Fine-Mapping And Identification of A Candidate Gene Controlling Seed Coat Color In Melon (*Cucumis Melo* L. Var. *Chinensis* Pangalo)

Zhicheng Hu

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Xueyin Shi

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Xuemiao Chen

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Jing Zheng

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Aiai Zhang

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Qiushi Fu (✉ fuqiushi@caas.cn)

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers <https://orcid.org/0000-0003-2377-7034>

Huaisong Wang

Chinese Academy of Agricultural Sciences Institute of Vegetables and Flowers

Research Article

Keywords: melon, seed coat color, fine-mapping, phenotypic analysis

Posted Date: July 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-720440/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Theoretical and Applied Genetics on November 26th, 2021. See the published version at <https://doi.org/10.1007/s00122-021-03999-5>.

Abstract

Seed coat color is related to flavonoid content which is closely related to seed dormancy. According to the genetic analysis of a six-generation population derived from two parents (IC2508 with a yellow seed coat and IC2518 with a brown seed coat), we discovered that the yellow seed coat trait in melon was controlled by a single dominant gene, named *CmBS-1*. Bulk segregant analysis sequencing (BSA-Seq) revealed that the gene was located at 11,860,000–15,890,000 bp (4.03 Mb) on Chr 6. The F₂ population was genotyped using insertion-deletions (InDels), from which cleaved amplified polymorphic sequence (dCAPS) markers were derived to construct a genetic map. The gene was then fine-mapped to a 233.98 kb region containing 12 genes. Based on gene sequence analysis with two parents, we found that the *MELO3C019554* gene encoding a homeobox protein (PHD transcription factor) had a nonsynonymous single nucleotide polymorphism (SNP) mutation in the coding sequence (CDS), and the SNP mutation resulted in the conversion of an amino acid (A→T) at residue 534. In addition, *MELO3C019554* exhibited lower relative expression levels in the yellow seed coat than in the brown seed coat. Furthermore, we found that *MELO3C019554* was related to 12 flavonoid metabolites. Thus, we predicted that *MELO3C019554* is a candidate gene controlling seed coat color in melon. The study lays a foundation for further cloning projects and functional analysis of this gene, as well as marker-assisted selection breeding.

Key Message

***MELO3C019554* encoding a homeobox protein (PHD transcription factor) is a candidate gene that involved in the formation of seed coat color in melon.**

Introduction

Seed coat color is an essential agronomic trait in melon. An earlier study revealed that seed coat color is related to antioxidant quantities and activities, as well as the biochemical characteristics of the seed (Malencic et al. 2012). Melons possess diverse seed coat colors. While yellow and brown seed coats are common in melon, white, yellowish-white, dark brown, reddish-brown, and other seed coat colors are also found. Flavonoids related to seed coat color formation, as well as the associated molecular mechanisms thereof, have been identified in *Arabidopsis* and the Brassicaceae family (Debeaujon et al. 2000; Li et al. 2010; Padmaja et al. 2014). Proanthocyanidin (PA), one type of flavonoid, was confirmed to be associated with seed coat color formation in earlier research (Dixon et al. 2005; Lepiniec et al. 2006). PAs are end products in the flavonoid biosynthesis pathway and have protective effects against pathogens (Winkel-Shirley 2001; Dixon et al. 2005).

There are far more studies on seed coat color in *Arabidopsis*, the Brassicaceae family, soybean, and other species than in melon. Research has reported that different genes control the formation of seed coat color in different species. In *Arabidopsis*, several genes related to the flavonoid biosynthesis pathway were found to be associated with seed coat color, including *chalcone synthase*, *chalcone isomerase*,

