Genome-Wide Survey of Flavonoid Biosynthesis Genes and Gene Expression Analysis between Black- and Yellow-Seeded *Brassica napus*

Cunmin Qu¹,²,³†, Huiyan Zhao¹,²†, Fuyou Fu⁴†, Zhen Wang¹,², Kai Zhang¹,², Yan Zhou¹,², Xin Wang¹, Rui Wang¹,², Xinfu Xu¹,², Zhanglin Tang¹,², Kun Lu¹,²* and Jia-Na Li¹,²*

¹ Chongqing Engineering Research Center for Rapeseed, College of Agronomy and Biotechnology, Southwest University, Chongqing, China, ² Engineering Research Center of South Upland Agriculture of Ministry of Education, Southwest University, Chongqing, China, ³ Food and Bioproduct Science, University of Saskatchewan, Saskatoon, SK, Canada, ⁴ Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN, USA

Flavonoids, the compounds that impart color to fruits, flowers, and seeds, are the most widespread secondary metabolites in plants. However, a systematic analysis of these loci has not been performed in Brassicaceae. In this study, we isolated 649 nucleotide sequences related to flavonoid biosynthesis, i.e., the *Transparent Testa* (*TT*) genes, and their associated amino acid sequences in 17 Brassicaceae species, grouped into *Arabidopsis* or Brassicaceae subgroups. Moreover, 36 copies of 21 genes of the flavonoid biosynthesis pathway were identified in *Arabidopsis thaliana*, 53 were identified in *Brassica rapa*, 50 in *Brassica oleracea*, and 95 in *B. napus*, followed the genomic distribution, collinearity analysis and genes triplication of them among Brassicaceae species. The results showed that the extensive gene loss, whole genome triplication, and diploidization that occurred after divergence from the common ancestor. Using qRT-PCR methods, we analyzed the expression of 18 flavonoid biosynthesis genes in 6 yellow- and black-seeded *B. napus* inbred lines with different genetic background, found that 12 of which were preferentially expressed during seed development, whereas the remaining genes were expressed in all *B. napus* tissues examined. Moreover, 14 of these genes showed significant differences in expression level during seed development, and all but four of these (i.e., *BnTT5, BnTT7, BnTT10*, and *BnTTG1*) had similar expression patterns among the yellow- and black-seeded *B. napus*. Results showed that the structural genes (*BnTT3, BnTT18*, and *BnBAN*), regulatory genes (*BnTTG2 and BnTT16*) and three encoding transfer proteins (*BnTT12, BnTT19*, and *BnAHA10*) might play an crucial roles in the formation of different seed coat colors in *B. napus*. These data will be helpful for illustrating the molecular mechanisms of flavonoid biosynthesis in Brassicaceae species.

**Keywords:** *Brassica napus* L., Brassicaceae species, expression profile, flavonoid biosynthesis pathway, phylogenetic analysis, seed coat color

**Abbreviations:** TT: Transparent Testa, Bn: Brassica napus L., TTG: TRANSPARENT TESTA GLABRA, PAL: Phenylalanine ammonia-lyase, C4H: cinnamate 4-hydroxylase, FLS: flavonol synthase, LDOX: leucoanthocyanidin dioxygenase, BAN: BANYULS, AHA10: H⁺-ATPase isof orm 10, PAs: proanthocyanidins, Bj: Brassica juncea, BLAST: basic local alignment search tool, BRAD: Brassica Database, PGDD: PLANT GENOME DUPLICATION DATABASE, DAP: days after pollination.
INTRODUCTION

Rapeseed (Brassica napus) is the most important source of edible vegetable oil and protein-rich meal in the world diet, and the seeds, which store protein and oil, are the major harvested plant organ (Nesi et al., 2008). However, the quality of rapeseed oil and meal is greatly affected by the pigments and polyphenols derived from flavonoid biosynthesis that remain after oil extraction. Rapeseed use is limited by the concentration of anti-nutritional factors, including phenolic compounds, lignin, tannins, and proanthocyanidins, it contains. Previous research showed that yellow-seeded B. napus has a thinner seed coat, less pigmentation, and higher protein and oil contents than does black-seeded B. napus in the same background, rendering it a more nutritional feed for livestock (Chen and Henene, 1992; Tang et al., 1997; Meng et al., 1998). Thus, selecting lines with a stable yellow-seed coat trait is one of the most important breeding aims for B. napus.

In plant kingdom, flavonoid biosynthesis pathway play important roles in the coloration of fruits, flowers, and seeds, and numerous evidences had showed that TT-type genes and their homologs are crucial for the accumulation of flavonoids and their derivatives (Nesi et al., 2001; Winkel-Shirley, 2002; Xie et al., 2003; Baudry et al., 2004; Hoffmann et al., 2006; Lepiniec et al., 2006; Kasai et al., 2007). In the model plant Arabidopsis thaliana, the formation of transparent and colorless testa (seed coat) were associated with tt loci that are disrupted the flavonoid synthesis pathway in the loss-of-function mutations (Wan et al., 2002; Winkel-Shirley, 2002; Baudry et al., 2004; Lepiniec et al., 2006). To date, 17 genes involved in this pathway have been cloned and functionally characterized, including eight structural genes (i.e., TT3, TT4, TT5, TT6, TT7, FLS1, LDOX, and BAN; Albert et al., 1997; Devic et al., 1999; Xie et al., 2003; Routaboul et al., 2006; Chiu et al., 2010), six regulatory genes (TT1, TT2, TT8, TTG1, TTG2, and TT16; Nesi et al., 2000, 2001; Baudry et al., 2006; Routaboul et al., 2006), and three encoding transfer proteins (TT12, TT19, and AHA10; Debeaujon et al., 2001; Baxter et al., 2005), which were also classified as Early Biosynthetic Genes (CHS, CHI, and F3H etc.) and Late Biosynthetic Genes (BAN, DFR, and TTG1, etc.) (Nesi et al., 2000, 2001; Winkel-Shirley, 2001; Lepiniec et al., 2006). Homologs of some of these genes, named Tt-type genes, have also been identified and shown to be involved in the flavonoid biosynthetic pathway. These genes are thus candidate genes for the molecular basis of seed color manifestation (Supplementary Table S1). TTG1 in Brassica rapa has the same gene function as its orthologs in A. thaliana, i.e., it influences root hairiness and the color of the seed coat (Zhang et al., 2009). BrTT8 was recently shown to regulate the accumulation of proanthocyanidins (PAs) in the seed coat and to regulate the expression of the late biosynthetic genes (LBGs) of the flavonoid pathway in B. rapa, and an analysis in the “sorbon” line of B. rapa showed that the yellow-seeded trait was caused by loss of BrTT8 function (Li X. et al., 2012). In addition, BjuA.TT8 and BjuB.TT8 co-segregated perfectly with the seed coat color phenotype in allotetraploid Brassica juncea (Padmaja et al., 2014). However, the inheritance of seed coat color is complex in B. napus. In previous studies, a stable major quantitative trait locus (QTL) for seed coat color of B. napus was detected in different generations and environments, and TT10 was considered as a candidate gene involved in seed coat color, based on microsynteny of this QTL with Arabidopsis genome sequences (Fu et al., 2007). The following findings showed that BrTT10 functions in proanthocyanidin polymerization and lignin biosynthesis, as well as seed coat pigmentation in B. napus (Zhang et al., 2013). Additionally, Chai et al. (2009) found that TT12 was also a candidate gene for seed coat color in B. napus. Moreover, several key loci isolated from B. napus by our group, such as F3’h, PAL1, TTG1, and TT2, showed no or limited down-regulation in the yellow-seeded lines (Wei et al., 2007; Xu et al., 2007; Ni et al., 2008; Lu et al., 2009). Based on a marker closely linked with a major QTL for seed fiber and color in B. napus, Stein et al. proposed that the transparent testa gene AHA10 has a strong effect on both seed color and lignin content (Stein et al., 2013). Undoubtedly, the inheritance of seed color in B. napus is also sensitive to environmental influences, such as lighting, temperature, maturity, and harvest time (Chen and Henene, 1992; Deynne et al., 1995). Therefore, the molecular mechanism underlying the yellow seed coat trait is unclear in Brassica species.

