides, the down-regulated pathways spliceosome (brp03040) and RNA transport (brp03013) were also enriched in lightcyan module. Two fatty acid metabolic pathways, fatty acid degradation (brp00071) and alpha-linolenic acid metabolism (brp00592), were enriched in turquoise module, implying that this module also participated in regulating the fatty acid metabolism (Supplementary Table S6).

Herein, the top 30 hub genes were selected for each of the 10 correlated modules and added them to known flavonoid biosynthetic genes to construct the gene co-expression network (Supplementary Tables S7 and S8). Results showed that genes in the same module tended to cluster together (Fig. 3). The blue and green modules were the most closely related with known flavonoid biosynthetic genes. In addition, five indices (degree, closeness, centroid value, betweenness and bridging centrality) were employed to measure how critical each gene was to a given network (Supplementary Table S8). As we expected, most known flavonoid biosynthetic genes had high degree and closeness values, indicating that they may play central roles in the regulation of gene expression within the network. Several known flavonoid biosynthetic genes were also associated with high centroid value, betweenness and bridging centrality values, such as *PHENYLALANINE AMMONIA-LYASE 1* (*BrPAL1a*), *UDP-GLUCOSYL TRANSFERASE 2* (*BrUGT2a*, *BrUGT2e*), *BrTT10b*, *BrTT18a*, *BrTT8*, *BrTT4b* and *BrTT6c*, suggesting that they may be crucial for maintaining connections across different modules. Meanwhile, several newly discovered genes were also identified from different modules, including *BraA03g034240.3C* (encoding protein of unknown function, blue module), *BraA04g030130.3C* (Carbohydrate-binding X8 domain superfamily protein, cyan module), *BraA06g007200.3C* (Glutathione S-transferase family protein17, cyan module), *BraA06g017750.3C* (Major facilitator super family protein, skyblue module) and

BraA10g006450.3C (MATE efflux family protein, salmon module), which showed high betweenness and bridging centrality values, making them potentially important nodes to connect their module to other modules or known flavonoid biosynthetic genes (Supplementary Table S8, Fig. 3).

Except for the 10 selected modules closely related with SC and the top 10 differential flavonoids, some flavonoids from the different stages of flavonoid biosynthesis pathway correlated with distinctive modules, such as C199 (Naringenin-7-O-glucoside) with black module, C126 (Kaempferol-3-O-sinapoylsophorotrioside-7-O-glucoside) with pink module, C223 (Isorhamnetin) and C196 (Petunidin-3-O-beta-glucopyranoside) with darkgreen module, C178 (Taxifolin) and C195 (Isorhamnetin-3-o-glucoside) with darkolivegreen module, etc. The top hub genes of these modules were also used for further candidate gene screening for different flavonoids metabolism (Supplementary Table S5).

Metabolome-based genome-wide association study for different flavonoids

A collection of 159 natural spring *B. rapa* accessions showing extensive variation in seed color were used to measure their flavonoids in seeds by LC/MS in different environment over three years and two locations, which were then used as traits for metabolite GWAS (mGWAS). The genotype, environment, genotype × environment interaction effect on different flavonoids, and the heritabilities of them were analyzed, the results showed that the environment have a significant influence on most of flavonoids, especially on the flavonoids from the early and middle stage of flavonoid biosynthesis pathway (Supplementary Table S4).

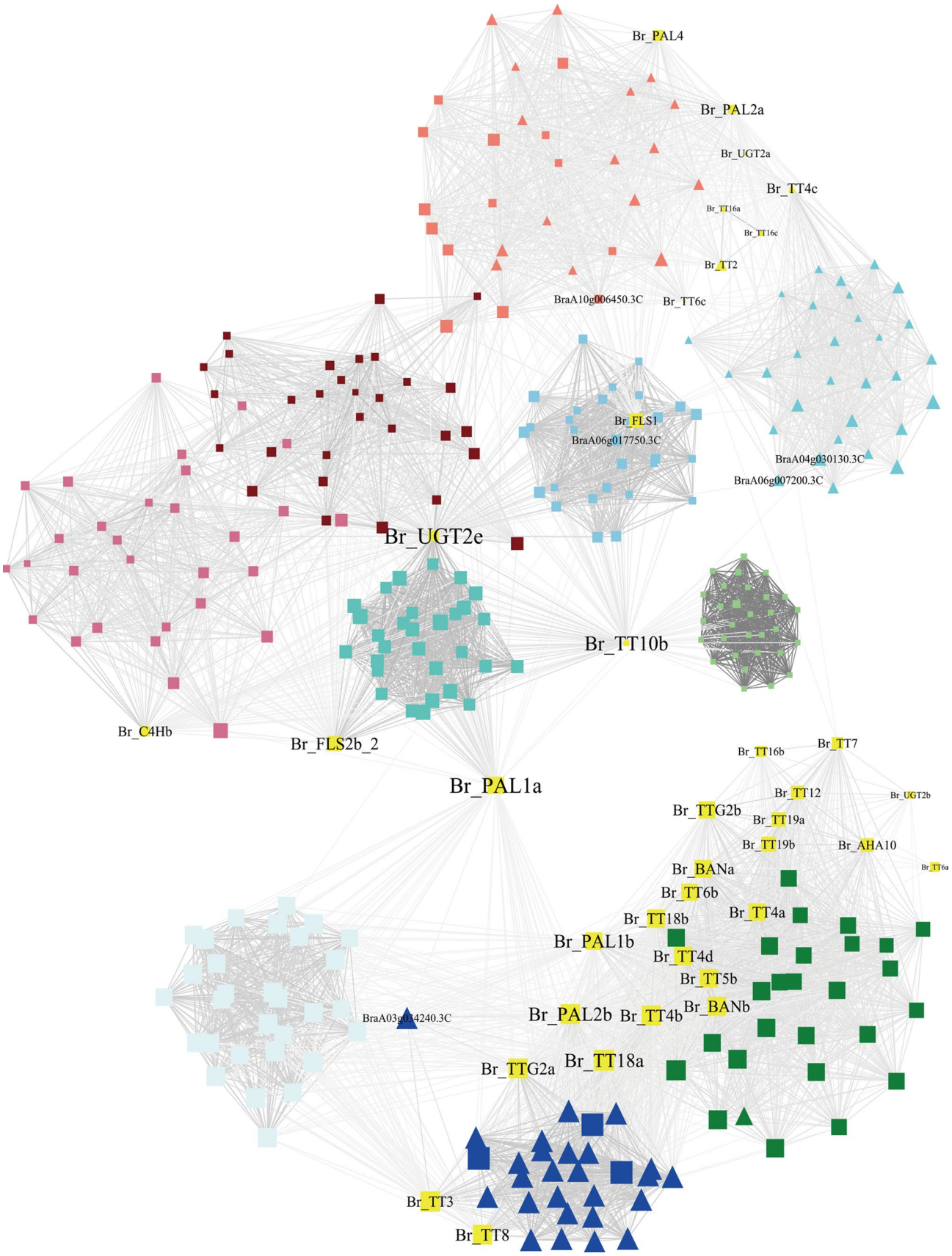
In total, 1,626 quantitative trait nucleotides (QTNs) were detected for 42 flavonoids by four multi-locus GWAS algorithms, and distributed on 10 chromosomes (A01–A10) and 151 scaffolds, explaining 0.17% to 61.66% of the phenotypic variation for different flavonoids, respectively (Supplementary Fig. S7). Of which, 500 QTNs were shared among different metabolites, GWAS algorithms or environments. We detected 306, 719 and 699 QTNs for metabolites from the early, middle and late stage of flavonoid biosynthesis pathway (Fig. 4B). Focusing on the top 10 differential flavonoids mentioned above (results section of WGCNA), we detected 517 closely linked QTNs in three years.