dihydroflavonol reductase (DFR), and other genes (Yu 2013). In soybean, *GmF3'H*, which encodes flavonoid 3' hydroxylase, is involved in the flavonoid pathway and results in pigmentation of the seed coat (Zabala and Vodkin, 2003). In addition, seed coat color is also regulated by some transcription factors. *BrTT8* encodes a basic/helix-loop-helix (bHLH) transcription factor related to the accumulation of PAs, and a large insertion in *BrTT8* is responsible for the yellow seed coat in *Brassica rapa* (Li et al. 2012). Ren et al. (2017) and Wang et al. (2016) found that *TTG1* was the most likely candidate gene regulating seed coat color in *Brassica rapa*. In *Arabidopsis*, genes encoding PAP1, GL3, FUSCA3, ANL2, KAN4, and other transcription factors were identified that control seed coat color formation (Qu et al. 2013; Padmaja et al. 2014). Other genes related to seed coat color formation have been discovered. In *Brassica juncea*, Huang et al. (2016) confirmed that *Bra036828* is likely linked to yellow seed coat formation. The gene encoding polyphenol oxidase (PPO) involved in black pigment synthesis was found to be a candidate gene of seed coat color in sesame (Wei et al. 2016). *Cla019481*, which encodes PPO, is the candidate gene for the formation of a black seed coat in watermelon (Li et al. 2019).

Melon (*Cucumis melo* L.) is an essential vegetable crop in the Cucurbitaceae family. With the rapid development of sequencing technology and molecular biology, gene mapping processes are becoming increasingly accurate. Following the release of the melon genome (Garcia-Mas et al. 2012; Valentino et al. 2018), the associated sequence information has been employed in marker development, fine-mapping, and gene cloning. Some markers tightly linked to peel spot (Lv et al. 2018), chilling tolerance (Hou et al. 2018), gummy stem blight (GSB) resistance (Hassan et al. 2018), male sterility mutant (Sheng et al. 2017), sweet and sour (Zhang et al. 2016), and other agronomic traits have been developed in melon. Genes controlling cucumber mosaic virus resistance (Giner et al. 2017), GSB resistance gene (Hassan et al. 2018), fruit shape (Gur et al. 2017), fruit firmness (Nimmakayala et al. 2016), and other important traits have been mapped and cloned in melon.

In this study, we used six-generation populations derived from IC2508 (yellow seed coat) and IC2518 (brown seed coat) to investigate the inheritance of melon seed coat color. We used bulked segregant analysis sequencing (BSA-Seq) and constructed a genetic map with InDels and dCAPS markers from the F₂ population to map the candidate gene. Based on gene function annotation analysis, sequence alignment, and quantitative real-time (qRT)-PCR, we predicted a candidate gene controlling melon seed coat color. The study contributes to the cloning of the *CmBS-1* gene and furthers our understanding of melon seed coat color formation.

Materials And Methods

Plant materials

The inbred line IC2508 with a yellow seed coat was the female parent, and the inbred line IC2518 with a brown seed coat was the male parent (Fig. 1). The seed coats of IC2508 and IC2518 at 30 days after pollination were used to measure the flavonoid content and were subjected to flavonoid metabolomic analysis. The F₁ generation was generated by crossing IC2508 and IC2518, and was self-pollinated to

produce the F_2 population. The F_1 (as the female) was backcrossed with IC2508 and IC2518 to obtain BC_1P_1 and BC_1P_2 . To validate the genetic inheritance and map the gene, in the 2018 spring season, we planted 30 P_1 , 30 P_2 , 50 F_1 , 258 F_2 , 98 BC_1P_1 , and 95 BC_1P_2 individuals. In the 2019 spring season, we screened the recombinant plants from a larger F_2 population (2262 individuals) to fine-map the target gene. Five inbred lines with a yellow seed coat color and five inbred lines with a brown seed coat color were planted to measure the relative expression of the target gene using qRT-PCR. The mature seeds of the 10 melon accessions are shown in Fig S1.

Phenotypic analysis

Self-pollinated seeds of the six-generation family were collected in the mature period, and the seed coat color trait was analyzed for genetic inheritance. The coat color of mature melon seeds was examined visually. In addition, a colorimeter (Konica Minolta CM-3700D, Osaka, Japan) was used to measurement of F_2 seed coat color. The color parameters including a^* , a greenness to redness parameter, b^* , a blueness to yellowness parameter, L^* , a darkness to lightness parameter, $E\text{-value}=(a^{*2}+b^{*2}+L^{*2})^{1/2}$. SAS version 9.4 (SAS Institute, Cary, NC, USA) was used to conduct Chi-square tests on the F_2 (252 individuals) and BC_1P_2 (94 individuals) populations.