The family Brassicaceae is well known for its large variation in chromosome numbers, common occurrence of polyploids and many reports of interspecific gene flow (Marhold and Lihova, 2006). Moreover, Brassicaceae plants arose form a common ancestor, of which B. napus (AACC, genome size ∼489.7 Mb) was allotetraploid species formed ∼7500 years ago by hybridization between B. rapa (AA, genome size ∼312 Mb) and Brassica oleracea (CC, genome size ∼540 Mb), followed by genome duplications and mergers during the evolutionary process (Chalhoub et al., 2014). Therefore, Brassica is an ideal model to increase knowledge of polyploid evolution (Parkin et al., 2005; Albertin et al., 2006), which is usually assumed that the physiology and developmental biology of TT genes in A. thaliana are highly similar to those of other Brassicaceae plants. Although dozens of genes involved in the flavonoid biosynthesis pathway of A. thaliana were identified based on tt mutations (Holton and Cornish, 1995; Devic et al., 1999; Wan et al., 2002; Xie et al., 2003; Baudry et al., 2006; Lepiniec et al., 2006; Routaboul et al., 2006; Saito et al., 2013), only some of these have been characterized in other Brassicaceae plants, and to date no comprehensive study of these genes has been reported. We previously conducted a systematic study of the expression profiles of related genes in B. napus seeds at different stages of development (Qu et al., 2013).

In this study, we identified 21 genes involved in the flavonoid biosynthesis pathway in 17 sequenced Brassicaceae species. We systematically analyzed the phylogenetic relationships and triplication events of these genes among the Brassicaceae plants. Additionally, the inheritance of flavonoid biosynthesis pathway in B. napus is quite complicated, and the regulatory mechanisms underlying the biosynthesis of the relevant genes were not well understood. Hence, using quantitative real-time PCR (qRT-PCR) analysis, we identified significant differences (Student's t-test, P < 0.05 or 0.01) in the expression patterns of 18 genes associated with the flavonoid biosynthesis pathway in the stems, leaves, buds, flowers, siliques, and pericarps, and at five different stages of seed development (10, 20, 30, 40, and 50 DAP) in six inbred rapeseed lines, which were used to represent typical yellow- and black-seeded genotypes of B. napus, have different genetic backgrounds (Figure 1, Table 1). These results provide useful
information for identifying key genes or regulatory nodes that control yellow seed coat formation, and provide insight into the inheritance of qualitative differences between the yellow- and black-seeded *B. napus*.

**MATERIALS AND METHODS**

**Plant Materials**

Plant materials were collected from yellow-seeded and black-seeded *B. napus* in different genetic backgrounds (Table 1, Figure 1). The yellow seed degree are calculated using the published reports (Li et al., 2012). The B1 and B2 were inbred lines ZY821 and ZS11, and Y1 and Y2 were GH06 and 05E258-1, which were used to represent typical black-seeded and yellow-seeded genetic types of *B. napus*, respectively. B3 and Y3 are representative two near isogenic lines of black- and yellow-seeded *B. napus*, selected from the successive backcross of the sixth generation with parent ZS11 and recurrent parent 06E241 (Qu et al., 2015). The plants were grown under normal field conditions at Chongqing Rapeseed Technology Research Center (CRTRC) in 2014. Field management essentially followed normal agronomic procedures. Various tissue organs, including stems (St), leaves (Le), flowers (Fl), buds (Bu), silique pericarps (SP), and seeds harvested at 10, 20, 30, 40, and 50 days after pollination (DAP) were sampled and stored at −80°C until used.

**Identification of Flavonoid Biosynthesis Pathway Gene Members in Brassicaceae Species**

The full genome sequences of Brassicaceae were downloaded from multiple databases, including the BRAD database (http://brassicadb.org/brad/ftpTrans.php; Cheng et al., 2011), the Phytozome database (http://phytozome.jgi.doe.gov/pz/portal.html#isearch; Goodstein et al., 2012), PGDD (PLANT GENOME DUPLICATION DATABASE; http://chibba.agtec.uga.edu/ duplication/index/files; Lee et al., 2013), the *Raphanus sativus* Genome DataBase (http://radish.kazusa.or.jp/; Kitashiba et al., 2014), and the *B. napus* database (http://www.genoscope.cns.fr/fr/brassicanapus/; Chalhoub et al., 2014). From amongst the species with full genome sequences, the following 17 species were selected: *Aethionema arabicum* (Aa), *Arabidopsis halleri* (Ah), *Arabidopsis lyrata* (Al), *A. thaliana* (At), *Brassica napus* L. (Bn), *B. oleracea* (Bo), *B. rapa* (Br), *Boechera stricta* (Bs), *Capsella grandiflora* (Cg), *Capsella rubella* (Cr), *Camelina sativa* (Cs), *Leavenworthia alabamica* (La), *Raphanus sativus* L. (Rs), *Sisymbrium irio* (Si), *Schrenkia parvula* (Sp), *Thellungiella halophila* (Th), and *Thellungiella salsuginea* (Ts). All coding sequences (CDSs) and amino acid sequences were stored in a local Brassicaceae database using Geneious Pro 4.8.5 software (http://www.geneious.com/). To identify the flavonoid biosynthesis pathway genes and their homologous genes, all amino acid sequences of flavonoid biosynthesis pathway genes in *A. thaliana* retrieved from TAIR 10 (http://www.arabidopsis.org/) were used as queries to search against the Brassicaceae protein models with HMMER3 (version 3.1b2 with Pfam HMM library Pfam 28.0; Finn et al., 2011). To identify the flavonoid biosynthesis genes, unique protein sequences of these genes from the *A. thaliana* genome were used as query using BLASTP program (Altschul et al., 1997) in the local Brassicaceae database developed in this research. All taxa were named using two-letter acronyms and gene type was used as the species gene name. Briefly, the first uppercase letter represents the genus, the second the species, and the following the gene name. A number at the end indicates the copy number. For example, *A. thaliana* phenylalanine ammonia lyase 1, which has only one copy in *A. thaliana*, is indicated by AtPAL1.

**Mapping of Flavonoid Biosynthesis Pathway Genes among Brassicaceae Species**

To assign the location of flavonoid biosynthesis pathway genes in the Brassicaceae species genomes, the GFF genome files were downloaded from the aforementioned databases. Then, MapChart 2.0 was used to draw graphic representations of their corresponding physical position on pseudo-molecular chromosomes of *Brassica* crops.