Screening candidate genes for the flavonoid biosynthesis pathway

To obtain high-confidence candidate genes for flavonoid biosynthesis and seed coat color formation, we calculated the QTN density over the entire genome in 500-kb windows and a step of 50 kb, yielding 209 candidate regions with high-density QTNs (QTN number ≥ 11), coalescing into 37 intervals on the chromosomes that contained 6,272 genes (Fig. 4B). In addition, 4,810 candidate genes were detected and located around 236 QTNs with higher contribution rates ($r^2 \geq 15\%$), and 13,160 genes were identified from the 500 sharing QTNs among different stages. Furthermore, we performed an overall analysis of combining mGWAS, RNA-seq results (DEGs and selected modules from WGCNA), and functional annotation of homologous genes to confirm the candidate genes. A total of 241 candidate genes were identified that might be involved in flavonoid biosynthesis and seed coat color formation in *B. rapa*, encoding transcription regulators, transporters, oxidoreductases, glycosyl transferases and many proteins of unknown function, etc. (Supplementary Table S9). Interestingly, the candidate genes were significant enriched in the flavonoid biosynthesis (brp00941), circadian rhythm (brp04712) and glutathione metabolism (brp00480) by KEGG pathway enrichment analysis, and the enriched genes included BraA02g039760.3C (*BrTT4c*), BraA03g005990.3C (*BrTT4a*), BraA02g003870.3C (*BrHY5*), BraA03g016820.3C (glutathione peroxidase), etc. (Supplementary Table S7 and S9). In addition, we found that 80, 134 and 113 candidate genes were predicted responsible for the early, middle and late stages of flavonoid biosynthesis pathway, respectively, such as BraA10g024990.3C (*BrTT4b*) and BraA04g032060.3C (*BrCPC*) for the early stage, BraA10g024430.3C (*BrMYB66*) and BraA02g017040.3C (*BrMYB90*) for the middle stage, BraA07g039090.3C (*BrBBX21*) for the late stage, etc. (Supplementary Table S9). Importantly, 107 candidate genes were highly associated with the top 10 differential flavonoids from the metabolites between the yellow-and black-seeded *B. rapa*, such as BraA08g009180.3C (*BrTT2*), BraA08g029360.3C (*BrAHA10*), BraA09g004490.3C (*BrMYB111*), etc. (Supplementary Table S9). However, whether these candidate genes are directly involved in the regulation of the corresponding metabolites biosynthesis has to be investigated in the future.

Interaction network prediction and functional analysis of candidate genes

The interaction network analysis plays a crucial role in many aspects of biological processes and helps comprehensively understand the metabolic mechanism in plants. Herein, we explored the potential interaction network of 241 candidate genes and known *TT* genes involved in flavonoid biosynthesis using the STRING database (http://string-db.org/). Results showed that the interaction network could be roughly clustered into two clusters by K-means clustering and enriched in the flavonoid biosynthesis pathway (brp00941; Supplementary Fig. S8). Furthermore, we found that 67 of candidate genes were verified in the identified interaction network. For example, the candidate


Br_PAL4
Br_PAL2a
Br_UGT2a
Br_TT4c
Br_TT16a
Br_TT16c
Br_TT2
BraA10g006450.3C
Br_TT6c
Br_FLS1
BraA06g017750.3C
BraA04g030130.3C
BraA06g007200.3C
Br_UGT2e
Br_TT10b
Br_C4Hb
Br_FLS2b_2
Br_PAL1a
Br_TT16b
Br_TT7
Br_TT12
Br_UGT2b
Br_TTG2b
Br_TT19a
Br_TT19b
Br_AHA10
Br_TT6a
Br_BANa
Br_TT6b
Br_TT18b
Br_TT4a
Br_PAL1b
Br_TT4d
Br_TT5b
BraA03g034240.3C
Br_PAL2b
Br_TT4b
Br_BANb
Br_TTG2a
Br_TT18a
Br_TT3
Br_TT8

◂**Fig. 3** Co-expression network analysis of the hub genes and known flavonoid biosynthetic genes. Known flavonoid biosynthetic genes are shown in yellow; other color points represent the candidate genes belonging to the 10 modules closely related to seed coat color and the top differential metabolites. Squares and triangles indicate that the expression of a gene is positively or negatively correlated with seed coat color, respectively; the size of squares and triangles indicate the gene significance (GS) value between the gene and seed coat color; the size of the genes' name indicates the genes' degree in the network

genes, including BraA03g020660.3C (*SPLAYED*, *SYD*), BraA04g032060.3C (*CAPRICE*, *CPC*), BraA02g003870.3C (*LONG HYPOCOTYL 5*, *HY5*), BraA02g041360.3C (*MYB86*), BraA09g004490.3C and BraA06g034790.3C (*MYB111*), BraA02g017040.3C (*PRODUCTION OF ANTHOCYANIN PIGMENT 1*, *PAP1*), BraA09g000420.3C (*ANTHOCYANINLESS 2*, *ANL2*), BraA08g015650.3C (*LESS ADHESIVE POLLEN 5*, *LAP5*), etc., play a crucial role in flavonoid biosynthesis, which could be interacted with known *TT* regulators (*TT1*, *TT8*, *TT2* and *TT19*). In present study, we noticed that two members of *BrMYB111*, BraA09g004490.3C and BraA06g034790.3C, were located nearby three significant QTNs (18C185_mM_31.81, 19C178_FM_2.11 and 19C178_FM_2.46) for taxifolin (C185, C178; Figs. 4 and 5A, Supplementary Fig. S7 and Table S9), implying that *BrMYB111* may be involved in the metabolism of taxifolin. Hence, functional analysis for these sequences was performed by transient expression in *N. benthamiana*. Meanwhile, 39 flavonoid metabolites were detected between the injected and untreated *N. benthamiana* leaves, and the levels of taxifolin were significantly increased in the leaves of overexpressed *BrMYB111* than that did in leaves treated with water (Fig. 5). Our findings suggest that *BrMYB111* may catalyze the key step of taxifolin biosynthesis and further confirm the validity of mGWAS results.