Flavonoid content measurement

We collected the seed coats of IC2508 and IC2518 at 30 days after pollination. The flavonoids were extracted and measured as described by Li et al. (2011), with three biological and technical replicates measured per sample.

Extraction and analysis of flavonoid metabolites

Metabolites were extracted from the melon seed coats of IC2508 and IC2518 at 30 days after pollination, with three replicates per sample. An ultra-performance liquid chromatography-electrospray ionization-tandem mass spectrometry (UPLC-ESI-MS/MS) system (UPLC, SHIMADZU Nexera X2, www.shimadzu.com.cn/; MS, Applied Biosystems 4500 Q TRAP, www.appliedbiosystems.com.cn/) was used to analyze the extracts. We used a combination of the variable importance in the projection (VIP) score extracted from the orthologous partial least squares-discriminant analysis (OPLS-DA) and absolute Log2FC (fold change) to determine the significant differential metabolites, with a screening criteria of $VIP \geq 1$ and $\text{Log2FC} \geq 1$. We annotated identified metabolites using the Kyoto Encyclopedia of Genes and Genomes (KEGG) compound database (<http://www.kegg.jp/kegg/compound/>) and then mapped the annotated metabolites to the KEGG pathway database (<http://www.kegg.jp/kegg/pathway.html>). Additionally, we fed the pathway into MSEA (metabolite sets enrichment analysis), with significance determined by the P -values of the hypergeometric test.

DNA extraction and PCR conditions

A modified cetyltrimethyl-ammonium bromide (CTAB) method (Murray and Thompson 1980) was applied to extract total genomic DNA from the young leaves of melon. The DNA concentration was quantified using a NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA) and was diluted to a suitable concentration. Electrophoresis in 1.0% agarose gel was used to evaluate the DNA.

The 10- μ L PCR reaction mixture contained 1 μ L DNA, 0.5 μ L forward primer (10 μ M), 0.5 μ L reverse primer (10 μ M), 3 μ L ddH₂O, and 5 μ L PCR mix and was performed under the following conditions: an initial 94°C for 5 min; followed by 35 cycles at 94°C for 15 s, 54°C for 15 s, and 72°C for 30 s; with a final extension at 72°C for 6 min; followed by 4°C. The dCPAS marker repeated the above steps, and the corresponding restriction endonucleases were added to the PCR-amplified products to digest the PCR products following the manufacturer's instructions. Amplicons were separated by 8% polyacrylamide gel electrophoresis (PAGE) at 150 V for 2–2.5 h and then silver-stained.

BSA-Seq and resequencing analysis

The two bulks were constructed with 30 yellow seed coat F₂ individuals and 30 brown seed coat F₂ individuals. The DNAs of IC2508, IC2518, the yellow bulk, and the brown bulk were extracted for resequencing on an Illumina HiSeq 2000 platform. The melon genome version 3.5.1 was used as the reference genome (https://melonomics.net/files/Genome/Melon_genome_v3.5.1/). Clean reads were obtained by filtering the raw reads and were aligned to the melon reference genome using BWA software (Li and Durbin 2009; Table S1). The SNPs and InDels were detected by GATK and Picard (McKenna et al. 2010). SnpEff software (Cingolani et al. 2012) was used to annotate the localization of the SNPs and InDels. Euclidean distance (ED) (median + 3SD as the threshold), Δ SNP index (99% quantile as the threshold), and Δ InDel index (95% quantile as the threshold) values were calculated to obtain a preliminary candidate interval associated with the seed coat color trait (Abe et al. 2012; Fekih et al. 2013; Hill et al. 2013; Takagi et al. 2013; Lian et al., 2019).

Primer design and genotyping

During preliminary mapping, we used the InDels (> 5 bp difference) and SNPs to design primers with Primer 3.0 software and dCAPS Finder 2.0 (<http://helix.wustl.edu/dcaps/dcaps.html>), respectively. The primers were used to screen the DNA of the parents and the two bulks to obtain polymorphic primers. The polymorphic primers were then used to identify the genotypes of the brown seed coat color individuals from the F₂ population. For each polymorphic primer, individuals with the IC2508 allele were marked as 'a', individuals with the IC2518 allele were marked as 'b', individuals with the F₁ allele were marked as 'h', and the unknown bands were marked as 'u'. The QTL IciMapping 4.1 software was used to construct the genetic map.