**TABLE 1 | List of B. napus genotypes used in this study.**

| No. | Genetic background | Yellow seed degree | Location |
|-----|--------------------|--------------------|----------|
| B1  | ZY821              | 22.23              | Chongqing, China |
| Y1  | GH06               | 126.43             | Chongqing, China |
| B2  | ZS9                | 63.00              | Chongqing, China |
| Y2  | 05E258-1           | 121.46             | Chongqing, China |
| B3  | ZS9/08E241(BC2F6)  | 54.21              | Chongqing, China |
| Y3  | ZS9/08E241(BC2F6)  | 131.68             | Chongqing, China |
TABLE 2 | Characterization of genes involved in the phenylpropane-flavonoid biosynthesis pathway of Brassicaceae species.

| Gene name | Length (aa) | No. of Brassicaceae Species | A. thaliana | B. rapa | B. oleracea | B. napus | Total No. |
|-----------|-------------|-----------------------------|-------------|--------|-------------|---------|---------|
| PAL1      | 263–728     | 18                          | 1           | 2      | 2           | 4       | 28      |
| PAL2      | 377–725     | 16                          | 1           | 3      | 4           | 5       | 31      |
| PAL3      | 107–698     | 6                           | 1           | 2      | 2           | 2       | 10      |
| PAL4      | 344–709     | 16                          | 1           | 1      | 1           | 2       | 22      |
| C4H       | 468–1197    | 17                          | 1           | 5      | 5           | 10      | 43      |
| TT1       | 287–323     | 17                          | 1           | 1      | 1           | 2       | 20      |
| TT2       | 156–265     | 17                          | 1           | 1      | 1           | 2       | 20      |
| TT3       | 244–387     | 18                          | 1           | 1      | 1           | 3       | 24      |
| TT4       | 393–734     | 15                          | 1           | 5      | 3           | 9       | 39      |
| TT5       | 197–254     | 18                          | 1           | 3      | 3           | 4       | 30      |
| TT6       | 274–548     | 16                          | 1           | 4      | 3           | 6       | 34      |
| TT7       | 384–687     | 17                          | 1           | 1      | 1           | 2       | 21      |
| TT8       | 497–613     | 17                          | 1           | 1      | 1           | 2       | 23      |
| TT10      | 489–1676    | 17                          | 1           | 2      | 1           | 3       | 22      |
| TT12      | 507–560     | 17                          | 1           | 1      | 1           | 2       | 22      |
| TT15      | 473–864     | 17                          | 1           | 1      | 1           | 2       | 24      |
| TT16      | 183–318     | 17                          | 1           | 3      | 3           | 6       | 33      |
| TT18      | 351–416     | 18                          | 1           | 2      | 2           | 4       | 28      |
| TT19      | 213–439     | 17                          | 1           | 2      | 2           | 5       | 27      |
| TTG1      | 120–521     | 17                          | 1           | 2      | 1           | 3       | 24      |
| TTG2      | 114–431     | 17                          | 1           | 2      | 3           | 4       | 31      |
| AHA10     | 918–985     | 16                          | 1           | 1      | 1           | 2       | 20      |
| BAN       | 216–1041    | 17                          | 1           | 2      | 2           | 4       | 27      |
| UGT2a     | 422–566     | 12                          | 1           | 1      | 2           | 2       | 14      |
| UGT2b     | 84–662      | 17                          | 1           | 4      | 3           | 5       | 35      |

Phylogenetic Analysis

Based on previously described methods, all sequence alignments for each flavonoid biosynthesis pathway gene superfamily were performed using ClustalW2 software (Larkin et al., 2007), and phylogenetic analysis was carried out using Molecular Evolutionary Genetics Analysis (MEGA) 6.0 (Tamura et al., 2013) with a maximum likelihood (ML). In the ML method, phylogenetic trees were constructed using the JTT+I+G substitution model in PhyML version 3.0.1. To ensure the accuracy of the phylogenetic tree, each tree was subjected to bootstrap analysis with 1000 replicates (Guindon et al., 2010). Finally, all the phylogenetic trees were visualized using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Expression Pattern Analysis of TT-Type Genes in B. napus

To characterize differences in expression of the 18 genes associated with the flavonoid biosynthesis pathway between the yellow- and black-seeded varieties, we designed the primers in consensus region based on alignment the gene sequence (Table 3). Then the total RNA was extracted from various tissues using the RNAprep Pure Plant Kit (TIANGEN BIOTECH, Beijing, China) according to manufacturer's instructions. Then, 1 µg RNA sample was reverse transcribed with the Oligo dT-Adaptor Primer using the RNA PCR Kit (AMV) Ver. 3.0 (TaKaRa, http://www.takara.com.cn). To monitor sample uniformity of initial RNA input and RT efficiency, Brassica napus 26S rRNA was used as the internal control according to the previously described method (Qu et al., 2013).

Real-time PCR was performed using SYBR® Premix Ex Taq™ II (Perfect Real Time) (TaKaRa, China) in a 20 µl volume that included 10 µl of SYBR® Premix Ex Taq™ II, 2 µl (100 ng) of template cDNA, and 0.4 µM of each PCR primer. All primer sequences used for the qRT-PCR are listed in Table 3, designed according to the methods described in our previous research (Qu et al., 2013). Then the specific primers used in this study and cycling conditions were 95°C for 2 min, followed by 40 cycles at 95°C for 10 s (denaturation) and 60°C for 20 s (annealing and extension). The melting curves of each PCR application were obtained using the following cycling conditions: 95°C for 10 s followed by a constant increase in temperature between 65 and 95°C at an increment of 0.5°C/cycle, and samples were run on the Bio-Rad CFX96 Real Time System (USA). The relative expression of the target genes was analyzed using the 2−ΔΔCt method with BnACTIN7 (EV116054) and BnUBC21 (EV086936) as internal controls (Wu et al., 2010).

Three biological replicates for each sample were used for real-time PCR analysis and three technical replicates were analyzed for each biological replicate. Then the values represent the average ± SD of three biological replicates with three technical
replicates of each tissue and organ. Relative gene expression levels were normalized according to the expression values in black-seeded at 10 DAP.

RESULTS

Characterization of Flavonoid Biosynthesis Pathway Gene Members in Brassicaceae Species

To identify members of the Brassicaceae gene subfamily that are involved in the flavonoid biosynthesis pathway, the whole genome sequences of 17 species were downloaded from multiple public databases (Materials and Methods). In total, 649 nucleotide sequences of 21 genes and their associated amino acid sequences were respectively were identified using the HMM nucleotide sequences of 21 genes and their associated amino acid sequences ranged from 84 aa (Bn_UGT2d) to 1676 aa (Sp_TT10; Supplementary Table S2). In addition, the number of gene families also varied by species; for example, only 10 copies of PAL3 were found in 6 species, but 43 C4H gene copies occurred in 17 species with as many as 10 orthologs (BnC4H) in B. napus (Supplementary Table S2). However, TT4, TT6, and AHA10 were not identified in all Brassicaceae species. For example, TT4 was not found in A. halleri and C. rubella, TT6 was absent from A. halleri, and AHA10 was not present in A. lyrata (Supplementary Table S2). We used the sequences of flavonoid biosynthesis genes from A. thaliana as query to search genome databases, and identified highly conserved sequences, including LESS ADHESIVE POLLEN/CRYPTOPHORIC POLLEN (LAPS/CPF) and TT4 in A. halleri and C. rubella, DMR-LIKE OXYGENASE1/2 (DLO1/2) and TT6 in A. halleri, and members of the Autoinhibited H(+)-ATPase (AH) superfamily, which might be homologs of TT4, TT6, and AHA10 and function in flavonoid biosynthesis in A. lyrata. Here, 36 copies of 21 genes of the flavonoid biosynthesis pathway were identified in A. thaliana, but 53 were identified in B. rapa, 50 in B. oleracea, and 95 in B. napus, in accordance with the fact that sequences present as a single copy in the A. thaliana were present in 2–8 copies in B. rapa (Cavell et al., 1998), and with the observation that excessive gene loss is typical after polyploidization in eukaryotes (Sankoff et al., 2010; Wang et al., 2011).