Discussion

The flavonoid biosynthesis pathway is one of the most important secondary metabolism pathways in plants. It produces chalcones, flavones, flavonols, flavandiols, anthocyanins and proanthocyanidins, which are pigments that give different plant organs their color, such as leaves, flowers and seeds (Koes et al. 2005, 1994; Routaboul et al. 2006; Winkel-Shirley 2001). Recent advances in transcriptome and metabolome profiling have opened the doors to the exploration of the molecular and biochemical basis behind the various colors displayed by plant organs, with the flavonoid biosynthesis pathway being often implicated (Lai et al. 2020; Li et al. 2020; Xue et al. 2020). How the seed coat acquires its color has been investigated in *B. rapa* (Ren et al. 2021), *B. napus* (Qu et al. 2020), *B. juncea* (Liu et al. 2013) and *B. carinata* (Li et al. 2010). In this study, our results support that proanthocyanidins (including epicatechin and its polymers with different degrees of polymerization) were the most significantly differentially accumulating pigments between yellow and black seeds in different *Brassica* species. Previous results showed that anthocyanins were not detected in *B. carinata* with yellow or black seeds, and dihydroquercetin, dihydrokaempferol, quercetin, and kaempferol were only present in *B. carinata* with yellow seeds (Marles et al. 2003). In *B. napus*, the contents of several isorhamnetin (isorhamnetin, isorhamnetin-3-O-glucoside, isorhamnetin-3-O-glucoside-7-O-glucoside) and quercetin (quercetin-3-O-sophoroside) derivatives were consistently different between yellow and black seeds in certain *B. napus* varieties (Lian et al. 2017; Qu et al. 2020; Wang et al. 2018), while kaempferol-3-O-glucoside-7-O-glucoside had different accumulation patterns between yellow and black seeds (Lian et al. 2017; Wang et al. 2018). In *B. rapa*, proanthocyanidins and several quercetin derivatives (quercetin deoxyhex, quercetin glucoside rhamnoside, quercetin glucuronside) were reported to differentially accumulate between *B. rapa* with yellow and brown seeds (Ren et al. 2021). Other derivatives of isorhamnetin, kaempferol and quercetin derivatives, including isorhamnetin dihexoside, isorhamnetin-pentoside, kaempferol-hexoside-glucoside, quercetin-hexoside and quercetin glucoside rhamnoside, were detected in both yellow and brown seeds, although at lower levels in yellow seeds (Ren et al. 2021). Herein, we found that proanthocyanidins was the most important differentially accumulating metabolites between *B. rapa* yellow and black seeds (Table 1). Other metabolites such as 3′ 4′ 5 7-tetrahydroxyflavanone, taxifolin, quercetin and some isorhamnetin derivatives were also significantly different between yellow and black seeds (Table 1). Aside from proanthocyanidins, intermediate products of the flavonoid biosynthesis pathway (flavonols, flavandiols and anthocyanins) accumulated to variable levels among the different varieties of the same Brassica species or among different *Brassica* species. The characterization of flavonoids in wild-type Arabidopsis seeds and various transparent testa mutants revealed the inherent plasticity of the flavonoid biosynthesis pathway (Routaboul et al. 2006), which might also apply to *Brassica* crops. Due to limits of detection, some flavonoids were likely missed, but they might be unlikely to substantially contribute to seed coat coloration. Rather, proanthocyanidins are the root cause of the yellow and black (or brown) seed coat color, while the differences of other flavonoids might lead to more abundant colors in rapeseed, implying the different regulation mechanism among different species.

Genes encoding enzymes involved in the flavonoid biosynthesis pathway were broadly identified from DEGs between yellow and black seeds and from pathway enrichment analyses in different Brassica species. The expression