Functional annotation and sequence analysis of the candidate gene

The functional annotation and sequence analysis of the genes in the mapping region were based on the melon (DHL92) genome (https://melonomics.net/files/Genome/Melon_genome_v3.5.1/). Amino acid

sequences were predicted using the ExPASy-Translate tool (<https://web.expasy.org/translate/>). The gene and amino acid sequences were aligned using MEGA 7 software and BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

Extraction of RNA and qRT-PCR

The specific primers of the candidate gene and *Actin* for qRT-PCR were designed using Primer 5.0 software based on the melon genomic database (Table S2). Total RNA was extracted from the melon seed coat tissue using an RNAprep Pure Plant Kit (TianGen, China). We used a PrimeScript RT Reagent Kit (TaKaRa, Japan) to prepare first-stand cDNA. The SYBR Primer Ex Taq Kit (TaKaRa, Japan) was used to conduct the qRT-PCR. The amount of DNA was monitored using the StepOne Plus Real-Time Fluorescent Quantitative PCR system (ABI, Foster City, CA, USA). Three biological replicates and five technical replicates of each sample were used for qRT-PCR. The $2^{-\Delta\Delta Ct}$ method was used to evaluate the quantification.

Phylogenetic analysis

According to the homologous gene sequence obtained by a BLAST search in the NCBI database (<https://www.ncbi.nlm.nih.gov/>), we constructed the phylogenetic tree in MEGA 7 software using the bootstrap method with 1000 replications (Kumar et al. 2016).

Results

Determination of flavonoid content

The flavonoid contents of the mature seed coat samples of the two parents were measured. The result showed that the flavonoid content of the yellow seed coat was $3.21 \text{ mg}\cdot\text{g}^{-1}$, while that of the brown seed coat was $8.09 \text{ mg}\cdot\text{g}^{-1}$ (Fig. 2). The flavonoid content of the brown seed coat was 1.5-fold higher than that of the yellow seed coat.

The flavonoid metabolite profiling analysis of the two parental lines indicated similar results. A total of 19 significantly differentially abundant metabolites, including anthocyanins, flavonols, flavonoids, chalcones, and flavonoid carbonosides, were detected between the IC2508 and IC2518 mature seed coat samples, with most metabolites upregulated (14 upregulated and 5 downregulated) in IC2518. The levels of kaempferol-3-*O*-sambubioside, apigenin-8-*C*-glucoside, diosmetin, isovitexin, chrysoeriol-5-*O*-glucoside, and chrysoeriol-7-*O*-(6"-acetyl)-glucoside were increased by at least three-fold (Table S3) in IC2518. We mapped the differentially abundant metabolites to KEGG biological pathways to identify the main pathways that the metabolites are involved in. There were three, one, and one significantly differentially abundant metabolites that were assigned to flavone and flavonol biosynthesis (ko00944), anthocyanin biosynthesis (ko00942), and flavonoid biosynthesis (ko00941), respectively, among others (Table S3). The results suggested that the difference in seed coat flavonoid content was responsible for the difference in the seed coat color.

Genetic analysis of melon seed coat color

The seed coat color of all 50 individuals in the F_1 ($P_1 \times P_2$) population was yellow, indicating that the yellow seed coat trait is dominant over the brown seed coat trait. In the F_2 population (252 individuals), 186 individuals exhibited a yellow seed coat and 66 individuals produced a brown seed coat. The Chi-square test suggested a 3:1 segregation ratio (Table 1). The analysis of E-value showed that the seed color of F_2 population conformed to bimodal distribution (Fig. S1). In the BC_1P_2 ($F_1 \times P_2$) population (96 individuals), 58 individuals showed a yellow seed coat and 38 individuals showed a brown seed coat. The Chi-square test suggested that the segregation ratio was 1:1. The seed coat color of all 95 BC_1P_1 ($F_1 \times P_1$) individuals was yellow (Table 1). The results confirmed that the yellow seed coat trait was controlled by a single dominant gene, which was named *CmBS-1*.