Genomic Distribution on Chromosomes of Brassicaceae Species

All of the flavonoid biosynthesis pathway genes in 17 Brassicaceae species were mapped onto pseudo-molecules or chromosomes using GFF files of their nucleotide sequences (Supplementary Table S3). Brassica crops are the ideal model for studying genome evolution (Wang et al., 2011). To intuitively assign the physical position to the chromosomes of B. rapa, B. oleracea, and B. napus, all gene members of flavonoid biosynthesis pathway were mapped to their chromosomes [52 (98.1%) gene copies in B. rapa, 39 (78.0%) gene copies in B. oleracea, and 84

### Table 3 | Primers of the flavonoid biosynthesis pathway genes and housekeeping gene used for qRT-PCR.

| Target gene | Forward sequence (5′–3′) | Reverse sequence (5′–3′) | GenBank | AGI number |
|-------------|--------------------------|--------------------------|---------|------------|
| BnTT4       | GACTACTAATCTTGCCATCCAAACAG | GCTACTAGTTAGGGACTTCAAGCC | AF076335 | AT5G13930  |
| BnTT5       | CTTCTCGTGGCGGCGAGGATG     | ACAAGTTCCTGTTACTTTCTTG   | EU042417 | AT3G55120  |
| BnTT6       | CTGTTCCAGGGTCAAGGGAGGAT   | TGTTGCACAGGTGATGTTGACG   | E3G1329  | AT3G51240  |
| BnTT7       | GCCATAGGGCAGCCAGGGGGA     | GCTCTCAGGCGGAGTACCTTCG   | DQ324379 | AT3G07990  |
| BnTT8       | AGACGCTGACTGACTACAGC      | AGACTGCTGCAGACAGGGGCA    | DQ767950 | AT3G42800  |
| BnTT9       | TTCTCTCCTAGCTGACTACCTTC   | GCCACCAAGTCGTTCAAGGTA    | HOMB0509 | AT5G48100  |
| BnTT10      | GAGCTGTCGCGCGAAGACGGGCT   | GCTGCTGCCTGCAAGAGCAGA    | NM_101587 | AT1G17260 |
| BnAHA10     | AGCTGTCGCGCGAAGACGGGCT   | GCTGCTGCCTGCAAGAGCAGA    | NM_101587 | AT1G17260 |
| BnTT12      | GCCTCAGAACGACTACAGC       | AGCTGTCGCGCGAAGACGGGCT   | E3G1329  | AT3G51240  |
| BnTT13      | GCCTCAGAACGACTACAGC       | AGCTGTCGCGCGAAGACGGGCT   | E3G1329  | AT3G51240  |
| BnTT14      | CACTTCTTCCTAGCTGACTACCTTC | GCCACCAAGTCGTTCAAGGTA    | HOMB0509 | AT5G48100  |
| BnTT15      | GCTGTCGCGCGAAGACGGGCT   | GCTGCTGCCTGCAAGAGCAGA    | NM_101587 | AT1G17260 |
| BnTT16      | TGGGTTTGCTGCTGTTCTCTCTCT | GCCACCAAGTCGTTCAAGGTA    | E3G1329  | AT3G51240  |
| BnTT17      | AAACTCGAGGGAAATGGCTGCTCGAACA | ACTTCCTTTGACTCTTGTTGCTCTG | EU320168 | AT2G37260  |
| BnTT18      | TGGGTTTGCTGCTGTTCTCTCTCT | GCCACCAAGTCGTTCAAGGTA    | HOMB0509 | AT5G48100  |
| BnTT19      | ACAAATGACCGCGAGTCAGGAGGCT | GCTGCTGACCTCGGCTGAGAGGCT | HOMB0509 | AT5G48100  |
| BnACTN7     | TGGGGTTGGCTGCTGTTCTCTCTCT | GCCACCAAGTCGTTCAAGGTA    | HOMB0509 | AT5G48100  |
| BnUBC21     | CACTTCTTCCTAGCTGACTACCTTC | GCCACCAAGTCGTTCAAGGTA    | HOMB0509 | AT5G48100  |

Primers for amplifying partial sequences of flavonoid biosynthesis genes were designed from conserved nucleotide regions identified by multiple alignments of sequences.
(88.4%) gene copies in *B. napus* and 22 (1 in *B. rapa*, 11 in *B. oleracea*, and 11 in *B. napus*) were distributed on the unanchored scaffolds (Figure 2, Supplementary Table S3), which showed strong collinearity between the A subgenomes from *B. rapa* and *B. napus* and the C subgenomes from *B. oleracea* and *B. napus* (Figure 2). This result indicates that the gene copies are distributed in orthologous blocks in each genome, and that substantial genome reshuffling had occurred. For example, copies of *TT10* and *TT16* were not identified on *B. napus* chromosome A02, and the differential gene copies located in orthologous blocks on chromosome C06 differed greatly between *B. oleracea* and *B. napus* (Figure 2). These findings are consistent with the fact that *Brassicaceae* genomes underwent *Brassicaceae*-lineage-specific whole genome triplication, followed by diplloidization after divergence from their common ancestor (Lysak et al., 2005; Town et al., 2006; Mun et al., 2009; Wang et al., 2011; Cheng et al., 2013).

**Phylogenetic Analysis and the Triplication of the Flavonoid Biosynthesis Pathway Gene Superfamily**

Despite different flavonoid biosynthesis pathway genes having similar functions in *Brassicaceae* species, the phylogenetic relationships of members of each superfamily have not been comprehensively studied. To gain insight into the functional roles of flavonoid biosynthesis pathway genes that may have arisen during evolution, we performed a phylogenetic analysis using the putative amino acid sequence based on each member of the flavonoid biosynthesis pathway superfamily in *Brassicaceae* species.

Initially, we conducted a BLASTP analysis against the local database, which included 17 *Brassicaceae* species, followed by HMM searching (Finn et al., 2011). We identified 649 sequences of 21 flavonoid biosynthesis pathway gene family members, and found that the number of gene copies varied from 1 to 10 in each family (Supplementary Table S2), which were mapped onto pseudo-molecules or chromosomes using GFF files of their nucleotide sequences (Supplementary Table S3). To intuitively assign the physical position to the chromosomes of *B. rapa*, *B. oleracea*, and *B. napus*, all gene members of flavonoid biosynthesis pathway were mapped to their chromosomes (52 (98.4%) gene copies in *B. rapa*, 39 (78.0%) gene copies in *B. oleracea*, and 84 (88.4%) gene copies in *B. napus*) and 22 (1 in *B. rapa*, 11 in *B. oleracea*, and 11 in *B. napus*) were distributed on the unanchored scaffolds (Figure 2, Supplementary Table S3), which showed strong collinearity between the A subgenomes from *B. rapa* and *B. napus* and the C subgenomes from *B. oleracea* and *B. napus* (Figure 2). This result indicates that the gene copies are distributed in orthologous blocks in each genome, and that substantial genome reshuffling had occurred. For example, copies of *TT10* and *TT16* were not identified on *B. napus* chromosome A02, and the differential gene copies located in orthologous blocks on chromosome C06 differed greatly between *B. oleracea* and *B. napus* (Figure 2). These findings are consistent with the fact that *Brassicaceae* genomes underwent *Brassicaceae*-lineage-specific whole genome triplication, followed by diplloidization after divergence from their common ancestor (Lysak et al., 2005; Town et al., 2006; Mun et al., 2009; Wang et al., 2011; Cheng et al., 2013).

In addition, we performed a phylogenetic analysis to identify each flavonoid biosynthesis pathway gene type, and constructed

![FIGURE 2 | Genomic distribution of the flavonoid biosynthesis pathway genes on *B. rapa*, *B. oleracea*, and *B. napus* chromosomes. The dissociation fraction of chromosomes indicates the normal single-copy locus mapped to the random chromosome; Ann and Cnn are the indeterminate Chromosome A and C. Numbers in parentheses are the physical distance of the scaffold for each gene. The collinearity of homologous genes were indicated by the ligatures among *Brassica* species. The homologous genes of flavonoid biosynthesis pathway were indicated by the same colors of the exact same fonts among different chromosomes of *Brassica* species.](https://www.frontiersin.org)
their corresponding phylogenetic trees using MEGA 6.0 with the ML method and modified the tree using FigTree v1.4.2. According to the phylogeny generated using the ML method, all representative sequences for each gene member from the Brassicaceae species formed a well-supported clade, which was classified into the Arabidopsis or Brassicaceae subgroups, but each gene was assigned to monophyletic clades using other substitution models (Figures 3–5, and Supplementary Figures S1–S10). For example, we identified 4 members of PAL, which encodes key enzymes of the phenylpropanoid pathway, that were clearly grouped into four subclades (Figure 3). The gene members, Bn_PAL1b and Si_PAL3 belonged to the same PAL2 subcategories, and Aa_PAL4 and Cs_PAL4c clustered alone or in the vicinity of the PAL3 and PAL4 group (Figure 3; PAL1, Gray; PAL2, Light blue; PAL3, Pink; PAL4, Green). We could thus predict the functions of specific flavonoid biosynthesis pathway genes in Brassicaceae species, because genes with similar functions tend to be retained with orthologous genes (Koonin, 2005). In addition, Brassica species are an ideal model for systematically studying polyploid genome evolution. We identified one clade of genes of the flavonoid biosynthesis pathway, and the copies in B. rapa, B. oleracea, and B. napus (e.g., C4H, TT4, TT6, TT18, TT19, and UGT2) were divided into different subclades or monophyletic subclades by phylogenetic analysis (Figures 4A–F). However, multiple gene copies of TT12, TTG2, and BAN were present in the subclades in B. rapa, B. oleracea, and B. napus, which have high levels of divergence from their common ancestor with A. thaliana (Figures 5A–C).

