Mapping and identification of yellow seed coat color genes in Brassica juncea L.

Zhen Huang (huang_zhen.8@163.com)
Northwest Agriculture and Forestry University

Yang Wang
Northwest Agriculture and Forestry University

Hong Lu
Agricultural Bureau of Feixian

Xiang Liu
Northwest Agriculture and Forestry University

Lu Liu
Qinghai Red Cross Hospital

Aixia Xu
Northwest Agriculture and Forestry University

Research article

Keywords: Yellow seed color, Mapping, Cloning, Brassica juncea (B. juncea)

DOI: https://doi.org/10.21203/rs.3.rs-25889/v1

Abstract

Background

Yellow seed breeding is an effective method to improve the oil content in rapeseed. Yellow seed coat color formation is influenced by various factors, and no clear mechanisms are known. In this study, Bulked segregant RNA-Seq (BSR-Seq) of BC_9 population of Wuqi mustard (yellow seed) and Wugong mustard (brown seed) was used to identity the candidate genes controlling the yellow seed color in *Brassica juncea* L.

Results

Yellow seed coat color gene was mapped to chromosome A09, and differentially expressed genes (DEGs) between brown and yellow bulks enriched in the flavonoid pathway. A significant correlation between the expression of *BjF3H* and *BjTT5* and the content of the seed coat color related indexes was identified. Two intron polymorphism (IP) markers linked to the target gene were developed around *BjF3H*. Therefore, *BjF3H* was considered as the candidate gene. The *BjF3H* coding sequences (CDS) of Wuqi mustard and Wugong mustard are 1071-1077bp, encoding protein of 356-358 amino acids. One amino acid change (254, F/V) was identified in the conserved domain. This mutation site was detected in four *Brassica rapa* (*B. rapa*) and six *Brassica juncea* (*B. juncea*) lines, but not in *Brassica napus* (*B. napus*).

Conclusions

The results indicated *BjF3H* is a candidate gene that related to yellow seed coat color formation in *Brassica juncea* and provided a comprehensive understanding of the yellow seed coat color mechanism.

Background

Rapeseed is one of the most crucial oil crops in China. In 2018, the total planting area of rapeseed in China is 6.551 million hectares, but it still cannot meet the needs of the market[1]. In recent years, breeders have tried to increase the value of rapeseed by improving the oil and protein content. A large number of studies have shown that the yellow seeded rapeseed has higher oil and protein content, lower erucic acid content, thinner seed coat, and higher nitrogen phosphorus potassium utilization rate compared with the brown seeded rapeseed at the same genetic background[2]. Therefore, cultivating high yield and good quality yellow seeded rapeseed is considered as one of the most important objectives for rapeseed breeding. However, few natural occurring yellow seeded genotypes have been reported in *Brassica napus* (*B. napus*) and the inheritance of the yellow seed traits is not stable*. Brassica juncea* (*B. juncea*), an allotetraploid species, has some natural yellow seeded germplasms with stable inheritance. For example, the landrace Wuqi mustard is one of the *Brassica juncea* (*B. juncea*) varieties, which has been cultivated in the north of Shaanxi in China. The yellow seed coat color is controlled by a single recessive gene[3], which can be used as the yellow seed coat color gene to create yellow seeded *Brassica napus* (*B. napus*).

In order to understand the mechanism of yellow seed coat color in rapeseed, some studies have been conducted on the pigment compounds associated with yellow seed coat color. The formation of yellow seed coat color in rapeseed is largely determined by the content of phenolic compounds[4–6]. The main phenolic compounds found in rapeseed are flavonoids[7]. Flavonoids include proanthocyanidins (PAs), anthocyanins and flavonols[8]. Marles[9]found that the pigments affecting the seed coat color are mainly proanthocyanins (PAs); however, PAs are rarely detected in mature yellow seeded rapeseed. PAs are the end-products of the flavonoid biosynthesis pathway[10, 11]. The flavonoid biosynthesis pathway has been extensively explored at the genetic and biochemical levels in model plants, including *Arabidopsis thaliana*, *Petunia* and *Zea mays*[12–14]. Some studies showed melanin has a decisive effect on seed coat color[15].

In *Arabidopsis*, the flavonoid biosynthesis pathway can be divided into two stages. The first stage is the phenylpropanoid pathway, *4-coumaroyl-CoA* is synthesized, as well as many secondary metabolites. The phenylpropanoid pathway is regulated by *PAL* and *C4H*. The second stage is the flavonoid pathway initiated by the condensation of one molecule of *4-coumaroyl-CoA* with three molecules of *malonyl-CoA*. So far, 17 genes involved in the flavonoid biosynthesis pathway have been characterized by functional analysis in *Arabidopsis*[16–18], including eight structural genes (*TT3/ DFR*, *TT4/ CHS*, *TT5*, *TT6*, *TT7*, *FLS*, *LDOX* and *BAN)*, six regulatory genes (*TT1*, *TT2*, *TT8*, *TTG1*, *TTG2* and *TT16*) and three genes encoding transfer proteins (*TT12*, *TT19* and *TT10*). A large number of studies have confirmed that there is a great genomic collinearity between *Arabidopsis* genome and *Brassica* genome[19, 20], which makes it possible to analyze the gene functions of *Brassica* species using the genomic information of *Arabidopsis thaliana*. Until now, the regulatory mechanism of flavonoid metabolism in *Brassica juncea* (*B.juncea*) remains unclear. Therefore, the information of these flavonoid biosynthesis genes in *Arabidopsis* contributes to the study of yellow seed coat color in *Brassica juncea* (*B.juncea*).

With the releasing of the genomic information of *Brassica*, it is possible to clone the seed coat color genes in *Brassica*. So far, many yellow seed color genes were successfully cloned, for example, *TTG1* and *TT8* in *Brassica rapa* (*B. rapa*), *TT10* in *Brassica napus* (*B. napus*)[21, 22]. However, only a few studies on the yellow seed color genes in *Brassica juncea* (*B.juncea*) were reported. Previous studies showed that the yellow seed color gene of Wuqi mustard was located on the A09 chromosome and a high resolution genetic and physical map around the yellow seed color gene has been constructed[3], but no gene controlling yellow seed coat color has been isolated. Some other studies showed that the yellow seed color genes in *Brassica* are related to the flavonoid biosynthesis pathway[21]. Whether the candidate genes controlling yellow seed color in our study also play a mojor role in the flavonoid biosynthesis pathway is yet to be understood. Therefore, the objects of this study include: 1) to map the yellow seed color gene in *Brassica juncea* (*B.juncea*); 2) to analyze the relationship between the accumulation of seed coat color related indexes and the gene expression; 3) to identify the candidate genes responsible for yellow seed coat color. These results will lay a great foundation for understanding the mechanism of yellow seed color in *Brassica juncea* L.