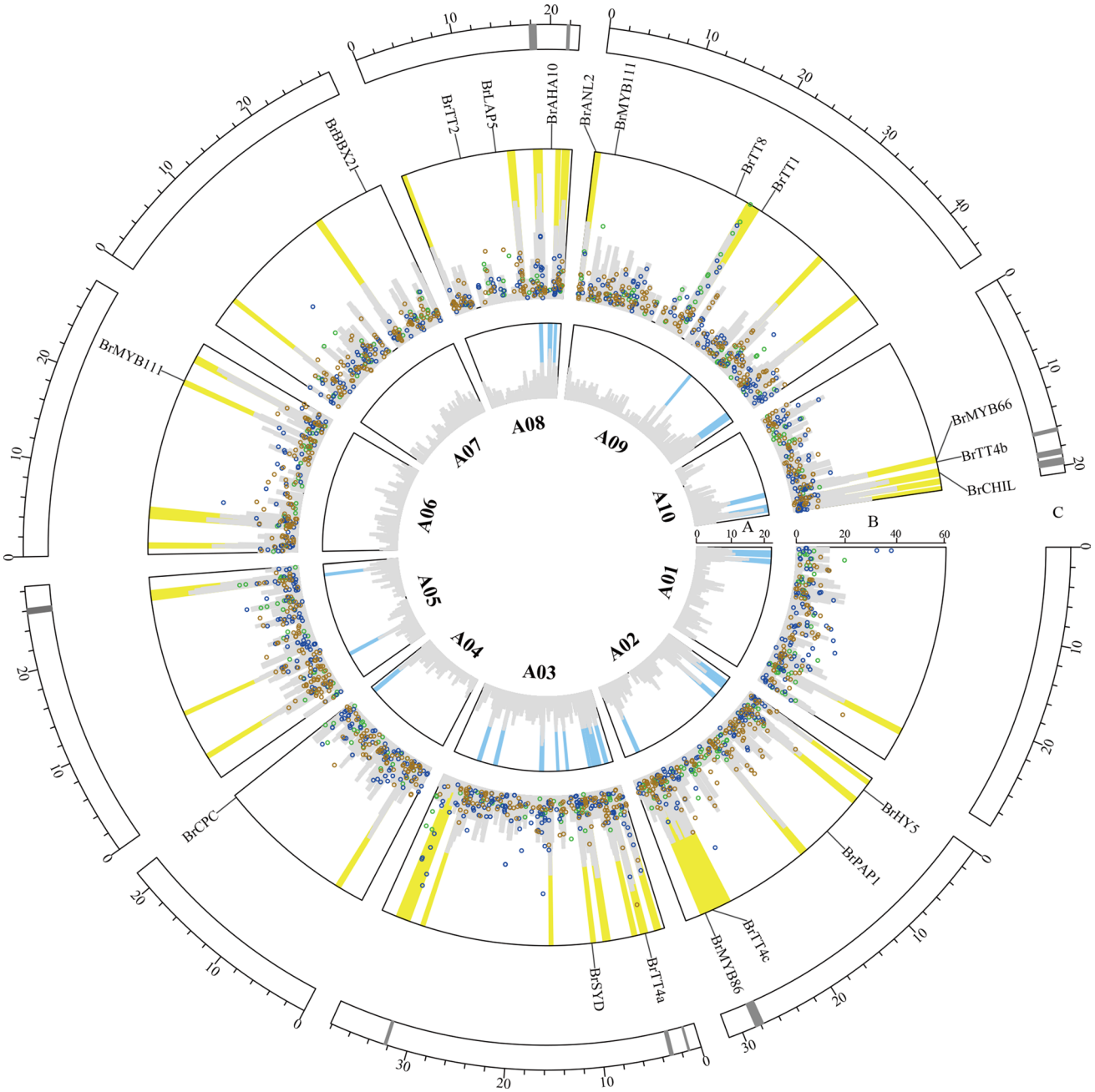


Fig. 4 Comparison of the mGWAS and DEGs in *B. rapa*. **A** Distribution of 2,499 DEGs (Fig. 2) between yellow and black seeds during all stages of seed development on 10 chromosomes, the lightgray bars and scale indicated the number of DEGs in 500 Kb intervals on the 10 chromosomes. 25 intervals containing high-density DEGs were marked as lightblue. **B** Distribution of the detected quantitative trait nucleotides (QTNs) and *TT* or candidate genes to different flavonoids on 10 chromosomes. The green, blue and brown dots indicated detected QTNs for flavonoids from the early, middle and late stage of flavonoid biosynthesis pathway. The scale indicated the contribution rate (%) of QTNs, and the the lightgray bars indicated the number of QTNs in 500 Kb intervals on the 10 chromosomes. 37 intervals containing high-density QTNs were marked as lightyellow; **C** Distribution of 10 overlapped intervals between 25 high-density intervals for DEGs and 37 high-density intervals for QTNs on the 10 chromosomes, which were marked as dark gray

of most *TT* genes that involved in the flavonoid biosynthesis pathway is significantly repressed in yellow seeds in different *Brassica* species, especially genes encoding enzymes from the late stage of the pathway, such as *TT3* (encoding DFR, dihydroflavonol reductase), *TT18* (ANS, anthocyanidin synthase), *BAN* (ANR, anthocyanidin reductase), *TT12* (proton antiporter) and *AHA10* (H^+-ATPase) (Hong et al. 2017; Jiang et al. 2013; Li et al. 2012; Padmaja et al. 2014;

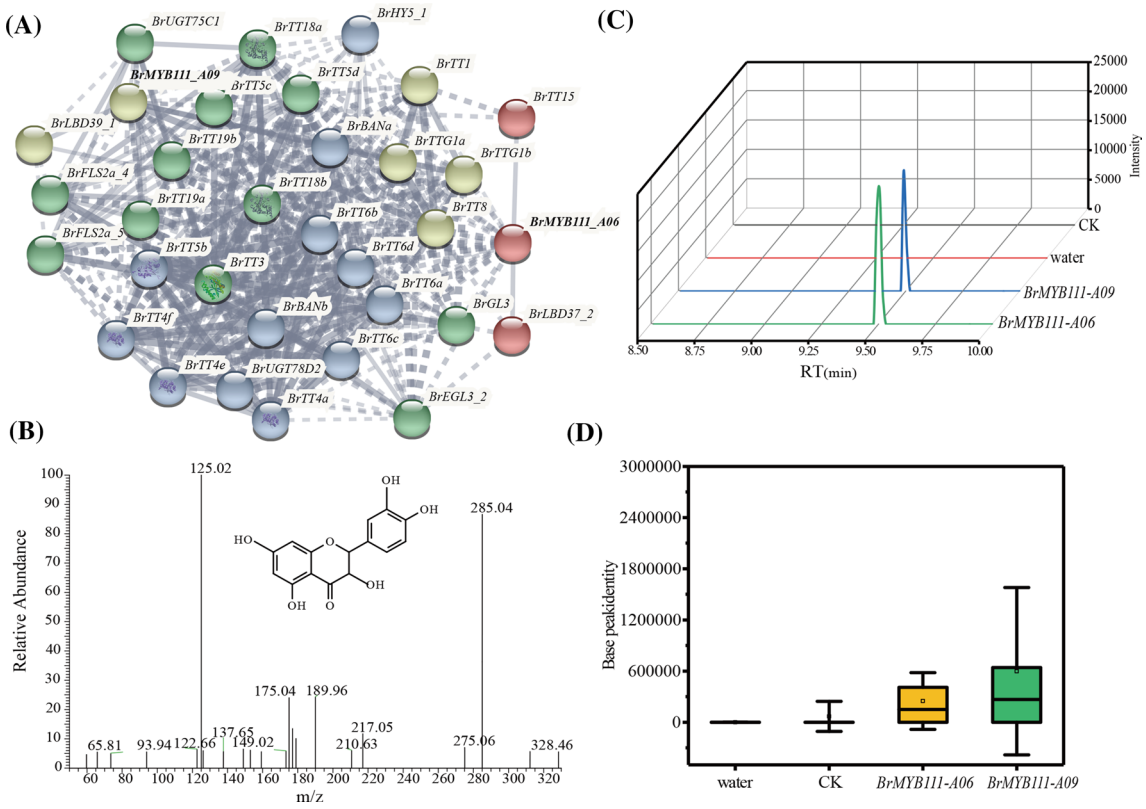


Fig. 5 Identification of the causal gene *BrMYB111* related to the biosynthesis of taxifolin. **A** The predictive sub-network of *BrMYB111*; *BrMYB111* and the potential interacting genes can be divided into four clusters (nodes with four different colors) by K-means clustering in STRING database. The dotted lines indicate interaction between clusters, and the solid lines indicate interaction of genes in one cluster. The darker the line is, the more evidence for the interaction between the two linked genes. Detailed information on protein–protein interaction can be found in Supplementary Table S10. The fragmentation pattern (MS/MS) (**B**) and total ion chromatogram (**C**) of taxifolin obtained from *N. benthamiana* leaves by in vivo injection test. **D** The relative contents of detected taxifolin in overexpression of *BrMYB111* in *N. benthamiana* leaves. BrMYB111-A09, BrMYB111-A06 and water indicated *N. benthamiana* leaves were pressure-infiltrated by suspended *Agrobacterium* with two members of *BrMYB111* and water. CK represents untreated *N. benthamiana* leaves. M111_A6 and M111_A9 represent the *N. benthamiana* leaves infiltrated by BrMYB111-A06 and BrMYB111-A09