Based on the collinearity of orthology, we identified syntenic orthologous genes of the flavonoid biosynthesis pathway among the genomes of the Brassicaceae species (Supplementary Table S4). In the genomes of B. rapa, B. oleracea, and B. napus, the copies of C4H were distributed in the LF, MF1, and MF2 subgenomes. Besides, the copies of C4H were also identified outside the genome triplication segments in the genomes of Brassica crops. However, 10 of 25 gene members had a single copy in the B. rapa and B. oleracea genomes, but multiple copies in B. napus. Moreover, we identified 10 pseudo-copies of five genes in seven species, i.e., 3 of which in C. rubella (C4H, TT10, and UGT2), 2 copies of TT1 in L. alabamica and S. irio, and 5 copies of TT6 in A. arabicum, B. rapa, B. oleracea, and B. napus (Supplementary Table S4). These findings can be used to systematically study gene retention in the triplicated genomes of Brassicaceae species, as well as to understand the evolutionary history of these orthologous genes among Brassicaceae species.

### Development- and Tissue-Specific Expression of TT-Type Genes in B. napus

In this study, we analyzed the expression patterns of 18 TT-type genes involved in flavonoid biosynthesis by qRT-PCR in various tissues of black- and yellow-seeded lines of B. napus with different genetic backgrounds. Except for BnTT4, BnTT5, BnTT6, and BnTT7, the expression level of 11 structural genes was much higher in developing seeds than in other tissues in B. napus (Figure 6). The expression level of BnBAN, BnTT12, BnTT19, and BnAHA10 peaked during the early stages of seed development, which were the EBGs (Figures 6G–I,K, Table 4). BnTT3, and BnTT18 were expressed in a similar pattern as BnBAN, BnTT12, BnTT19, and BnAHA10, but peaked later in development, which were the LBGs (Figures 6E,F,J, Table 4). Thus, these genes may be regulated by the same upstream gene or they may have a synergistic effect on the flavonoid biosynthetic pathway during seed development. In addition, the expression levels of BnTT5 and BnTT7 did not differ significantly (Student’s t-test, P > 0.05) among the rapeseed tissues, whereas BnTT4, BnTT5, BnTT6, and BnTT7 expression was higher in the buds and flowers of B. napus (Figures 6A–D, Table 4). Furthermore, the expression level of five regulatory genes, BnTT2, BnTT8, BnTT16, BnTTG2, and BnTT1, was much higher in developing B. napus seeds, and peaked during the early and middle stages of seed development, which may be the EBGs (Figures 6L,M,O,P, Table 4). However, BnTT15 was expressed at higher levels in flowers than in other organs, indicating that this gene may play an important role in flower formation (Figure 6R).

### Differential Expression Patterns of TT-Type Genes in B. napus

In this study, most of the key genes involved in flavonoid biosynthesis were found to exhibit similar expression patterns, and the expression levels of genes varied greatly between the yellow- and black-seeded lines of B. napus (Figure 7). The structural genes involved in flavonoid biosynthesis, i.e., BnTT3, BnTT18, BnBAN, BnTT12, BnTT19, and BnAHA10, had similar expression patterns in the developing seeds of different genetic backgrounds, but had significant differences (Student’s t-test, P < 0.01) in expression level between the black- and yellow-seeded lines (Figures 7C,E–I,K, Table 4). BnTT4, BnBAN, BnTT12, BnTT19, and BnAHA10 expression peaked at 20 DAP, which were EBGs (Figures 7A,G–I,K, Table 4), and BnTT6, BnTT3, and BnTT18 expression peaked at 40 DAP, were belong the LBGs (Figures 7E,F, Table 4). These genes were expressed at higher levels in the black-seeded than in the yellow-seeded lines, especially at 20 DAP (Figures 7A,H,K, Table 4), 30 DAP (Figures 7G,I), and 40 DAP (Figures 7B,E,F, Table 4), but the expression patterns of these genes were hardly affected by genetic background. Additionally, BnTT7, BnTT5, and BnTT10 expression peaked at different developmental stages both in the black- and yellow-seeded lines (Figures 7B,D,J, Table 4, and Supplementary Table S5), indicating that the expression of these genes was affected by the genetic background. As for the expression patterns of structural genes, regulatory genes, such as BnTT2, BnTT8, BnTTG2, BnTT1, and BnTT15, had similar expression patterns in the black- and yellow-seeded B. napus, with the genetic background having a minimal effect, and the expression of these genes peaked during the early and middle developmental stages (20 and 30 DAP, EBGs; Figures 7L,M,O–R, Table 4). In addition, the expression patterns and levels of BnTTG1 showed obvious variations during seed development of different lines of B. napus, suggesting that they were significantly affected by genetic background (Figure 7N).

Although the expression patterns of most of the genes involved in flavonoid biosynthesis did not significantly differ
FIGURE 3 | Phylogenetic relationships of PAL gene family of Brassicaceae species. \( \text{PAL1, PAL2, PAL3, and PAL4} \) were indicated by gray, light blue, purple, and green color. The \( \text{Bn\_PAL1b and SI\_PAL3} \) were denoted by red font that may be synonymous of \( \text{PAL2} \). The Red color diamond, triangle, circle, and rectangle were denoted the gene copies in \( \text{A. thaliana, B. rapa, B. oleracea, and B. napus} \), respectively. Scale bar (the numbers) indicates the estimated number of amino acid substitutions per site.

(Student’s \( t \)-test, \( P > 0.05 \)) between the developing seeds of the black- and yellow-seeded \( B. \ napus \), the expression levels of these genes did differ and peaked at different developmental stages (Figure 7), indicating that these genes can also be classified as Early Biosynthetic Genes and Late Biosynthetic Genes, as in \( \text{Arabidopsis} \) (Lepiniec et al., 2006). The expression of genes that function upstream in the flavonoid biosynthesis pathway peaked before those that functioned later, suggesting that the downstream genes were influenced by the upstream genes. For example, the expression of \( \text{BnTT4} \), which encodes a protein that catalyzes the first committed step of flavonoid biosynthesis (Albert et al., 1997; Tang et al., 1997), peaked at 20 DAP, which was before the expression of the downstream genes \( \text{BnTT5, BnTT6, and BnTT7} \) peaked (Figure 7). All flavonoids and isoflavonoids are derived from the naringenin chalcone generated by this first enzyme. Moreover, the expression...
**FIGURE 4 | Phylogenetic relationships of C4H, TT4, TT6, TT18, TT19, and UGT2 gene family of Brassicaceae species.** The red color diamond, triangle, circle, and rectangle were denoted the gene copies in *A. thaliana*, *B. rapa*, *B. oleracea*, and *B. napus*, respectively. Scale bar (the numbers) indicates the estimated number of amino acid substitutions per site. (A–F) were indicated *C4H*, *TT4*, *TT6*, *TT18*, *TT19*, and *UGT2* gene family, respectively.
of most of the regulatory genes also peaked at 20 DAP (Figure 7), suggesting that they may be essential for regulating the expression of genes involved in flavonoid biosynthesis during the early stages of seed development. Although, the expression of BnTTG2 peaked later than that of BnTT2 and BnTT8, the expression of BnTT2 was greater in yellow-than in black-seeded B. napus, and BnTTG1 expression was also largely affected by the genetic background (Figures 7L–N,P), indicating that the mechanism of BnTTG2 is different from that of its orthologs in A. thaliana (Lepiniec et al., 2006). In addition, the expression of BnTT6, BnTT18, and BnTT2 peaked sooner in the yellow-seeded than in the black-seeded lines (Figures 7C,F,L).