Results

Data filtering and short reads assembly

We selected one vital seed development stage to compare the color difference between the Wuqi mustard and the Wugong mustard. At 30 DAP, large differences of the seed color between the two parents were observed (Supplemental Fig. 1). During this stage, the seeds of the Wuqi mustard were yellow, while those of the Wugong mustard began to turn brown. Therefore, the seeds at 30 DAP were prepared for BSR-Seq in *Brassica juncea* (*B.juncea*). Total 41.62-53.90 million pairs of 150 bp raw reads were generated from BSR-Seq. After the data filtering, 40.79-52.82million clean reads remained. Because ribosome contamination affects subsequent analysis, we removed the reads could be mapped to the Ribosomal RNA . The remained reads were used to map to the reference genome (Ensembl release 36 bra), and the Cufflinks Software was used for reference annotation based transcripts assemble.

Gene expression in developing seeds at 30 DAP

Based on RSEM software, the FPKM value was calculated between BB bulk and BY bulk. A total of 169 differentially expressed genes (DEGs) were identified (|log2 FlodChange| > 1, P<0.05), out of which, 157 were significantly up-regulated and 12 were significantly down-regulated in BY seeds compared to BB seeds at 30 DAP. KEGG functional annotation indicated that 17 (10.83%), 13 (8.28%) and 9 (5.73%) of the up-regulated genes were related to flavonoid biosynthesis, phenylpropanoid biosynthesis and phenylalanine, respectively, and no down-regulated genes were related to these three pathways, which is consistent with the previous study that yellow seed coat color is related to flavonoid biosynthesis in rapeseed[21]. Therefore, the DEGs involved in flavonoid biosynthesis were selected as the potential genes in the regulation of seed color, including *BjBAN*, *BjDFR*, *BjLDOX*, *BjTT5*, *BjC4H*, *BjTT8*, *BjTT4*, *BjPLA*, and *BjF3H* (Table 1). Obviously, the expression of the up-regulated DEGs in flavonoid biosynthesis expressed in the BB bulk were significantly higher than that in the BY bulk (Table 1).

Discovery of polymorphic SNPs

Reads from the pooled BB and BY bulks were extracted separately for SNP discovery. A total of 6825 polymorphic SNPs between the pooled BB and BY samples were identified (Supplemental Table 1). A total of 118 SNPs were mapped to scaffolds, and the remaining 6707 SNPs were in 10 chromosomes. The number of SNPs on chromosomes varied with the length of the chromosomes. The number of SNPs on chromosome A09 and A04 was the highest and the lowest, respectively. The mapping of the *candidate genes* was carried out using 6825 SNPs through BSR-Seq. The linkage probability of each SNP was plotted against its physical coordinate in the *Brassica rapa* (*B. rapa)* reference genome (Chiifu-401). Each of the SNP that had a high probability of being linked to the candidate gene clustered on chromosome A09 (Fig. 1). Therefore, the flavonoid biosynthesis DEGs on the A09 chromosome were selected as the candidate genes, including *BjDFR*, *BjF3H*, *BjTT5*, *BjTT4* and *BjTT8*.

Differential expression patterns of candidate genes in *B. juncea*

The majority of the flavonoid biosynthesis DEGs on the A09 chromosome had different expression patterns, and the expression level of the genes was significantly different between the yellow and brown seeded lines. *BjF3H* and *BjDFR* had similar expression patterns in the developing seeds whose expression peaked at 9 DAP. The expression of the *BjTT5* gene peaked at 23 DAP. The expression of *BjTT4* and *BjTT8* peaked at different stages in the yellow and brown-seeded parents, their expression peaked at 16 DAP in the Wugong mustard, and peaked at 23 and 9 DAP in the Wuqi mustard, respectively. The specific information is shown in Fig. 2.

Analysis of candidate genes in *Brassica juncea* (*B. juncea*)

Differences in the content of the seed coat color related indexes in different seed developmental stages were observed. During the development stages of seeds, the flavonoid (except 9DAP and 16DAP), anthocyanidin, melanin and total phenol contents of Wugong mustard were significantly higher than those in Wuqi mustard (p<0.01). The content of flavonoids, melanin and total phenol in both Wuqi mustard and Wugong mustard steadily increased along the seed development stage, with the maximum value detected at 45 DAP. However, the synthesis of these indexes occurred earlier in Wugong mustard than in Wuqi mustard. The difference of flavonoid and anthocyanin content in the two parents peaked at 38 DAP and 30 DAP, respectively. And the difference of melanin and total phenol between the two parents peaked at 45 DAP. The detailed information was shown in Table 2.

As shown in Table 3, the expression of *BjTT5* and *BjF3H* had good correlation with the seed coat color related indexes in Wugong mustard and Wuqi mustard. For *BjTT5*, the expression showed significant or extremely significant positively correlated with the content of anthocyanins and total phenols in the two parents with a correlation coefficient of 0.817, 0.939, 0.848 and 0.811 respectively. For *BjF3H*, the expression was significantly negatively correlated with the content of total phenol in Wugong mustard with a correlation coefficient of -0.874. In Wuqi mustard, the expression of *BjF3H* was also negatively correlated with the content of the four seed coat color related indexes with a correlation coefficient of -0.900, -0.827, -0.836, and -0.878, respectively (Table 3).

In order to determine which one is the most likely candidate gene, twenty pairs of IP primers based on the sequences of chromosome 5 in *Arabidopsis* were designed, among which, two IP primers, IP-at3g51220 and IP-at3g51280, successfully amplified polymorphic bands between the two parents. Thus, these two primers were used to screen the small population and the entire BC_9 population. As a result, all of them showed polymorphism, and no recombinants were identified, which co-segregated with the yellow seed coat color genes. As these two IP markers were from chromosome 5 of *Arabidopsis*, and close to *F3H* (*AT3G51240*), combined with the results of correlation between the expression of the key genes involved in flavonoid biosynthesis and seed coat color related indexes in the two parents, it was possible that *BjF3H* was the candidate gene controlling the yellow seed coat color.