Table 1 Genetic analysis of seed coat color trait in six-generation family

Generation	Total no. of individuals	No. of yellow individuals	No. of brown individuals	Expected ratio	χ^2 ($\alpha=0.05$)	<i>P</i> -value ($\alpha=0.05$)
P_1	50	50	0	-	-	-
P_2	50	0	50	-	-	-
F_1	50	50	0	-	-	-
F_2	252	186	66	3:1	0.1905	0.6625
BC_1P_1	97	97	0	-	-	-
BC_2P_2	94	56	38	1:1	3.4468	0.0634

Preliminary mapping of *CmBS-1* by BSA-Seq

We used BSA-Seq to rapidly identify the genomic region of *CmBS-1*. A total of 60.06 Gb clean reads were generated by re-sequencing, with an average sequencing depth of 33.37×. The clean reads were aligned to the melon reference genome, and the coverage of the reference genome was 98.02%. A total of 123,456 SNPs and 49,004 small InDels were identified between the two parents, while 30,950 SNPs and 13,894 small InDels were identified between the two DNA bulks. The details of the SNPs and InDels between the two bulks as well as the parents are shown in Tables S4 and S5. We used the ED method (the threshold was 0.29) and Δ SNP index (the confidence value was 99%) to analyze the SNPs and found that *CmBS-1* was mapped to a 4.03 Mb region from 11,860,000 to 15,890,000 bp on Chr 6 (Fig. 3). Similarly, we used the ED method (the threshold was 0.30) and Δ InDel index (the confidence value was 95%) to analyze the small InDels, and the result suggested that *CmBS-1* was mapped to an interval of 11,770,000–16,270,000 bp (4.50 Mb) on Chr 6 (Fig. 4). The two regions overlapped. Therefore, we

confirmed that *CmBS-1* was located at the merged region from 11,860,000–15,890,000 bp on Chr 6 at a physical distance of 4.03 Mb.

Fine mapping of *CmBS-1* using InDel and dCAPS markers

In the preliminary mapping region, based on the re-sequencing data, we designed 1498 InDel primers and randomly chose 240 InDel primers to screen in the two parents and the two bulks to obtain polymorphic primers. Fifty-two of the 240 primers were polymorphic, and the polymorphism rate was 21.7%. We then chose eight co-dominant polymorphic markers that had clear bands and demonstrated stable amplification to map *CmBS-1* in a smaller F₂ population (252 individuals) with 12 recombinant individuals (Tables S6 and S7). The result suggested that the *CmBS-1* gene was mapped in a 970.78 kb interval between M255 and L274 (11,109,174–12,079,958 bp).

In the interval of M255 and L274, we designed 35 InDel markers and two dCAPS markers based on the re-sequencing data, and eight of them were polymorphic markers. We then used the eight co-dominant polymorphic markers to fine-map the *CmBS-1* in a larger F₂ population (2270 individuals) with 10 recombinant individuals (Tables S6 and S7). The result indicated that SNP2 and ZH39 were the closest markers to *CmBS-1*. The physical distance between the two markers was 233.98 kb (11,763,814–11,997,798 bp). We used QTL IciMapping 4.1 software to construct the genetic map with 16 markers (Fig. 5a).

Prediction and sequence analysis of the candidate gene

Based on the melon genome database, 12 genes were identified in the 233.98 kb region. The position and functional annotation of these genes are shown in Table 2. We analyzed the sequences of the 12 genes in the candidate region based on the parental re-sequencing data. The gene sequence alignment of the 12 genes between the yellow seed coat and brown seed coat parental lines revealed that one nonsynonymous SNP and two InDel mutations were present in *MELO3C019554*. We found that both of the two InDels were in the intron region at 11,965,686 bp (TA→T) and 11,966,210 bp (G→GA) on Chr 6, respectively (Table S8). The one nonsynonymous SNP was in the fifth exon region at 11,967,142 bp (C→T) on Chr 6, resulting in the conversion of an alanine (A) to threonine (T) at residue 534 (Fig. 5 b–d; Table S8). However, we found that there were no mutations in exon of the other 11 genes between the two parental lines. The promoter sequence analysis of the 12 genes between the two parental lines showed that there were no mutations in any of the 12 genes. According to the melon genome database, *MELO3C019554* encodes a homeobox protein that is a PHD transcription factor. Therefore, we predicted that *MELO3C019554* was the candidate gene for the yellow seed coat and brown seed coat in melon.