Qu et al. 2013; Ren et al. 2021; Wang et al. 2017; Zhao et al. 2019). Meanwhile, some other *TT* genes from the early and middle stages of the flavonoid biosynthesis pathway consisted of multigene families, such as *TT4* (CHS, chalcone synthase), *TT5* (CHI, chalcone isomerase), *TT6* (F3H, flavanone 3-hydroxylase) and *FLS* (DFR, flavonol synthase), with each member exhibiting a distinct expression profile. However, the expression patterns of individual genes are similar in yellow and black seeds of *B. rapa* and *B. napus*, even for homologs from the A subgenome (for example, the *TT4a* gene, which is *BraA03g005990.3C* in *B. rapa* and *BnaA03g04590D* in *B. napus*), indicating that the function of important structural *TT* genes is conserved among *Brassica* species (Hong et al. 2017; Ren et al. 2021; Wang et al. 2017; Zhao et al. 2019). In addition, regulatory transcription factors, such as *TT1* (zinc finger protein), *TT8* (*bHLH42*), *TT2* (*MYB123*) and *TTG1* (WD40 domain containing protein), have been reported to play major roles in seed coat color determination in *B. rapa*, *B. napus* and *B. juncea* (Li et al. 2012; Padmaja et al. 2014; Wang et al. 2017; Xie et al. 2020; Zhai et al. 2020; Zhang et al. 2009). However, only *BrTT8* was significantly down-regulated in yellow seeds compared to black seeds, which was consistent with other studies of the seed coat color trait in *B. rapa* (Niu et al. 2020; Ren et al. 2021; Wang et al. 2017; Zhao et al. 2019). Nevertheless, the *B. napus* homologs of *TT1*, *TT8*, *TT2* and *TTG1* were down-regulated in yellow seeds in different studies (Hong et al. 2017; Qu et al. 2013, 2020), hinting at the existence of different regulatory mechanisms for seed coat color trait between *B. rapa* and *B. napus*.

Previous results showed that the down-regulated flavonoid biosynthetic genes were associated with the yellow seeds (Niu et al. 2020; Hong et al. 2017; Xie et al. 2020; Zhai et al. 2020; Liu et al. 2013), while the up-regulated DEGs were rarely reported in Brassica species. In this study, DEGs enriched in the ribosome (brp03010) pathway were significantly up-regulated and mostly detected by both transcriptome and WGCNA analysis (Supplementary Table S3

and S6), in accordance with the previous results (Niu et al. 2020; Hong et al. 2017; Liu et al. 2013). Subsequently, we further found that these candidate genes were mainly co-expressed with *TT* genes (Supplementary Fig. S9) in *B. rapa*. Our findings will provide the new insight into understanding the mechanisms of seed color in *Brassica* species.

In rapeseed, the yellow seed trait is always linked with better quality of seeds, such as higher oil and protein content, lower hull content, etc. (Jiang et al. 2015; Jönsson 1977). However, in this study, the advantages of yellow seed disappeared when compared with black seeds (Fig. 1). We attributed it to the diverse genetic backgrounds of the plant materials we used here. Under the same genetic background, the seeds of yellow-seeded lines of *Brassica* crops had the lower fiber content and higher protein and oil content than that in black-seeded lines (Zhao et al. 2019; Hong et al. 2017). As complex quantitative traits, the quality traits of rapeseed, such as oil and protein content, are not only determined by the seed coat color trait, which are also influenced by different metabolic processes other than flavonoid biosynthesis pathway in *B. rapa*.

Based on metabolome and transcriptome analysis, we constructed a presumptive flavonoid biosynthesis pathway in *B. rapa* (Fig. 6). In this pathway, chalcones, flavones and flavandiols are produced in the early stage of flavonoid biosynthesis; three flavonols (kaempferol, quercetin and isorhamnetin) are produced in the middle stage; and anthocyanins and proanthocyanidins are produced in the late stage. Except for the known *TT* genes, new transcription factors and structural genes were predicted for flavonoid biosynthesis. Forward genetic analysis for seed coat color in some *B. rapa* lines suggested that two to seven genes might be responsible for this trait (Rahman 2001; Schwetka 1982; Stringam 1980). One major locus for seed coat color was located on chromosome A09 in different studies, as well as several minor loci on chromosomes A03, A05, A06 and A08 by quantitative trait locus (QTL) mapping (Bagheri et al. 2013; Kebede et al. 2012; Lou et al. 2007; Rahman et al. 2014; Zhang et al. 2019). In this study, many QTNs with a high contribution rate also mapped to chromosome A09 between 21.4 to 23.2 Mb, with high-confidence candidate genes *BrTT1* (23.6 Mb) and *BrTT8* (19.59 Mb) nearby. Therefore, we compared the sequence and gene arrangement of A09 chromosomes from different *B. rapa* accessions for collinearity analysis, revealing the major locus located in the centromere region with large structural variation and sequence differences (Supplementary Fig. S10 and Table S11) (Zhang et al. 2018), suggesting that new candidate genes contributing to seed coat color and flavonoid metabolism might reside there. In agreement with previous reports, we also noticed a high density of important QTNs on chromosomes A03, A05, A06 and A08, as well as additional intervals on other chromosomes (Fig. 4B). The distribution of 2,499 DEGs on the chromosomes was also uneven. We determined the density of