Candidate gene cloning

BjF3H was amplified in the two parents, and the two genes were named as *Wuqi-F3H* and *Wugong-F3H* with the full-length of 3161bp and 3211bp respectively, containing three exons and two introns. Two CDS sequences in Wuqi mustard and one CDS sequence in Wugong mustard were obtained, named *Wuqi-F3H.1*, *Wuqi-F3H.2* and *Wugong-F3H*, separately. The ORF sequences were 1077bp, 1071 bp and 1077 bp, encoded protein of 358, 356 and 358 amino acids.

The isoelectric points of *Wuqi-F3H.1*, *Wuqi-F3H.2* and *Wugong-F3H* were 5.45, 6.73, 5.45, the molecular weights were 40.15kD, 39.86kD, 40.10kDa, respectively. Seven single nucleotide differences between *Wuqi-F3H.1* and *Wugong-F3H* were detected, resulting in one amino acid changes (254, F/V) (Supplemental Fig. 2). Two typical domains, DIOX_N super family domain and 2OG-FeII_Oxy super family domain, were identified between residues 38-154 and 193-294, respectively (Fig. 3), these amino acid changes were identified in the conserved domain between two parents.

BjF3H gene cloning in *Brassica napus* (*B. napus*), *Brassica juncea* (*B. juncea*) and *Brassica rapa* (*B. rapa*)

We cloned *F3H* sequence from 2 yellow *Brassica rapa* (*B. rapa*) genotypes□2 brown *Brassica rapa* (*B. rapa*) genotypes, 3 yellow *Brassica napus* (*B. napus*) genotypes, 3 brown *Brassica napus* (*B. napus*) genotypes, 3 yellow *Brassica juncea* (*B. juncea*) genotypes and 3 brown *Brassica juncea* (*B. juncea*) genotypes. The sequencing results showed that two copies were obtained in the *Brassica napus* (*B. napus*) lines BnB1, BnB2, BnY1, BnY2 and BnY3, and one copy was cloned in *Brassica napus* (*B. napus*) line BnB3. However, only one copy was obtained in all *Brassica juncea* (*B. juncea*) lines. Two copies were cloned in BrY1 and BrY2 of *Brassica rapa* (*B. rapa*), and a copy was cloned in BrB1 and BrB2 of *Brassica rapa* (*B. rapa*). The ORF length of these copies was 1077 bp, encoding a protein of 358 amino acids (Supplemental Table 2).

The results of SNP analysis showed that 42, 8 and 16 SNPs were detected in *Brassica napus* (*B. napus*), *Brassica juncea* (*B. juncea*) and *Brassica rapa* (*B. rapa*), respectively, resulting in nine (7: T/N, 29: V/A, 66: N/S, 161: N/S, 188: N/S, 242: K/E, 324: K/R, 331: I/L and 341: K/Q), one (254: V/F) and one (254: V/F) amino acids changes, respectively. The four *Brassica rapa* (*B. rapa*) lines, six *Brassica juncea* (*B. juncea*) lines and the two parents (Wuqi mustard and Wugong mustard) had the same mutation sites of amino acids (254: V/F) (Supplemental Table 3).

Phylogenetic tree analysis

The BLAST tool developed by BRAD (http://brassicadb.org/brad/blastPage.php) was used to conduct the phylogenetic analysis. *Bra036828* from *Brassica rapa* (*B. rapa*) was clustered together with *Wugong-F3H*; *GSBRNA2T00129251001* from *Brassica napus* (*B. napus*) was clustered together with *Wuqi-F3H.1*; *Bol044664* from *Brassica oleracea* (*B. oleracea*) was clustered together with *Wuqi-F3H.1* (Fig 4). All the genes are involved in flavonoid synthesis. Therefore, it is presumed that the candidate gene can respond to seed coat color. We also found *F3H* from *Brassica napus* (*B. napus*) was clustered together with *F3H* from *Brassica rapa* (*B. rapa*), and most of *F3H* from *Brassica napus* (*B. napus*) were clustered (Fig. 4), indicating that the genetic relationship of *Brassica rapa* (*B. rapa*) and *Brassica napus* (*B. napus*) is closer

than that of *Brassica rapa* (*B. rapa*) and *Brassica juncea* (*B. juncea*) in China. A sequence of 1500 bp upstream of the ATG of *Wuqi-F3H* and *Wugong-F3H* was cloned using the primer F3H-Up and analyzed for *cis*-acting elements using Plant CARE. It showed that a number *cis*-acting elements were identified, including AT-rich, ATBP-1, 5UTR Py-rich stretch, MYB binding site, MBS, ACE, GCN4_motif and TGA, among which, the auxin-responsive element TGA was specific for the upstream of *Wugong-F3H* gene sequence (Table 4).

Discussion

With the development of next-generation sequencing technology and the decline of sequencing cost, the genome sequence of most important species has been obtained. BSR-Seq is a cheap and effective approach for gene mapping, especially for species publishing reference genomes. Contrast with the traditional methods, BSR-Seq could provide adequate genes information, such as SNPs, gene expression data, which greatly accelerate the process of gene mapping. And now the BSR-Seq mapping strategy was widely used to map the key genes in Brassica, maize, wheat and other species[23–26]. This study BSR-Seq approach was used to locate the yellow seed color gene to A09 chromosome and the information of gene expression on A09 chromosome was also provided, which is helpful for screening candidate genes, therefore BSR-Seq is a powerful method for gene mapping and cloning.

Great progress has been made in the understanding of the physiological and biochemical mechanism of the yellow seed coat color in rapeseed. However, the consensus has not been reached. Melanin is considered the main factor in color variation of the seed coat, and polyacid plays an important role in the difference of seed coat color[27]. But some studies have shown that the seed coat color difference between the black seed and the yellow seed is mainly related to polyphenols, flavonoids, anthocyanins and melanin[15, 28, 29]. The main difference in the yellow and black seed in the mustard type rapeseed is the accumulation level of proanthocyanidins[30], which is consistent with our results. Therefore, the mechanism of yellow seed coat color of *Brassica juncea* (*B. juncea*) can be explained according to the variation of these indexes.