Table 2 Annotation of genes in the candidate region

Gene ID	Position	Functional annotation
<i>MELO3C019545</i>	11771441..11774714 (-strand)	Reticulon-like protein
<i>MELO3C019546</i>	11775819..11778327 (-strand)	PRKR-interacting 1
<i>MELO3C019547</i>	11783292..11783498 (-strand)	MGDG synthase type A family protein
<i>MELO3C019548</i>	11851782..11853751 (-strand)	Alcohol dehydrogenase, putative
<i>MELO3C019549</i>	11868402..11868584 (-strand)	Galactose oxidase/kelch repeat superfamily protein
<i>MELO3C019550</i>	11892173..11895412 (+strand)	Methylthioribose kinase 1
<i>MELO3C019551</i>	11897392..11897999 (-strand)	FAD-binding Berberine family protein
<i>MELO3C019552</i>	11900852..11902620 (+strand)	Coffea canephora DH200=94 genomic scaffold, scaffold_1
<i>MELO3C019553</i>	11924765..11929982 (+strand)	Alpha-L-fucosidase 2
<i>MELO3C019554</i>	11955708..11969635 (-strand)	Homeobox protein
<i>MELO3C019555</i>	11956237..11956524 (+strand)	Zinc finger protein 1
<i>MELO3C019556</i>	11972549..11973087 (-strand)	DNA polymerase epsilon catalytic subunit A

Expression analysis for verifying the candidate gene

To verify the candidate gene, we performed expression analysis of *MELO3C019554* using qRT-PCR in the two parents, F_1 , and 10 melon germplasms with different seed coat colors. As indicated in Fig. 6a, the relative expression level of *MELO3C019554* in the seed coat of IC2508 and F_1 was significantly lower than that in the seed coat of IC2518 at 5, 10, 15, and 20 d after pollination, while the expression level did not differ significantly between $P_1(F_1)$ and P_2 on the other days. In IC2508, IC2518, and F_1 , the relative expression level of *MELO3C019554* was highest at the color accumulation stage (10 d after pollination). Thus, we analyzed the relative expression of *MELO3C019554* using qRT-PCR in five yellow seed coat melon germplasms and five brown seed coat melon germplasms at 10 d after pollination. The result showed that *MELO3C019554* exhibited significantly higher relative expression in the brown seed coat melon germplasms than in the yellow seed coat melon germplasms (Fig. 6b). Within the same seed coat color, the differing expression level may be due to the different shades of the coat (Fig. S2). In conclusion, *MELO3C019554* exhibited different expression levels in the seed coat, which were associated with the

yellow and brown seed coats. The result further revealed that *MELO3C019554* was the likely candidate gene associated with seed coat color in melon.

The candidate gene is related to flavonoid metabolites

Combined candidate gene and metabolite analysis was used to better elucidate the relationship between *MELO3C019554* and flavonoids. We found that *MELO3C019554* was related to 12 metabolites (10 metabolites upregulated and two metabolites downregulated in IC2518; Table S9).

Phylogenetic analysis

We constructed a neighbor-joining tree to evaluate the relationships between the protein (XP 008456177.1) encoded by *MELO3C019554* and its homologs. The result showed that the protein encoded by *MELO3C019554* was closely phylogenetically related to the Cucurbitaceae family, including *Cucurbita moschata*, *Cucumis sativus*, *Cucurbita maxima*, and *Cucurbita pepo* (Fig. 7a), which suggested that the protein was highly conserved in the Cucurbitaceae family. The sequence alignment analysis suggested that the region with the amino acid mutation (A→T) at residue 534 was conserved in the homologs (Fig 7b).