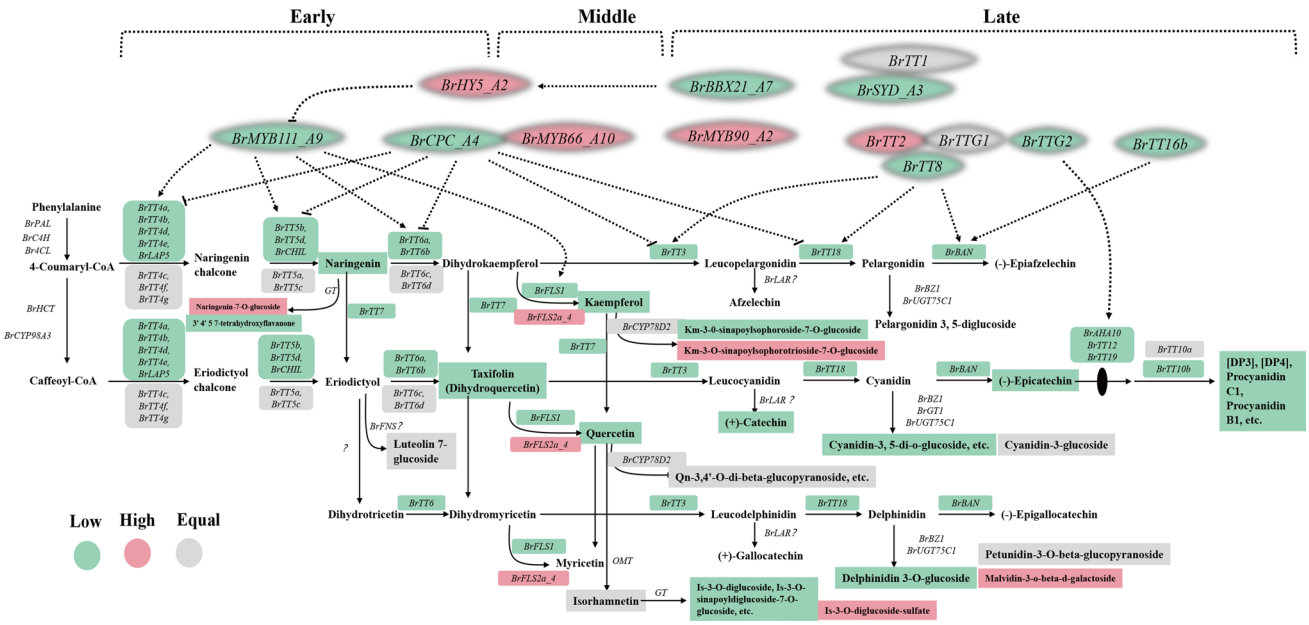


Fig. 6 Comparison of transcriptional levels of the flavonoid biosynthesis pathway in *B. rapa*. Red indicates up-regulated genes, and green indicates down-regulated genes or flavonoids with lower levels in yellow seeds. Gray indicates no change in expression of metabolite levels between yellow and black seeds. Solid lines with arrows indicate metabolite biosynthesis by the listed enzyme. Dotted lines with arrows or short horizontal lines indicate transcriptional activation or repression of the indicated gene. Early, middle and late indicate the three general stages of the flavonoid biosynthesis pathway with approximate boundaries

2,499 DEGs along the chromosomes (500-kb windows and a step of 50 kb, DEGs number ≥ 10), yielding 25 intervals on 10 chromosomes (Fig. 4A). Comparing with the 37 intervals on chromosomes where QTNs gathered, 10 intervals overlapped on A02 (27.7–28.55 Mb), A03 (1.2–1.35 Mb, 2.9–3.25 Mb, 32.15–32.4 Mb), A05 (25.7–26.25 Mb), A08 (17.85–18.55 Mb, 21.7–21.95 Mb) and A10 (16.2–16.4 Mb, 18.3–18.8 Mb, 19.2–19.85 Mb) (Fig. 4C). In plants, the genes participating in the same metabolic pathway might form a gene cluster due to their physical proximity (Liu et al. 2020; Qi et al. 2004). The high concentration of DEGs in the overlapped intervals on the chromosomes in this study might, therefore, represent important metabolic gene clusters for flavonoid biosynthesis. We selected several new candidate genes that might play roles in seed coat color determination in *B. rapa* by combining metabolome and transcriptome data and include homologs to some previously reported regulators of flavonoid biosynthesis in other plants, such as *AtMYB111*, *AtHY5*, *AtBrBBX21* and *AtCPC* (Stracke et al. 2007; Wang et al. 2016; Xu et al. 2016; Zhu et al. 2009). The candidate genes' expression values were extracted from two extremely yellow-seeded and two extremely black-seeded *B. rapa* varieties, which should be verified in more varieties with different kinds of seed coat colors in *B. rapa* to narrow down the candidates. In addition, the candidate intervals can be refined by combining the mGWAS results with other seed coat color mapping results in different segregation populations. The candidate genes that are repeatedly mapped or show constant differential expression between yellow and black seeds deserve further detailed function analysis. In conclusion, this study elucidates the overall molecular mechanisms of flavonoid metabolites and seed coat color in *B. rapa* and provide a theoretical basis for rapid genetic improvements in Brassica species.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-022-04099-8.

Acknowledgements This research was funded by the National Science Foundation of China (31830067, 32072093), Qinghai Natural Science Foundation Program–innovation team (2022-ZJ-902), Innovation and Entrepreneurship Training Program for Undergraduates (202110635058), Natural Science Foundation of Chongqing (cstc2020jcyj-bshX0016), Qinghai Innovation Platform Construction Project (2022–ZJ–Y01, 2022–ZJ–Y13), the 111 Project (B12006) and China Agriculture Research System of MOF and MARA. We thank Professor Guoyong Tang and Professor Zhigang Zhao from Spring Rape Research Institute of Qinghai University for providing important plant materials, and Professor Xiao Qiu (University of Saskatchewan), and Plant editors for their grammar editing on this manuscript.

Authors contribution statement JL and CQ designed the experiments; DD and KL gave important suggestions to the research; HZ, GS, NY, SC, SS, HJ, YT, FS, YZ, YN, ZZ, LX performed experiments. HZ and GS analyzed the data and wrote the manuscript.

Funding This study was financially supported by the National Science Foundation of China (31830067, 32072093), Qinghai Natural Science Foundation Program–innovation team (2022-ZJ-902), Innovation and Entrepreneurship Training Program for Undergraduates (202110635058), Natural Science Foundation of Chongqing (cstc2020jcyj-bshX0016), Qinghai Innovation Platform Construction Project (2022–ZJ–Y01, 2022–ZJ–Y13), the 111 Project (B12006) and China Agriculture Research System of MOF and MARA.

Data availability The datasets used in the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

Auger B, Marnet N, Gautier V, Maia-Grondard A, Leprince F, Renard M, Guyot S, Nesi N, Routaboul JM (2010) A detailed survey of seed coat flavonoids in developing seeds of *Brassica napus* L. J Agric Food Chem 58:6246–6256

Bagheri H, Pino-Del-Carpio D, Hahnart C, Bonnema G, Keurentjes J, Aarts MGM (2013) Identification of seed-related QTL in *Brassica rapa*. Span J Agric Res 11:1085–1093

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant 13:1194–1202

Chen W, Gao Y, Xie W, Gong L, Lu K, Wang W, Li Y, Liu X, Zhang H, Dong H, Zhang W, Zhang L, Yu S, Wang G, Lian X, Luo J (2014) Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. Nat Genet 46:714–721

Pang Z, Chong J, Zhou G, Morais D, Chang L, Barrette M, Gauthier C, Jacques PE, Li S, Xia J (2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucl Acids Res 49:W388–W396

Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403

Daun JK, Declercq DR (1988) Quality of yellow and dark seeds in *Brassica campestris* canola varieties Candle and Tobin. J Am Oil Chem Soc 65:122–126