Flavonoids are the main secondary metabolites of plants, and *F3H* (flavanone hydroxylase) is a key enzyme in the biosynthesis pathway of plant flavonoids. In 1996, the *F3H* gene was first cloned in *Arabidopsis*[31], and has been reported in a variety of crops, but not yet reported in rapeseed. It was found that when the *F3H* gene of tea tree is transformed into *Arabidopsis thaliana*, the flavonol and proanthocyanidin content shows a significant increase[32]. In seed coats, leaves and stems, the content of anthocyanins and their downstream metabolites in *F3H* mutant is lower than that of the wild type[33]. A previous study also showed that *F3H* plays an important role in plant biotic and abiotic stresses[34]. Overexpression of *F3H* gene can improve drought resistance in tobacco and increase salt tolerance in *Arabidopsis*[35, 36]. *F3H* also catalyzes the formation of dihydroflavonol (DHK) in the flavonoid pathway, which interacts synergistically with different genes in a variety of species[31]. Studies have also shown that some transcription factors have a regulatory effect on the *F3H* gene, such as *MYB*, which can interact with *F3H* to fully reveal the function of *F3H* in mustard-type canola[31]. Therefore, the additional

function of the *F3H* gene in the *Brassica* species should be studied to fully reveal the function of the *F3H* gene in plants.

In this study, two copies were obtained in *Brassica rapa* (*B. rapa*) and *Brassica napus* (*B. napus*), respectively, and many SNPs between these two copies were also identified. It was found that one copy was highly consistent with the genome of *Brassica rapa* (*B. rapa*). Therefore, we speculate that one copy was from the A genome, and the other copy was from the C genome. The results of the phylogenetic tree showed that the genetic relationship of *F3H* between *Brassica rapa* (*B. rapa*) and *Brassica napus* (*B. napus*) was extremely close. According to the “triple hypothesis”, there should be three *F3H* copies in *Brassica rapa* (*B. rapa*) and six *F3H* copies in *Brassica napus* (*B. napus*) and *Brassica juncea* (*B. juncea*). However, we only obtained one or two *F3H* sequences by the homologous cloning method. Although not all copies of the *F3H* gene were obtained in this study, the phylogenetic relationship of the *F3H* gene in different types of rapeseed was reflected to some extent, and similar results have also been reported in previous studies[37].

Conclusions

A total of 157 significantly up-regulated differentially expressed genes (DEGs) were identified in yellow bulk compared to brown bulk, out of which, 17 up-regulated DEGs involved in flavonoid biosynthesis pathway. 6825 polymorphic SNPs between the pooled BB and BY samples were identified through BSR-Seq. Each of the SNP on chromosome A09 had a high probability of being linked to the candidate gene clustered on chromosome A09. Therefore, up-regulated DEGs located on chromosome A09 and involved in flavonoid biosynthesis pathway were considered as candidate genes based on transcriptome profiles. These genes were *BjDFR*, *BjF3H*, *BjTT5*, *BjTT4* and *BjTT8*.

The flavonoid, anthocyanidin, melanin and total phenol contents of Wugong mustard (brown seed)were significantly higher than those in Wuqi mustard (yellow seed) during the seed development stages. We found the expression of *BjF3H* and *BjTT5* have correlation with yellow seed related index. Two intron polymorphism (IP) markers linked to the target gene were developed around *BjF3H*. One amino acid change (254, F/V) between Wuqi mustard and Wugong mustard was identified in the conserved domain. This mutation site was detected in four *Brassica rapa* (*B. rapa*) and six *Brassica juncea* (*B. juncea*) lines. Therefore, *BjF3H* was possible the candidate gene controlling the yellow seed coat color. As far as we know, this work was the first time to provide a comprehensive transcriptome proliles of back crossing population.

Although further research on the functional analysis of *BjF3H* is required, our findings offer great insights into the molecular moleculare mechanisms underlying yellow seed coat color formation.

Materials And Methods

Plant materials and collection of tissues and organs

Two *Brassica juncea* (*B. juncea*) lines, Wuqi mustard and Wugong mustard, were used for gene expression, seed coat color related indexes analysis and candidate gene cloning. Wuqi mustard is yellow-seeded, Wugong mustard is brown-seeded. The plants were grown in a greenhouse (16h/8h, 25 °C/20 °C) at the Northwest A&F University, Yangling, Shaanxi, China in 2015. The seeds (9, 16, 23, 30, 38, and 45 days after pollination) of the two lines were sampled, and immediately frozen in liquid nitrogen and stored at -80 °C for RNA extraction. 10 plants for each line were measured per replicate, and three replicates were assessed. A BC_9 population consisting of 360 individuals derived from these two lines were used to develop the molecular markers linked to the yellow seed coat color genes, as stated in the method described by Huang et al[38].

RNA-Seq and data analyses

A BC_9 population consisting of 180 individuals derived from Wuqi mustard and Wugong mustard was used for RNA-Seq. Wuqi mustard was used as the recurrent parent to cross with the brown seeded individuals of the BC population in each generation. At 30 d post-pollination, seeds from 30 yellow seeded (BY) and 30 brown seeded plants (BB) were combined to form BY and BB bulks, respectively. The two bulks comprised one biological replicate as described. In total, three replicates were assessed. A RNA prep Pure Plant Kit (TIAN GEN, Beijing, China) was used for RNA extraction from each sample following the manufacturer's instructions. The RNA concentration and quality were checked using a NanoDrop 2000c Spectrophotometer and an Agilent Bioanalyzer (RIN) for each sample.

The following analysis was performed using the services of SAGENE (http://www.sagene.com.cn/). The clean reads obtained by sequencing each sample were aligned to the reference genome (Ensembl release 36 bra) using the TopHat2 software as paired-end fragments with ≤5 mismatches, and then the short reads were assembled using the Cufflinks software to obtain the assembly results of each sample. The results for these three biological replicates for BB and BY were combined by the Cuffmerge software. The FDR and log2FC values were used to identify DEGs between the BY and BB pool with the edgeR software. The DEGs were functionally classified using the KEGG database (https://www.genome.jp/kegg/).

Seed coat color related indexs analysis

Four seed coat color related indexes including flavonoids, anthocyanins, melanin and total phenol were measured. The concentrations of flavonoids, anthocyanin and total phenol were measured by the hydrochloric acid methanol method[15, 27]. The brief procedure was as follows: the fresh leaf (0.1g) was grinded in liquid nitrogen with 5mL acetone, and then treated in a 30°C water bash for 1h to extract chlorophyll and lutein. Subsequently, the mixture was centrifuged at 3500r/min for 15min and then the supernatant was discarded. The precipitate was re-suspended in 5mL methanol containing 5% hydrochloric acid, and incubated in a 60°C water bash for 1h. The supernatant was collected and the precipitate was re-extracted three times following the above procedure; the extracts were combined and diluted in methanol containing 5% hydrochloric acid. The absorbance of the mixture was determined at A_{280}nm using a solution containing methanol and 5% hydrochloric acid as a blank control. The total

phenol concentration in the sample was calculated based on a standard curve showing the correlation between gallic acid concentration and absorbance. The flavonoid concentrations were determined using the absorbance at A_{325}nm and the anthocyanin concentrations were determined using the absorbance at A_{530}, A_{620} and A_{630}nm. For melanin, the measurement is basically the same as the method described above, except the last step. After incubated in a 60°C water bash for 1h, the precipitate was immersed in 2 mL 2% NaOH, and then incubated in a 70 °C water bash until the color of the material completely faded. The absorbance of the mixture was determined at A_{290}nm.