Discussion

We measured the flavonoid content of the two parental lines. The result showed that the flavonoid content of IC2518 (brown seed coat) was significantly higher than that of IC2508 (yellow seed coat), which revealed that melon seed coat color is related to flavonoid content. In *Arabidopsis* and *Brassica juncea*, it has been verified that flavonoid content causes a difference in seed coat color (Dixon et al. 2005; Lepiniec et al. 2006; Huang et al. 2016). In plants, flavonoids play an important role as developmental regulators and signaling molecules (Lee et al. 2018). In addition, flavonoids enhance insect and disease resistance, as well as abiotic stress tolerance (Fukushima et al. 2017; Tu et al. 2018; Dai et al. 2019; Zhao et al. 2019). The brown seeds accumulated more flavonoids than the yellow seeds, and therefore the brown seeds may show higher resistance to insects, diseases, and other stresses than the yellow seeds.

In traditional forward genetics research, it is necessary to develop many markers and use a sufficiently large population in order to obtain quality experimental results (Zhu et al. 2017). In our study, the parents had a close affinity, which made it difficult to develop a sufficient number of polymorphic markers. Thus, we took advantage of the parents and the resequencing data of the two bulked populations to obtain a large number of InDels and SNPs, from which we developed abundant markers. BSA-Seq was used to rapidly identify the preliminary region of *CmBS-1*. With the development of sequencing technology, resequencing and BSA-Seq have promoted gene mapping and cloning studies (Yang et al. 2017). For instance, using whole-genome resequencing BSA, Lv et al. (2018) fine-mapped the spotted to non-spotted trait gene in melon. Huang et al. (2017) used BSA and resequencing to fine-map the gene controlling sparse panicle in *Setaria viridis*, while Zhu et al. (2017) performed BSA-Seq and virus-induced gene

silencing strategies in cotton to fine-map and clone the virescent-1 gene. Song et al. (2017) used next-generation sequencing-based bulked segregant analysis to rapidly map two qualitative genes that control cotyledon color in soybean, and the result showed that BSA-Seq could also accelerate the mapping of more than one locus controlling a trait.

Based on BSA-Seq and genotyping, we mapped the *CmBS-1* gene in a 233.98 kb candidate region on Chr 6, which contained 12 predicted genes. We detected a nonsynonymous SNP mutation that resulted in the change of an amino acid residue in the fifth exon region of the candidate gene *MELO3C019554*. Further gene function annotation and relative expression analysis showed that *MELO3C019554* encoded a PHD transcription factor, which may play a vital role in the seed coat color accumulation stage. In early research, genes encoding transcription factors were found to control the formation of seed coat color. For instance, Chandler et al. (1989) identified the first bHLH transcription factor regulating the anthocyanin biosynthesis pathway in maize seed. A further study then confirmed that the bHLH protein controls PA biosynthesis in the pericarp in rice (Sweeney et al. 2006). Nesi et al. (2000) revealed that the bHLH protein TT8 was essential for the expression of two flavonoid late-biosynthetic genes, including *DIHYDROFLAVONOL 4-REDUCTASE (DFR)* and *BANYULS (BAN)*, and TT8, TT2, and TRANSPARENT TESTA GLABRA1 (TTG1) were also found to control flavonoid metabolism through forming a ternary complex in the *Arabidopsis* seed coat. In soybean, *Glyma09g36990* encoding an R2R3 MYB transcription factor was the candidate gene for the brown seed coat phenotype, and its gene expression was closely related to an UDP-glucose: flavonoid 3-*O*-glucosyltransferase (UF3GT) gene involved in the final step of anthocyanin biosynthesis (Gillman et al. 2011). Hong et al. (2017) identified one bHLH gene and two MYB-related genes as seed coat color candidate genes in *Brassica napus*. In cowpea, MYB domain proteins control the expression of genes that encode enzymes in anthocyanin biosynthesis, and the genes encoding MYB transcription factors were identified as candidate genes associated with a black seed coat (Herniter et al. 2018). The PHD transcription factor has been proved to be involved in abiotic stress, seed germination, and male-sterility in different species. Wei et al. (2009) found six *GmPHD* genes that encode PHD finger proteins that respond to cold, salt, drought, and abscisic acid stress. PHD-domain proteins regulate seed germination and early seedling growth in *Arabidopsis* (Molitor et al. 2014). Ye and Zhao (2016) revealed that the gene encoding the PHD protein was responsible for seed germination inhibition. Thu et al. (2019) fine-mapped the male-sterile gene and identified *Glyma.02G243200*, which encodes a PHD protein, as the candidate gene in soybean. Currently, no research has suggested that the PHD transcription factor is responsible for seed coat color formation. In our study, the genes regulated by the PHD transcription factor and the mechanism by which the PHD transcription factor controls seed coat color were unclear and thus need further exploration.