Debeaujon I, Nesi N, Perez P, Devic M, Grandjean O, Caboche M, Lepiniec L (2003) Proanthocyanidin-accumulating cells in Arabidopsis testa: regulation of differentiation and role in seed development. Plant Cell 15:2514–2531

Fu FY, Liu LZ, Chai YR, Chen L, Yang T, Jin MY, Ma AF, Yan XY, Zhang ZS, Li JN (2007) Localization of QTLs for seed color using recombinant inbred lines of *Brassica napus* in different environments. Genome 50:840–854

Hong M, Hu K, Tian T, Li X, Chen L, Zhang Y, Yi B, Wen J, Ma C, Shen J, Fu T, Tu J (2017) Transcriptomic analysis of seed coats in yellow-seeded *Brassica napus* reveals novel genes that influence proanthocyanidin biosynthesis. Front Plant Sci 8:1674

Jiang J, Shao Y, Li A, Lu C, Zhang Y, Wang Y (2013) Phenolic composition analysis and gene expression in developing seeds of yellow- and black-seeded *Brassica napus*. J Integr Plant Biol 55:537–551

Jiang J, Wang Y, Xie T, Rong H, Li A, Fang Y, Wang Y (2015) Metabolic characteristics in meal of black rapeseed and yellow-seeded

progeny of *Brassica napus-Sinapis alba* hybrids. Molecules 20:21204–21213

Jönsson R (1977) Breeding for improved oil and meal quality in rape (*Brassica napus* L.) and turnip rape (*Brassica campestris* L.). Hereditas 87:205–218

Kebede B, Cheema K, Greenshields DL, Li C, Selvaraj G, Rahman H (2012) Construction of genetic linkage map and mapping of QTL for seed color in *Brassica rapa*. Genome 55:813–823

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37:907–915

Koes R, Verweij W, Quattrocchio F (2005) Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci 10:236–242

Koes RE, Quattrocchio F, Mol J (1994) The flavonoid biosynthesis pathway in plants: Function and evolution. BioEssays 16:123–132

Lai JL, Liu ZW, Luo XG (2020) A metabolomic, transcriptomic profiling, and mineral nutrient metabolism study of the phytotoxicity mechanism of uranium. J Hazard Mater 386:121437

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinfo 9:559

Li BJ, Zheng BQ, Wang JY, Tsai WC, Lu HC, Zou LH, Wan X, Zhang DY, Qiao HJ, Liu ZJ, Wang Y (2020) New insight into the molecular mechanism of colour differentiation among floral segments in orchids. Commun Biol 3:89

Li X, Chen L, Hong M, Zhang Y, Zu F, Wen J, Yi B, Ma C, Shen J, Tu J, Fu T (2012) A large insertion in bHLH transcription factor *BrTT8* resulting in yellow seed coat in *Brassica rapa*. PLoS One 7:e44145

Li X, Westcott N, Links M, Gruber MY (2010) Seed coat phenolics and the developing silique transcriptome of *Brassica carinata*. J Agric Food Chem 58:10918–10928

Lian J, Lu X, Yin N, Ma L, Lu J, Liu X, Li J, Lu J, Lei B, Wang R, Chai Y (2017) Silencing of *BnTT1* family genes affects seed flavonoid biosynthesis and alters seed fatty acid composition in *Brassica napus*. Plant Sci 254:32–47

Liao Y, Smyth GK, Shi W (2013) The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucl Acids Res 41:e108

Liu X, Lu Y, Yuan Y, Liu S, Guan C, Chen S, Liu Z (2013) De novo transcriptome of *Brassica juncea* seed coat and identification of genes for the biosynthesis of flavonoids. PLoS One 8:e71110

Liu Z, Cheema J, Vigouroux M, Hill L, Reed J, Paajanen P, Yant L, Osbourn A (2020) Formation and diversification of a paradigm biosynthetic gene cluster in plants. Nat Commun 11:5354

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}{}_{T}$ Method. Methods 25:402–408

Lou P, Zhao J, Kim JS, Shen S, Del Carpio DP, Song X, Jin M, Vreugdenhil D, Wang X, Koornneef M, Bonnema G (2007) Quantitative trait loci for flowering time and morphological traits in multiple populations of *Brassica rapa*. J Exp Bot 58:4005–4016

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550

Marles MS, Gruber MY, Scoles GJ, Muir AD (2003) Pigmentation in the developing seed coat and seedling leaves of *Brassica carinata* is controlled at the dihydroflavonol reductase locus. Phytochemistry 62(5):663–672

Niu Y, Wu L, Li Y, Huang H, Qian M, Sun W, Zhu H, Xu Y, Fan Y, Mahmood U, Xu B, Zhang K, Qu C, Li J, Lu K (2020) Deciphering the transcriptional regulatory networks that control size, color, and oil content in *Brassica rapa* seeds. Biotechnol Biofuels 13:90

Padmaja LK, Agarwal P, Gupta V, Mukhopadhyay A, Sodhi YS, Pental D, Pradhan AK (2014) Natural mutations in two homoeologous *TT8* genes control yellow seed coat trait in allotetraploid *Brassica juncea* (AABB). Theor Appl Genet 127:339–347

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33:290–295

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Qi X, Bakht S, Leggett M, Maxwell C, Melton R, Osbourn A (2004) A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. P Natl Acad Sci USA 101:8233–8238

Qu C, Fu F, Lu K, Zhang K, Wang R, Xu X, Wang M, Lu J, Wan H, Tang Z, Li J (2013) Differential accumulation of phenolic compounds and expression of related genes in black- and yellow-seeded *Brassica napus*. J Exp Bot 64:2885–2898

Qu C, Yin N, Chen S, Wang S, Chen X, Zhao H, Shen S, Fu F, Zhou B, Xu X, Liu L, Lu K, Li J (2020) Comparative analysis of the metabolic profiles of yellow- versus black-seeded rapeseed using UPLC-HESI-MS/MS and transcriptome analysis. J Agric Food Chem 68:3033–3049

Rahman M, Li G, Schroeder D, McVetty PBE (2010) Inheritance of seed coat color genes in *Brassica napus* (L.) and tagging the genes using SRAP, SCAR and SNP molecular markers. Mol Breeding 26:439–453

Rahman M, Mamidi S, McClean P, Chevre AM (2014) Quantitative trait loci mapping of seed colour, hairy leaf, seedling anthocyanin, leaf chlorosis and days to flowering in F_2 population of *Brassica rapa* L. Plant Breed 133:381–389

Rahman MH (2001) Inheritance of petal colour and its independent segregation from seed colour in *Brassica rapa*. Plant Breed 120:197–200

Ren Y, Zhang N, Li R, Ma X, Zhang L (2021) Comparative transcriptome and flavonoids components analysis reveal the structural genes responsible for the yellow seed coat color of *Brassica rapa* L. PeerJ 9:e10770