RNA extraction and cDNA synthesis

A RNA prep Pure Plant Kit (TIAN GEN, Beijing, China) was used for total RNA extraction from various tissues at different growth stages. The integrity of RNA was assessed by 1% agarose gel electrophoresis. The concentration and purity were assessed on a Thermo NanoDrop 2000 spectrophototometry. cDNA synthesis was performed following the instructions of the TIAN GEN FastQuant RT Kit (Beijing, China). All cDNAs were stored at -20°C for gene cloning and expression.

Expression pattern analysis of related genes

Quantitative real-time PCR (qRT-PCR) was used to analyze the expression of the genes in different tissues from the two parents. Primers for qRT-PCR were designed from conserved nucleotide regions identified by multiple alignments of sequences from the *Arabidopsis* information resource (TAIR: http://www.arabidopsis.org/). Twelve pairs of primers from twelve genes were designed for gene expression using Primer 5.0 (Supplemental Table 4), and synthesized by Sangon Biotech (Shanghai, China). The TIPS-41 gene was used as an internal control[39].

The qRT-PCR experiments were performed according to the instructions of the SYBR® Premix Ex Taq™ II (TliRNaseH Plus) (TaKaRa, DaLian) using a reaction system of 25μL containing 10μL SYBR Premix Ex Taq II (TliRNaseH Plus) (2×), 0.8μL PCR forward primer (10μM), 0.8μL PCR reverse primer (10μM), 0.4μL ROX Reference Dye II (50×), 2μL cDNA (2mg/μL), and 6μL ddH_2O. The amplification was performed by an initial incubation at 95°C for 30s, then 95°C for 5s for a total of 40 cycles, followed by 60 °C for 30s. The expression level of every gene was calculated by the $2^{-\triangle\triangle Ct}$ comparative threshold cycle (Ct) method. The Ct values were generated using an Applied Biosystems 7500 Fast Real-Time PCR System (Life Technologies). All samples were repeated for three times.

Correlation between the expression of key genes and seed coat color related indexes

SPSS 17.0 was used to analyze the correlation between the expression of key genes in the flavonoid biosynthesis pathway and the content of seed coat color related indexes in different seed development stages (9, 16, 23, 30, 38, and 45 days after pollination). Pearson coefficient was adopted in the correlation analysis.

Analysis of candidate genes in *Brassica juncea* (*B. juncea*)

In a previous study, the yellow seed coat color gene has been mapped to chromosome 5 in *Arabidopsis*[3]. Thus, the sequence of chromosome 5 in *Arabidopsis* was randomly selected for designing intron polymorphism (IP) primers, especially the sequences around *TT* genes. The detailed method can be referred to Huang et al[38]. These pairs of primers were used to amplify a small population including 10 yellow and 10 brown individuals derived from the BC_9 population. The primers that showed polymorphism were used to screen the recombinants of BC_9. Additionally, through the correlation comparison of gene expression and the content of seed coat color related indexes, the genes with better correlation were selected. The candidate genes were determined based on the analysis of the expression and mapping of genes in our research.

Amplification of the candidate gene

Genomic DNA was extracted from fresh leaves of the two parents by the cetyl trimethyl ammonium bromide (CTAB) method. The final DNA concentration was adjusted to 50ng/μL. The full-length DNA sequence was amplified using the genomic DNA as a template; the coding sequence (CDS) was amplified using the cDNA as a template. The 20μL PCR reaction mixture contained 2μL DNA template, 10μL 2×Es TaqMasterMix, 1μL forward primer (10μM), 1μL reverse primer (10μM) and 6μL ddH_2O. The amplified reaction was as follows: 95°C for 5 min; 95°C for 30s for a total of 35 cycles; 54°C for 30s; 72°C for 75s; 72°C for 5min. The PCR products were separated on a 1% agarose gel, and purified by a TIANgel Midi Purification Kit (TIAN GEN, Beijing, China). The purified products were connected to a pMD 18-T vector (TaKaRa, Japan), and transformed into *Escherichia coli* DH5a cells. A total of 5 positive clones were sent to Invitrogen for sequencing. The specific primers for the candidate genes were designed by Primer Premier 5.0 according to the reference genome of *B. rapa*. For *Wugong-F3H* and *Wuqi-F3H*, the forward primer sequence was 5'TCCATTACATATTCTCTCTCTTTTTCTT3' and the reverse primer sequence was 5'CGACTTACACAACCGAACCAAAC3'. For upstream of *Wugong-F3H* and Wuqi-*F3H*, the forward primer F3H-Up sequence was 5'GTTGCTTTAGTTGTATGAGTTTTGTGGTAT3', the reverse primer sequence was 5'GCGGAGTGTTAGCGGGAAGAAGGAA3'. All primers were synthesized by Shanghai Sangon Biological Engineering Technology. The open reading frames (ORFs) were predicted using the ORF Finder tool online (http://www.ncbi.nlm.nih.gov/gorf/gorf.html); the CDS sequences were translated into proteins by the DNAMAN5.0 software; the theoretical isoelectric point (PI) and molecular weight (MW) of the candidates were calculated using the computer PI/Mw tool online (http://web.expasy.org/compute_pi/); the conserved domains of amino acids were analyzed by the conserved domain database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) of NCBI; multiple sequence alignments were performed using the Clustal W tool. To investigate the evolutionary relationship of the candidate genes in *Arabidopsis thaliana* (*A.thaliana*), *Brassica rapa* (*B. rapa*), *Brassica napus* (*B. napus*), *Brassica oleracea* (*B. oleracea*), *Brassica nigra* (*B. nigra*) *and Brassica juncea* (*B. juncea*), a phylogenetic tree was constructed by the maximum likelihood (ML) using the MEGA 6.0 program. The bootstrap test method was adopted and the replicate was set to 1000. A sequence of 1500bp upstream of the ATG of *Wugong-F3H* and *Wuqi-F3H* was analyzed for *cis*-acting elements using Plant CARE.