**DISCUSSION**

Flavonoids are secondary metabolites that are extensively distributed in the plant kingdom. They not only play an important role in color formation in fruits and flowers, but are also well-known for their positive effect on health, due to their antioxidant and antitumor properties (Winkel-Shirley, 2001, 2002; Lepiniec et al., 2006; Routaboul et al., 2006). Genome-wide analyses of gene superfamilies have been widely performed after the completion of numerous plant genome projects. For instance, numerous genome-wide studies of gene superfamilies have been performed in the diploid species, B. rapa and B. oleracea (Song et al., 2013; Duan et al., 2014; Li et al., 2014; Diehn et al., 2015; Lu et al., 2015). Several gene superfamilies have also been reported in the allotetraploid species B. napus (Sun et al., 2014; Raboanatahiry et al., 2015). However, a comprehensive analysis of the superfamily of genes underlying the flavonoid biosynthesis pathway in Brassica has not been reported to date. Here, we identified 649 nucleotide sequences of 21 flavonoid biosynthesis pathway genes and their corresponding amino acid sequences in 17 Brassicaceae species (Table 2, Supplementary Table S2). The gene copy numbers differed between species; for example, 10 copies of PAL3 were found in 6 species, whereas 43 copies of C4H were identified in 17 species (Table 2, Supplementary Table S2). Although we did not detect orthologs for some genes involved in flavonoid biosynthesis (e.g., TT4, TT6, and AHA10), in accordance with the fact that the excessive gene loss is typical after polyploidy formation in eukaryotes (Sankoff et al., 2010; Wang et al., 2011), we identified orthologs of many of these genes by querying the Brassica genome databases with the sequences of A. thaliana genes involved in flavonoid biosynthesis. In addition, it has been widely suggested that the genome structures are highly conserved among Brassica species (Krishnamurthy et al., 2014; Thamilarasan et al., 2014; Dong et al., 2016). Each of the gene copies was found to be distributed in orthologous blocks by collinearity analysis between the A and C subgenomes (Figure 2). Not all gene members could be accurately annotated on chromosomes and the number of gene copies varied greatly in the orthologous blocks (Figure 2, Supplementary Table S3), indicating that may be associated with Brassicaceae-lineage-specific whole genome triplication, followed by diploidization after divergence from the common ancestor (Lysak et al., 2005; Town et al., 2006; Mun et al., 2009; Wang et al., 2011; Cheng...
et al., 2013). Although, subgenome sequences present higher levels of conservation in extensive collinear genome blocks among Brassicaceae species, we found that all genes of the flavonoid biosynthesis pathway were phylogenetically classified into two major subcategories (Arabidopsis and Brassicaceae species; Figures 3–5, Supplementary Figures S1–S11), consistent with the functional divergence of orthologous gene groups between Arabidopsis and Brassicaceae species during evolution. These results revealed that diversification occurred among the flavonoid biosynthesis pathway gene family members, likely indicating that functional divergence of orthologous gene groups occurred between Arabidopsis and Brassicaceae species during evolution. These findings provide insight into the functional divergence of these genes among Brassicaceae species. In addition, the identification of conserved genomic blocks will provide useful phylogenetic, polyploidization, and comparative genomics information (Schranz et al., 2006; Cheng et al., 2013). Subgenomes can be classified based on gene density into the following three groups: least fractionated (LF), medium fractionated (MF1), and most fractionated (MF2) (Wang et al., 2011; Cheng et al., 2013). We then performed the triplication of flavonoid genes in the whole genomes of Brassicaceae species. Furthermore, 10 pseudo-copies of five flavonoid biosynthesis pathway genes were identified, such as C4H, TT10, UGT2, TT1, and TT6 (Supplementary Table S4). These results will provide detailed information for systematic studies of the functions and roles of these genes in flavonoid biosynthesis pathway at the molecular level.

Seed coat color was previously reported to involve a similar mechanism in Brassica and Arabidopsis species (Marles and Gruber, 2004). Hence, identifying candidate genes by cloning Brassica TT genes involved in the flavonoid biosynthetic pathway and conducting comparative studies of these genes is a reasonable approach, and many homologs of these genes have also been identified in B. napus (Wei et al., 2007; Xu et al., 2007; Ni et al., 2008; Chai et al., 2009; Lu et al., 2009; Chen et al., 2013). However, little is known about the mechanism underlying seed color formation in B. napus. Using three groups of B. napus plants in different genetic backgrounds, we showed that the 12 genes (BnTT3, BnTT18, BnBAN, BnTT12, BnTT19, BnTT10, BnAHA10, BnTT2, BnTT8, BnTTG1, BnTT16, BnTTG2, and BnTT1) investigated in this study were highly expressed and showed...
TABLE 4 | The tissue specificity and expression stages of flavonoid biosynthesis pathway genes in B. napus and A. thaliana.

| Name   | Tissue specificity | Expressed during |  |
|--------|--------------------|------------------|---|
|        | B. napus           | A. thaliana      |  |
| TT4    | Except Silique pericarps | Whole plant     | EBGs |
| TT5    | Whole plant        | Whole plant      | EBGs |
| TT6    | Whole plant        | Whole plant      | NA  |
| TT7    | Whole plant        | Whole plant      | NA  |
| TT3    | Development seeds  | Whole plant      | EBGs |
| TT18   | Development seeds  | Development seeds| EBGs |
| BAN    | Development seeds  | Development seeds| EBGs |
| TT10   | Development seeds  | Development seeds and floral organs | NA |
| AHA10  | Development seeds and stem | Development seeds and stem | EBGs |
| TT2    | Development seeds  | Development seeds| EBGs |
| TT8    | Development seeds  | Development seeds| EBGs |
| TTG1   | whole plant        | whole plant      | LBGs |
| TT16   | Development seeds and flowers | Development seeds and flowers | EBGs |
| TTG2   | Whole plant        | Whole plant      | LBGs |
| TT1    | Development seeds and floral organs | Development seeds | EBGs |
| TT15   | Whole plant        | Whole plant      | LBGs |

EBGs and LBGs means the Early Biosynthetic Genes and the Late Biosynthetic Genes, respectively. NA indicates that the expression stage of genes were inconclusive during the developmental seeds of B. napus.

clear divergence in organ specificity in the developing seed (Figure 6), suggesting that these genes play an important role in seed development and are involved in the accumulation of seed pigmentation. Furthermore, BnTT15 was highly expressed in flowers (Figure 6R). By contrast, BnTT5, BnTT7, and BnTTG1 expression did not differ significantly (Student's t-test, \( P > 0.05 \)) among B. napus organs (Figures 6B,D,N). Similar expression profiles were observed for orthologs of these genes in A. thaliana (Schmid et al., 2005), suggesting evolutionary conservation of the regulatory mechanism governing flavonoid accumulation. Our study lays the foundation for future research aimed at deciphering the expression profiles of different gene copies in B. napus. These findings provide insight into the characteristics and functions of flavonoid pathway genes in B. napus.