Routaboul JM, Kerhoas L, Debeaujon I, Pourcel L, Caboche M, Einhorn J, Lepiniec L (2006) Flavonoid diversity and biosynthesis in seed of *Arabidopsis thaliana*. Planta 224:96–107

Schwetka A (1982) Inheritance of seed colour in Turnip Rape (*Brassica campestris* L.). Theor Appl Genet 62:161–169

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504

Shao Y, Jiang J, Ran L, Lu C, Wei C, Wang Y (2014) Analysis of flavonoids and hydroxycinnamic acid derivatives in rapeseeds (*Brassica napus* L. var. napus) by HPLC-PDA-ESI(–)-MS(n)/HRMS. J Agric Food Chem 62:2935–2945

Shirai K, Matsuda F, Nakabayashi R, Okamoto M, Tanaka M, Fujimoto A, Shimizu M, Shinozaki K, Seki M, Saito K, Hanada K (2017) A highly specific genome-wide association study integrated with transcriptome data reveals the contribution of copy number variations to specialized metabolites in *Arabidopsis thaliana* accessions. Mol Biol Evol 34:3111–3122

Somers DJ, Rakow G, Prabhu VK, Friesen KRD (2001) Identification of a major gene and RAPD markers for yellow seed coat colour in *Brassica napus*. Genome 44:1077–1082

Stein A, Wittkop B, Liu L, Obermeier C, Friedt W, Snowdon RJ, Chevre AM (2013) Dissection of a major QTL for seed colour and fibre content in *Brassica napus* reveals colocalization with candidate genes for phenylpropanoid biosynthesis and flavonoid deposition. Plant Breed 132:382–389

Stracke R, Ishihara H, Huep G, Barsch A, Mehrtens F, Niehaus K, Weisshaar B (2007) Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation

in different parts of the *Arabidopsis thaliana* seedling. Plant J 50:660–677

Stringam GR (1980) Inheritance of seed color in turnip rape. Cann J Plant Sci 60:331–335

Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucl Acids Res 39:D561–D568

Teutonico RA, Osborn TC (1994) Mapping of RFLP and qualitative trait loci in *Brassica rapa* and comparison of the linkage maps of *B. napus*, *B. oleracea*, and *Arabidopsis thaliana*. Theor Appl Genet 89:885–894

Voorrips RE (2002) MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. J Hered 1:77–78

Wang Y, Meng G, Chen S, Chen Y, Jiang J, Wang YP (2018) Correlation of phenolic contents and antioxidation in yellow- and black-seeded *Brassica napus*. Molecules 23:1815

Wang Y, Wang Y, Song Z, Zhang H (2016) Repression of MYBL2 by both microRNA858a and HY5 leads to the activation of Anthocyanin biosynthetic pathway in Arabidopsis. Mol Plant 9:1395–1405

Wang Y, Xiao L, Dun X, Liu K, Du D (2017) Characterization of the BrTT1 gene responsible for seed coat color formation in Dahuang (Brassica rapa L. landrace). Mol Breeding 37:137

Winkel-Shirley B (2001) Flavonoid biosynthesis, a colorful model for genetics, biochemistry, cell biology and biotechnology. Plant Physiol 126:485–493

Xiao L, Zhao Z, Du D, Yao Y, Xu L, Tang G (2012) Genetic characterization and fine mapping of a yellow-seeded gene in Dahuang (a *Brassica rapa* landrace). Theor Appl Genet 124:903–909

Xie T, Chen X, Guo T, Rong H, Chen Z, Sun Q, Batley J, Jiang J, Wang Y (2020) Targeted knockout of *BnTT2* homologues for yellow-seeded *Brassica napus* with reduced flavonoids and improved fatty acid composition. J Agric Food Chem 68:5676–5690

Xu D, Jiang Y, Li J, Lin F, Holm M, Deng XW (2016) BBX21, an *Arabidopsis* B-box protein, directly activates *HY5* and is targeted by COP1 for 26S proteasome-mediated degradation. P Natl Acad Sci USA 113:7655–7660

Xue Q, Fan H, Yao F, Cao XX, Liu MM, Sun J, Liu YJ (2020) Transcriptomics and targeted metabolomics profilings for elucidation of pigmentation in *Lonicera japonica* flowers at different developmental stages. Ind Crops Prod 145:111981

Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16:284–287

Zeng X, Yuan H, Dong X, Peng M, Jing X, Xu Q, Tang T, Wang Y, Zha S, Gao M, Li C, Shu C, Wei Z, Qimei W, Basang Y, Dunzhu J, Li Z, Bai L, Shi J, Zheng Z, Yu S, Fernie AR, Luo J, Nyima T (2020) Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B adaptation of Qingke. Mol Plant 13:112–127

Zhai Y, Yu K, Cai S, Hu L, Amoo O, Xu L, Yang Y, Ma B, Jiao Y, Zhang C, Khan MHU, Khan SU, Fan C, Zhou Y (2020) Targeted mutagenesis of *BnTT8* homologs controls yellow seed coat development for effective oil production in *Brassica napus* L. Plant Biotechnol J 18:1153–1168

Zhang J, Lu Y, Yuan Y, Zhang X, Geng J, Chen Y, Cloutier S, McVetty PB, Li G (2009) Map-based cloning and characterization of a gene controlling hairiness and seed coat color traits in *Brassica rapa*. Plant Mol Biol 69:553–563

Zhang L, Cai X, Wu J, Liu M, Grob S, Cheng F, Liang J, Cai C, Liu Z, Liu B, Wang F, Li S, Liu F, Li X, Cheng L, Yang W, Li MH, Grossniklaus U, Zheng H, Wang X (2018) Improved *Brassica rapa* reference genome by single-molecule sequencing and chromosome conformation capture technologies. Hortic Res 5:50

Zhang Y, Sun Y, Sun J, Feng H, Wang Y (2019) Identification and validation of major and minor QTLs controlling seed coat color in *Brassica rapa* L. Breed Sci 69:47–54

Zhang Y, Tamba C, Wen Y, Li P, Ren W, Ni Y, Gao J, Zhang Y (2020) mrMLM v4.0.2: An R Platform for Multi-locus Genome-wide Association Studies. Genomics Proteomics Bioinform 18(4):481–487

Zhao H, Basu U, Kebede B, Qu C, Li J, Rahman H (2019) Fine mapping of the major QTL for seed coat color in Brassica rapa var. Yellow Sarson by use of NIL populations and transcriptome sequencing for identification of the candidate genes. PLoS One 14:e0209982

Zhu HF, Fitzsimmons K, Khandelwal A, Kranz RG (2009) *CPC*, a single-repeat R3 MYB, is a negative regulator of anthocyanin biosynthesis in Arabidopsis. Mol Plant 2:790–802