Seed Coat Color Related Indexs Analysis

Four seed coat color related indexes including flavonoids, anthocyanins, melanin and total phenol were measured. The concentrations of flavonoids, anthocyanin and total phenol were measured by the hydrochloric acid methanol method[15, 27]. The brief procedure was as follows: the fresh leaf (0.1 g) was grinded in liquid nitrogen with 5 mL acetone, and then treated in a 30 °C water bash for 1 h to extract chlorophyll and lutein. Subsequently, the mixture was centrifuged at 3500r/min for 15 min and then the supernatant was discarded. The precipitate was re-suspended in 5 mL methanol containing 5% hydrochloric acid, and incubated in a 60 °C water bash for 1 h. The supernatant was collected and the precipitate was re-extracted three times following the above procedure; the extracts were combined and diluted in methanol containing 5% hydrochloric acid. The absorbance of the mixture was determined at A_{280}nm using a solution containing methanol and 5% hydrochloric acid as a blank control. The total phenol concentration in the sample was calculated based on a standard curve showing the correlation between gallic acid concentration and absorbance. The flavonoid concentrations were determined using the absorbance at A_{325}nm and the anthocyanin concentrations were determined using the absorbance at A_{530}, A_{620} and A_{630}nm. For melanin, the measurement is basically the same as the method described above, except the last step. After incubated in a 60 °C water bash for 1 h, the precipitate was immersed in 2 mL 2% NaOH, and then incubated in a 70 °C water bash until the color of the material completely faded. The absorbance of the mixture was determined at A_{290}nm.

Expression Pattern Analysis Of Related Genes

Quantitative real-time PCR (qRT-PCR) was used to analyze the expression of the genes in different tissues from the two parents. Primers for qRT-PCR were designed from conserved nucleotide regions identified by multiple alignments of sequences from the *Arabidopsis* information resource (TAIR: http://www.arabidopsis.org/). Twelve pairs of primers from twelve genes were designed for gene expression using Primer 5.0 (Supplemental Table 4), and synthesized by Sangon Biotech (Shanghai, China). The TIPS-41 gene was used as an internal control[39].

The qRT-PCR experiments were performed according to the instructions of the SYBR® Premix Ex Taq™ II (TliRNaseH Plus) (TaKaRa, DaLian) using a reaction system of 25μL containing 10μL SYBR Premix Ex Taq II (TliRNaseH Plus) (2×), 0.8μL PCR forward primer (10 μM), 0.8μL PCR reverse primer (10 μM), 0.4μL ROX Reference Dye II (50×), 2μL cDNA (2 mg/μL), and 6μL ddH_2O. The amplification was performed by an initial incubation at 95 °C for 30 s, then 95 °C for 5 s for a total of 40 cycles, followed by 60 °C for 30 s. The expression level of every gene was calculated by the $2^{-\triangle\triangle Ct}$ comparative threshold cycle (Ct) method. The Ct values were generated using an Applied Biosystems 7500 Fast Real-Time PCR System (Life Technologies). All samples were repeated for three times.

Correlation between the expression of key genes and seed coat color related indexes

SPSS 17.0 was used to analyze the correlation between the expression of key genes in the flavonoid biosynthesis pathway and the content of seed coat color related indexes in different seed development stages (9, 16, 23, 30, 38, and 45 days after pollination). Pearson coefficient was adopted in the correlation analysis.

Analysis of candidate genes in Brassica juncea **(**B. juncea**)**

In a previous study, the yellow seed coat color gene has been mapped to chromosome 5 in *Arabidopsis*[3]. Thus, the sequence of chromosome 5 in *Arabidopsis* was randomly selected for designing intron polymorphism (IP) primers, especially the sequences around *TT* genes. The detailed method can be referred to Huang et al[38]. These pairs of primers were used to amplify a small population including 10 yellow and 10 brown individuals derived from the BC_9 population. The primers that showed polymorphism were used to screen the recombinants of BC_9. Additionally, through the correlation comparison of gene expression and the content of seed coat color related indexes, the genes with better correlation were selected. The candidate genes were determined based on the analysis of the expression and mapping of genes in our research.

Abbreviations

PAs: proanthocyanidins;

PAL: phenylalanine ammonia lyase;

C4H: cinnamic acid 4-hydroxylase;

CHS: chalcone synthase;

CHI: chalcone isomerase;

DFR: dihydroflavonol-4-reductase;

F3H: flavonol 3-hydroxylase;

LDOX: leucoanthocyanidin dioxygenase;

TT: transparent testa;

TTG: Transparent testa glabra;

BAN: banyuls;

ANR: anthocyanidinreductase;

FLS: flavonol synthase;

qRT-PCR: quantitative real-time PCR;

CDS: conserved domains;

DAP: days after pollination;

Bj: Brassica juncea;

ORF: open reading frame;

IP: intron polymorphism;

DEGs: different expressed genes

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

All data generated during this study are included in this published article and its additional files.

Competing interests

The authors have no conflict of interest to declare

Funding

The research was supported by the National key research and development program (2018YFD0100603), Zhongying Youth Scholar Program of Northwest A&F University. There is no role of the funding body in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

ZH and AXX were responsible for designing this study and drafting the manuscript; YW carried out gene expression and gene cloning. HL carried out seed coat color related index analysis; LL and XL collected important background information and provided assistance for data acquisition, data analysis and statistical analysis. All authors have read and approved the content of the manuscript.

Acknowledgements

Not Applicable.