As in A. thaliana, the flavonoid biosynthesis pathway has been characterized mainly using \( tt \) mutants that exhibited a transparent and colorless testa (seed coat; Yu, 2013). Moreover, much research has focused on identifying the seed pigments involved in the formation of seed coat color in B. napus (Theander et al., 1977; Marles and Gruber, 2004; Akhov et al., 2009; Qu et al., 2013). Homologous genes in the B. napus flavonoid biosynthesis pathway have also been cloned and characterized (Wei et al., 2007; Xu et al., 2007; Ni et al., 2008; Akhov et al., 2009; Auger et al., 2009; Chai et al., 2009; Lu et al., 2009; Chen et al., 2013). However, a few of these genes was comprehensively functionalized in B. napus. Using qRT-PCR analysis, we now examined the temporal and spatial expression patterns of 18 flavonoid biosynthesis genes in the developing seeds of black- and yellow-seeded B. napus, sourced from different backgrounds. We found that the majority of genes had similar expression patterns in the developing seed, suggesting that these genes not only participated in the flavonoid pathway, but also might be regulated by an upstream regulatory gene involved in seed coat color formation in B. napus. In addition, we classified these genes based on the time at which their expression peaked. The first group of genes with similar expression patterns in black- and yellow-seeded lines (BnTT4, BnBAN, BnTT12, BnTT19, BnAHA10, BnTT8, BnTT16, BnTTG2, and BnTT15) was expressed at different levels in different tissues and showed the highest expression levels at 20 DAP, and was considered as the early biosynthetic genes (Figures 7, 8). By contrast, BnTT4, BnTT12, and BnAHA10 have higher expression levels in black-seeded than in yellow-seeded lines (Figures 7A,H,K). The first dedicated step for flavonoid biosynthesis of plant is catalyzed by TT4, which produces naringenin chalcone, and thus TT4 critically influences many important flavonoid-related characteristics, such as seed coat color, flower color, and pigmentation of the stem and leaf surface (Hoffmann et al., 2006; Kasai et al., 2007). Chai et al. (2009) proposed BnTT12 a potential candidate gene for seed coat color formation in B. napus. Stein et al. (2013) found that the transparent testa gene AHA10 strongly affected both seed color and lignin content using a marker that was closely linked to a major QTL for seed fiber and color in B. napus. Moreover, TT12 and AHA10 in Arabidopsis were both found to be related to the vacuolar transport of proanthocyanidin in seed coats (Debeaujon et al., 2001; Baxter et al., 2005). Thus, the difference in seed coat color between black- and yellow-seeded rapes seems to be related to the reduction in precursor accumulation following down-regulation of the encoding gene (Figure 8). Another group of genes (BnTT5, BnTT7, BnTT18, BnBAN, BnTT19, BnTT2, and BnTTG2) exhibited drastic differences...
FIGURE 7 | Expression pattern analysis of genes involved in flavonoid biosynthesis in the developing seeds of different *B. napus* lines by qRT-PCR. The yellow-seeded lines include Y1, Y2, and Y3, and the black-seeded lines include B1, B2, and B3 (A–R). Expression was monitored at five stages of seed development (10, 20, 30, 40, and 50 DAP). Values represent the average ± SD of three biological replicates with three technical replicates of each developmental stage. Error bars denote standard error of the mean (SEM) of three experiments. Relative gene expression levels were normalized according to the expression values in black-seeded *B. napus* at 10 DAP.
FIGURE 8 | The pathway of flavonoid biosynthesis, modification, and transport in *B. napus*. *Bn*, *Brassica napus*; *BAN*, BANYULS; AHA10, Autoinhibited H^+^-ATPase isoform 10; TT(G), TRANSPARENT TESTA (GLABRA); UGT, UDP flavonoid glucosyl transferase. The Early Biosynthetic Genes (EBGs) and the Late Biosynthetic Genes (LBGs) are red and blue colors, respectively. The bold means that genes might play an crucial roles in the formation of different seed coat colors in *B. napus.*

during the middle to late stages of seed coat development between the black- and yellow-seeded *B. napus* (Figures 7B,D,F,G,I,L,P). Additionally, TTG1, TT2, and TT8 were found to modulate the activity of proteins encoding TT3 (DFR), TT18 (LDOX), BAN, and TT12 in the proanthocyanidin subpathway, respectively. Moreover, we previously found that PAs and polyphenol compounds gradually increase during seed maturation and result in significant differences (Student's *t*-test, *P* < 0.05 or 0.01) in the colors of black- and yellow-seeded *B. napus* seed coats (Qu et al., 2013). However, *BnTT5* and *BnTTG1* were found to possess different expression patterns among the rapeseed seeds, suggesting that they were largely influenced by the genetic background (Figures 7B,N). These data suggest that numerous compounds accumulate in the seed coat during later development stage and confer color to the mature seed. Therefore, inhibiting the expression of genes involved in their biosynthesis can lead to lighter color seeds during the middle and late development stages in *B. napus*, and improve the nutritional quality of rapeseed oil and meal. Moreover, this can also explain why the expression of upstream genes peaked sooner than or simultaneously with the downstream genes in the developing seeds (Figures 7, 8), indicating that the upstream genes not only control the downstream genes, but also that the upstream genes cooperate in the flavonoid biosynthesis pathway. Together, the flavonoid biosynthetic pathway of *Brassica* species is much more complex than that in *A. thaliana*, with the former not only having more synthesis-related genes, but also exhibiting interactions with other genes involved in flavonoid biosynthesis at multiple loci (Figure 8). These findings provide insight into the molecular and biochemical mechanism of seed coat color development in *B. napus.*

**AUTHOR CONTRIBUTIONS**

CQ and FF designed and wrote the manuscript. HZ and ZW performed the data mining and gene expression analysis. YZ and XW collected the flavonoid gene sequences and bioinformatics.
ACKNOWLEDGMENTS

This work was supported by the National Science Foundation of China (31401412, U1302266, 31571701), the utilization of heterosis and selection of strong advantage of hybrid (2016YFD0101300), Projects in the National Science and Technology Pillar Program (2013BAD01B03-12), the Chongqing Project of Main Crop Variety Innovation (cstc2012gbB00088), the 111 Project (B12006), Chongqing Basic Scientific and advanced technology Research (cstc2015jcyjB20001), and Fundamental Research Funds for the Central Universities (XDJK2016B030).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2016.01755/full#supplementary-material
Kasai, A., Kasai, K., Yamamoto, S., and Senda, M. (2007). Structural features of GmIRChs, candidate of the Ig gene inhibiting seed coat pigmentation in soybean: implications for inducing endogenous RNA silencing of chalcone synthase genes. *Plant Mol. Biol.* 64, 467–479. doi: 10.1007/s11103-007-9169-4

Kitashiba, H., Li, F., Hirakawa, H., Kawanebe, T., Zou, Z., Hasegawa, Y., et al. (2014). Decipher sequences of the radish (*Raphanus sativus* L.) genome. *DNA Res.* 21, 481–490. doi: 10.1093/dnares/dsu014

Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics. *I. Annu. Rev. Genet.* 39, 309–338. doi: 10.1146/annurev.genet.39.070303.114725

Krishnamurthy, P., Hong, J. K., Kim, J. A., Jeong, M.-J., Lee, Y.-H., and Lee, S. I. (2014). Genome-wide analysis of the expansion gene superfamily reveals *Brassica rapa*-specific evolutionary dynamics upon whole genome triplication. *Mol. Genet. Genomics* 290, 521–530. doi: 10.1007/s00438-014-0945-0

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1900

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Lee, T.-H., Tang, H., Wang, X., and Paterson, A. H. (2013). PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res.* 41, D1152–D1158. doi: 10.1093/nar/gks1104

Lepiniec, L., Debeaujon, I., Routaboul, J.-M., Baudry, A., Pourcel, L., Nesi, N., et al. (2006). Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* 57, 405–430. doi: 10.1146/annurev.arplant.57.032905.105252

Li, J., Gao, G., Xu, K., Chen, B., Yang, L., Li, F., et al. (2014). Genome-wide survey and expression analysis of the putative non-specific lipid transfer proteins in *Brassica rapa* L. *PLoS ONE* 9:e88456. doi: 10.1371/journal.pone.0084556

Li, X., Chen, L., Hong, M., Zhang, Y., Zu, F., Wen, J., et al. (2012). A large insertion in HLH transcription factor *BrTT8* resulting in yellow seed coat in *Brassica rapa*. *PLoS ONE* 7:e44145. doi: 10.1371/journal.pone.0044145

Li, Y., Liu, X., Li, J., and Xu, X. (2012). Construction of near – infrared reflectance spectroscopy model for seed color of rapeseed. *Chin. J. Oil Crop Sci.* 34, 533–536.