References

1. National Bureau of Statistics. 2018.
2. Anderson DM, MacPherson MJ, Collins SA, MacIsaac PF. Yellow-and brown-seeded canola (Brassica napus), camelina (Camelina sativa) and Ethiopian mustard (Brassica carinata) in practical diets for rainbow trout fingerlings. Journal of applied aquaculture. 2018;30:187–95.
3. Huang Z, Ban Y, Yang L, Zhang Y, Li H, Xiao E, et al. Fine mapping of the yellow seed locus in Brassica juncea L. Genome. 2011;55:8–14.
4. Akhov L, Ashe P, Tan Y, Datla R, Selvaraj G. Proanthocyanidin biosynthesis in the seed coat of yellow-seeded, canola quality Brassica napus YN01-429 is constrained at the committed step catalyzed by dihydroflavonol 4-reductase. Botany. 2009;87:616–25.
5. Marles MAS, Gruber MY. Histochemical characterisation of unextractable seed coat pigments and quantification of extractable lignin in the Brassicaceae. Journal of the Science of Food and Agriculture. 2004;84:251–62.
6. Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L. The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in Arabidopsis siliques. The Plant Cell. 2000;12:1863–78.
7. Yang S-C, Arasu MV, Chun J-H, Jang Y-S, Lee Y-H, Kim IH, et al. Identification and determination of phenolic compounds in rapeseed meals (Brassica napus L.). Journal of Agricultural Chemistry and Environment. 2015;4:14.
8. Huang Y, Gou J, Jia Z, Yang L, Sun Y, Xiao X, et al. Molecular cloning and characterization of two genes encoding dihydroflavonol-4-reductase from Populus trichocarpa. PloS one. 2012;7:e30364.
9. Auger B, Marnet N, Gautier V, Maia-Grondard A, Leprince F, Renard M, et al. A detailed survey of seed coat flavonoids in developing seeds of Brassica napus L. Journal of Agricultural and Food Chemistry. 2010;58:6246–56.
10. Winkel-Shirley B. Biosynthesis of flavonoids and effects of stress. Current opinion in plant biology. 2002;5:218–23.
11. Lepiniec L, Debeaujon I, Routaboul J-M, Baudry A, Pourcel L, Nesi N, et al. Genetics and biochemistry of seed flavonoids. Annu Rev Plant Biol. 2006;57:405–30.
12. Bradley JM, Davies KM, Deroles SC, Bloor SJ, Lewis DH. The maize Lc regulatory gene up-regulates the flavonoid biosynthetic pathway of Petunia. The Plant Journal. 1998;13:381–92.
13. Park J-S, Kim J-B, Kim K-H, Ha S-H, Han B-S, Kim Y-H. Flavonoid biosynthesis: Biochemistry and metabolic engineering. Journal of Plant Biotechnology. 2002;29:265–75.
14. Saito K, Yonekura-Sakakibara K, Nakabayashi R, Higashi Y, Yamazaki M, Tohge T, et al. The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. Plant Physiology and Biochemistry. 2013;72:21–34.

15. Ye X, Li J, Tang Z, Liang Y, Chen L. Study on seedcoat color and related characters of Brassica napus. Zuo wu xue bao. 2001;27:550–6.
16. Arsovski AA, Haughn GW, Western TL. Seed coat mucilage cells of Arabidopsis thaliana as a model for plant cell wall research. Plant signaling & behavior. 2010;5:796–801.
17. Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L. TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana. The Plant Journal. 2004;39:366–80.
18. Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, et al. Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. Proceedings of the National Academy of Sciences. 2010;107:8063–70.
19. Lagercrantz U, Lydiate DJ. Comparative genome mapping in Brassica. Genetics. 1996;144:1903–10.
20. Love C, Logan E, Erwin T, Spangenberg G, Edwards D. Analysis of the Brassica'A'and'C'genomes and comparison with the genome of Arabidopsis thaliana. In: Proceedings of the Joint Meeting of the XIVth Crucifer Genetics Workshop and IVth ISHS Symposium on Brassicas. International Society for Horticultural Science; 2006. p. 99–104.
21. Li X, Chen L, Hong M, Zhang Y, Zu F, Wen J, et al. A large insertion in bHLH transcription factor BrTT8 resulting in yellow seed coat in Brassica rapa. PLoS One. 2012;7:e44145.
22. Zhang J, Lu Y, Yuan Y, Zhang X, Geng J, Chen Y, et al. Map-based cloning and characterization of a gene controlling hairiness and seed coat color traits in Brassica rapa. Plant molecular biology. 2009;69:553–63.
23. Huang Z, Peng G, Liu X, Deora A, Falk KC, Gossen BD, et al. Fine mapping of a clubroot resistance gene in Chinese cabbage using SNP markers identified from bulked segregant RNA sequencing. Frontiers in plant science. 2017;8:1448.
24. Song J, Li Z, Liu Z, Guo Y, Qiu L-J. Next-generation sequencing from bulked-segregant analysis accelerates the simultaneous identification of two qualitative genes in soybean. Frontiers in plant science. 2017;8:919.
25. Wang Y, Xie J, Zhang H, Guo B, Ning S, Chen Y, et al. Mapping stripe rust resistance gene YrZH22 in Chinese wheat cultivar Zhoumai 22 by bulked segregant RNA-Seq (BSR-Seq) and comparative genomics analyses. Theoretical and Applied Genetics. 2017;130:2191–201.
26. Quarrie SA, Lazić-Jančić V, Kovačević D, Steed A, Pekić S. Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. Journal of experimental botany. 1999;50:1299–306.
27. Ye X, Li X, Li J. Mechanism of melanin synthesis in seed coat of Brassica napus L. Zuo wu xue bao. 2002;28:638–43.
28. YE X, LI J, TANG Z, CHEN L. Difference of seedcoat color between black-and yellow-seeded in B. napus with seed development □. Changes of melanin, tyrosine and tyrosinase and their correlation analyses [J]. Chinese Journal of Oil Crop Scieves. 2001;3.

29. YE X, LI J, TANG Z, CHEN L. Difference of seedcoat color between black-and yellow-seeded in B. napus with seed development . Changes of anthocyanin, phenylalanine and phenylalaine ammonia-lyase and their correlation analyses [J]. Chinese Journal of Oil Crop Scieves. 2001;2.
30. Dixon RA, Xie D, Sharma SB. Proanthocyanidins–a final frontier in flavonoid research? New phytologist. 2005;165:9–28.
31. Pelletier MK, Burbulis IE, Winkel-Shirley B. Disruption of specific flavonoid genes enhances the accumulation of flavonoid enzymes and end-products in Arabidopsis seedlings. Plant molecular biology. 1999;40:45–54.
32. Han Y, Huang K, Liu Y, Jiao T, Ma G, Qian Y, et al. Functional analysis of two flavanone-3-hydroxylase genes from Camellia sinensis: a critical role in flavonoid accumulation. Genes. 2017;8:300.
33. Zuker A, Tzfira T, Ben-Meir H, Ovadis M, Shklarman E, Itzhaki H, et al. Modification of flower color and fragrance by antisense suppression of the flavanone 3-hydroxylase gene. Molecular Breeding. 2002;9:33–41.
34. Jeong ST, Goto-Yamamoto N, Kobayashi S, Esaka M. Effects of plant hormones and shading on the accumulation of anthocyanins and the expression of anthocyanin biosynthetic genes in grape berry skins. Plant Science. 2004;167:247–52.
35. Li C, Liu S, Yao X, Wang J, Wang T, Zhang Z, et al. PnF3H, a flavanone 3-hydroxylase from the Antarctic moss Pohlia nutans, confers tolerance to salt stress and ABA treatment in transgenic Arabidopsis. Plant growth regulation. 2017;83:489–500.
36. Song X, Diao J, Ji J, Wang G, Guan C, Jin C, et al. Molecular cloning and identification of a flavanone 3-hydroxylase gene from Lycium chinense, and its overexpression enhances drought stress in tobacco. Plant physiology and biochemistry. 2016;98:89–100.
37. Warwick SI, Black LD. Molecular systematics of Brassica and allied genera (subtribe Brassicinae, Brassiceae)—chloroplast genome and cytodeme congruence. Theoretical and Applied Genetics. 1991;82:81–92.
38. Huang Z, Liu L, Lu H, Lang L, Zhao N, Ding J, et al. Development of IP and SCAR markers linked to the yellow seed color gene in Brassica juncea L. Breeding science. 2016;66:175–80.
39. Chandna R, Augustine R, Bisht NC. Evaluation of candidate reference genes for gene expression normalization in Brassica juncea using real time quantitative RT-PCR. PloS one. 2012;7:e36918.

Tables

Table 1 Identification of KEGG involved in flavonoid biosynthesis pathway of DEGs.

gene	log2(FC)	*B.rapa* Gene ID	Chr
BAN	14.27	Bra021318	A01
BAN	12.33	Bra031403	A01
DFR	10.09	Bra027457	A09
LDOX	8.18	Bra013652	A01
LDOX	6.84	Bra019350	A03
TT5	5.3	Bra007145	A09
C4H	3.89	Bra021637	A04
TT8	3.63	Bra037887	A09
TT4	3.52	Bra023441	A02
PAL1	3.05	Bra005221	A05
TT4	3.04	Bra008792	A10
TT4	2.92	Bra006224	A03
F3H	2.61	Bra012862	A03
F3H	2.08	Bra036828	A09
PAL1	1.99	Bra017210	A04
TT4	1.74	Bra036307	A09
C4H	1.22	Bra018311	A05

Table 2 Analysis of seed coat color related indexes during the development stages of seeds.

	Flavonoids(Unit/g DW)			Anthocyanins(Unit/g DW)			Melanin(OD/g DW)			Total phenolics(mg GAE/g DW)		
	Wugong	Wuqi	Difference	Wugong	Wuqi	Difference	Wugong	Wuqi	Difference	Wugong	Wuqi	Difference
9DAP	121.47 E	127.85 E	-6.38	238.20 F	69.17 E	169.03 **	26.68 F	17.60 F	9.07 **	230.20 F	120.60 E	109.60 **
16DAP	109.51 E	126.12 E	-16.61	726.75 E	75.33 D	651.42**	59.89 E	48.22 E	11.66 **	319.30 E	180.90 W	138.40 **
23DAP	523.20 D	337.39 D	185.81 **	2217.00 B	182.80 C	2034.20**	109.66 D	87.50 D	22.15 **	793.30 D	255.60 D	537.70 **
30DAP	887.38 C	770.32 C	117.07 **	2337.33 A	247.83 B	2089.50 **	224.68 C	136.51 C	88.16 **	1070.40 C	544.00 C	526.40 **
38DAP	1926.05 B	1400.00 B	526.05 **	1256.34 C	305.43 A	950.91 **	397.66 B	195.95 B	201.71 **	1157.40 B	655.50 B	501.90 **
45DAP	2035.07 A	1778.57 A	256.50 **	1195.00 D	300.08 A	894.92 **	698.76 A	203.55 A	495.21 **	1229.70 A	868.00 A	601.70 **

** indicates significance at the level of 1%.

Table 3 Correlation between the key genes expression and seed coat color related indexes.

		BjTT4	*BjTT5*	*BjF3H*	*BjDFR*	*BjTT8*
Wugong mustard	Flavonoids	-0.623	0.726	-0.66	-0.882	-0.64
	Anthocyanins	-0.569	0.817*	-0.806	-0.507	-0.608
	Melanin	-0.522	0.624	-0.583	-0.776	-0.531
	Total phenolics	-0.757	0.939**	-0.874*	-0.951**	-0.784
Wuqi mustard	Flavonoids	-0.326	0.595	-0.900*	-0.33	-0.322
	Anthocyanins	-0.098	0.848*	-0.827*	0.273	-0.415
	Melanin	-0.209	0.776	-0.836*	0.14	-0.414
	Total phenolics	-0.324	0.811*	-0.878*	0.264	-0.345

Table 4 Analysis of cis-acting elements in the upstream of *Wuqi-F3H* and *Wugong-F3H*.

Cis-acting element	Core sequence	Function
AT-rich element	ATAGAAATCAA	binding site of AT-rich DNA binding protein (ATBP-1)
5UTR Py-rich stretch	TTTCTTCTCT	cis-acting element conferring high transcription levels
AC-I	CCCACCTACC	element conferring enhanced xylem expression
MBS	TAACTG	MYB binding site involved in drought-inducibility
ACE	ACTACGTTGG/ACGTGGA	cis-acting element involved in light responsiveness
ATCT-motif	AATCTGATCG	part of a conserved DNA module involved in light responsiveness
AT1-motif	ATTAATTTTACA	part of a light responsive module
TC-rich repeats	GTTTTCTTAC	cis-acting element involved in defense and stress responsiveness
GCN4_motif	TGAGTCA	cis-regulatory element involved in endosperm expression
Skn-1_motif	GTCAT	cis-acting regulatory element required for endosperm expression
TGA*	AACGAC	auxin-responsive element
ABRE	CACGTG/ACGTGGC	cis-acting element involved in the abscisic acid responsiveness
CE1	TGCCACCGG	cis-acting element associated to ABRE, involved in ABA responsiveness
P-box	CCTTTTG	gibberellin-responsive element
CGTCA-motif	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness
TGACG-motif	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness
O_2-site	GATGATGTGG/GATGACATGG	cis-acting regulatory element involved in zein metabolism regulation

Figures

Figure 1

Distribution of transcriptional sequence of yellow seed gene in B. juncea SNP variation.

Figure 2

Expression analysis of five flavonoid pathway DEGs at different seed development stages in parents.

Figure 3

Sequence comparison of BjF3H between Arabidopsis and Brassica species. The red rectangle indicates the DIOX_N super family domain, the orange rectangle indicates the 2OG-FeII_Oxy super family domain.

Figure 4

Phylogenetic tree analysis of the candidate genes. Numbers are bootstrap values indicating frequencies of respective furcations found in 1,000 replications of subset tree calculations.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SITables.docx
- SupplementalFigure1.pdf
- SupplementalFigure2.pdf