Lu, J., Li, J., Lei, B., Wang, S., and Chai, Y. (2009). Molecular cloning and characterization of two *Brassica napus* TTG1 genes reveal genus-specific nucleotide preference, extreme protein-level conservation and fast divergence of organ-specificity. *Genes Genomics* 31, 129–142. doi: 10.1007/BF03191146

Lu, K., Guo, W., Lu, J., Yu, H., Qu, C., Tang, Z., et al. (2015). Genome-wide survey and expression profile analysis of the Mitogen-Activated Protein Kinase (MAPK) gene family in *Brassica rapa*. *PLoS ONE* 10:e0132051. doi: 10.1371/journal.pone.0132051

Lysak, M. A., Koch, M. A., Pecinka, A., and Schubert, I. (2014). Comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends Plant Sci.* 11, 535–542. doi: 10.1016/j.tplants.2006.09.002

Song, X., Li, Y., and Hou, X. (2015). Genome-wide analysis of the AP2/ERF transcription factor superfamily in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *BMC Genomics* 14:573. doi: 10.1186/1471-2164-14-573

Stein, A., Wittkop, B., Liu, L., Obermeier, C., Friedt, W., and Snowden, R. J. (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. doi: 10.1111/pbr.12073

Sun, Y., Wang, C., Yang, B., Wu, F., Hao, X., Liang, W., et al. (2014). Identification and Phylogenic analysis of kelch motif containing ACPB in *Brassica rapa*. *BMC Genomics* 16:512. doi: 10.1186/1471-2164-15-375-6

Thamilarasan, S. K., Park, J.-I., Jung, H.-J., and Nou, I.-S. (2014). Genome-wide survey and expression analysis of the putative non-specific lipid transfer proteins in *Brassica napus* L. *Genes. 6*, 1215–1229. doi: 10.3390/genes6041215

Rabonatnathiry, N. H., Yin, Y., Chen, L., and Li, M. (2015). Genome-wide identification and Phylogenetic analysis of kelch motif containing ACPB in *Brassica rapa*. *BMC Genomics* 16:512. doi: 10.1186/1471-2164-15-375-6

Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., et al. (2013). The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity. *Plant Physiol. Biochem.* 72, 21–34. doi: 10.1016/j.plaphy.2013.02.001

Sankoff, D., Zheng, C., and Zhu, Q. (2010). The collapse of gene complement following whole genome duplication. *BMC Genomics* 11:313. doi: 10.1186/1471-2164-11-313

Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., et al. (2005). A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* 37, 501–506. doi: 10.1038/ng1543

Schranz, M. E., Lysak, M. A., and Mitchell-Olds, T. (2006). The ABCs of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends Plant Sci.* 11, 535–542. doi: 10.1016/j.tplants.2006.09.002

Sun, Y., Wang, C., Yang, B., Wu, F., Hao, X., Liang, W., et al. (2014). Identification and functional analysis of mitogen-activated protein kinase kinase kinase (MAPKKK) genes in *Brassica napus* L. *J. Exp. Bot.* 65, 2171–2188. doi: 10.1093/jxb/eru092

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Tang, Z., Li, J., Zhang, X., Chen, L., and Wang, R. (1997). Genetic variation of yellow-seeded rapeseed lines (*Brassica napus* L.) from different genetic sources. *Plant Breed.* 116, 471–474. doi: 10.1111/j.1439-0523.1997.tb01033.x

Thamilarasan, S. K., Park, J.-I., Jung, H.-J., and Nou, I.-S. (2014). Genome-wide analysis of the distribution of AP2/ERF transcription factors reveals duplication and CBP1 genes elucidate their potential function in *Brassica oleracea*. *BMC Genomics* 15:422. doi: 10.1186/1471-2164-15-422

Theander, O., Aman, P., Miksche, G. E., and Yasuda, S. (1977). Carbohydrates, polyphenols, and lignin in seed hulls of different colors from turnip rapeseed. *J. Agric. Food Chem.* 25, 270–273. doi: 10.1021/jf00210a042
Town, C. D., Cheung, F., Maiti, R., Crabtree, J., Haas, B. J., Wortman, J. R., et al. (2006). Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. *Plant Cell Online* 18, 1348–1359. doi: 10.1105/tpc.106.041665

Wan, L., Xia, Q., Qiu, X., and Selvaraj, G. (2002). Early stages of seed development in *Brassica napus*: a seed coat-specific cysteine proteinase associated with programmed cell death of the inner integument. *Plant J.* 30, 1–10. doi: 10.1046/j.1365-313X.2002.01262.x

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., et al. (2011). The genome of the mesopolyploid crop species *Brassica rapa*. *Nat. Genet.* 43, 1035–1039. doi: 10.1038/ng.919

Wei, Y. L., Li, J. N., Lu, J., Tang, Z. L., Pu, D. C., and Chai, Y. R. (2007). Molecular cloning of *Brassica napus* TRANSPARENT TESTA 2 gene family encoding potential MYB regulatory proteins of proanthocyanidin biosynthesis. *Mol. Biol. Reports* 34, 105–120. doi: 10.1007/s11033-006-9024-8

Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 126, 485–493. doi: 10.1104/pp.126.2.485

Winkel-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. *Cur. Opin. Plant Biol.* 5, 218–223. doi: 10.1016/S1369-5266(02)00256-X

Wu, G., Zhang, L., Wu, Y., Cao, Y., Lu, C. (2010). Comparison of five endogenous reference genes for specific PCR detection and quantification of *Brassica napus*. *J. Agric. Food Chem.* 58, 2812–2817. doi: 10.1021/jf904235b

Xie, D. Y., Sharma, S. B., Paiva, N. L., Ferreira, D., and Dixon, R. A. (2003). Role of anthocyanidin reductase, encoded by RANYULS in plant flavonoid biosynthesis. *Science* 299, 396. doi: 10.1126/science.1078540

Xu, B. B., Li, J. N., Zhang, X. K., Wang, R., Xie, L. L., and Chai, Y. R. (2007). Cloning and molecular characterization of a functional flavonoid 3′-hydroxylase gene from *Brassica napus*. *J. Plant Physiol.* 164, 350–363. doi: 10.1016/j.jplph.2006.03.001

Yu, C. Y. (2013). Molecular mechanism of manipulating seed coat coloration in oilseed *Brassica* species. *J. Appl. Genet.* 54, 135–145. doi: 10.1007/s13353-012-0132-y

Zhang, J., Lu, Y., Yuan, Y., Zhang, X., Geng, J., Chen, Y., et al. (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. doi: 10.1007/s11103-008-9437-y

Zhang, K., Lu, K., Qu, C., Liang, Y., Wang, R., Chai, Y., et al. (2013). Gene Silencing of BrTT10 family genes causes retarded pigmentation and lignin reduction in the seed coat of *Brassica napus*. *PLoS ONE* 8:e61247. doi: 10.1371/journal.pone.0061247

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Qu, Zhao, Fu, Wang, Zhang, Zhou, Wang, Wang, Xu, Tang, Lu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.