Declarations

Acknowledgement

This work was supported by the National Key Research and Development Program of China (2020YFD1000300), China Agriculture Research System of MOF and MARA (CARS-25), the Agricultural

Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-ASTIP-IVFCAAS).

Author contribution statement

HW and QF conceived and designed the research, ZH, JZ and AZ performed the experiment, ZH, XS and XM analyzed the data. ZH, HW and QF wrote the manuscript. All authors read and approved the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

Ethical standards The authors declare that the study complies with the current laws of China.

References

1. Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, Matsumura H, Yoshida K, Mitsuoka C, Tamiru M, Innan H, Cano L, Kamoun S, Terauchi R (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat Biotechnol* 30:174–178
2. Chandler VL, Radicella JP, Robbins TP, Chen J, Turks D (1989) Two regulatory genes of the maize anthocyanin pathway are homologous: isolation of B utilizing R genomic sequences. *Plant Cell* 1:1175–1183
3. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6:80–92
4. Dai Z, Tan J, Zhou C, Yang X, Yang F, Zhang S, Sun S, Miao X, Shi Z (2019) The OsmiR396-OsGRF8-OsF3H-flavonoid pathway mediates resistance to the brown planthopper in rice (*Oryza sativa*). *Plant Biotechnol J* 17:1657–1669
5. Debeaujon I, Leon-Kloosterziel KM, Koornneef M (2000) Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*. *Plant Physiol* 122:403–414
6. Dixon RA, Xie DY, Sharma SB (2005) Proanthocyanidins—a final frontier in flavonoid research. *New Phytol* 165:9–28
7. Fekih R, Takagi H, Tamiru M, Abe A, Natsume S, Yaegashi H, Sharma S, Sharma S, Kanzaki H, Matsumura H, Saitoh H, Mitsuoka C, Utsushi H, Uemura A, Kanzaki E, Kosugi S, Yoshida K, Cano L, Kamoun S, Terauchi R (2013) MutMap+: genetic mapping and mutant identification without crossing in rice. *PLoS One* 8:e68529
8. Fukushima A, Iwasa M, Nakabayashi R, Kobayashi M, Nishizawa T, Okazaki Y, Saito K, Kusano M (2017) Effects of combined low glutathione with mild oxidative and low phosphorus stress on the metabolism of *Arabidopsis thaliana*. *Front Plant Sci* 8:1464

9. Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, Gonzalez VM, Henaff E, Camara F, Cozzuto L, Lowy E, Alioto T, Capella-Gutierrez S, Blanca J, Canizares J, Ziarsolo P, Gonzalez-Ibeas D, Rodriguez-Moreno L, Droege M, Du L, Alvarez-Tejado M, Lorente-Galdos B, Mele M, Yang L, Weng Y, Navarro A, Marques-Bonet T, Aranda MA, Nuez F, Pico B, Gabaldon T, Roma G, Guigo R, Casacuberta JM, Arus P, Puigdomenech P (2012) The genome of melon (*Cucumis melo* L.). *Proc Natl Acad Sci U S A* 109:11872–11877
10. Gillman JD, Tetlow A, Lee JD, Shannon JG, Bilyeu K (2011) Loss-of-function mutations affecting a specific Glycine max R2R3 MYB transcription factor result in brown hilum and brown seed coats. *BMC Plant Biol* 11:155
11. Giner A, Pascual L, Bourgeois M, Gyetvai G, Rios P, Pico B, Troadec C, Bendahmane A, Garcia-Mas J, Martin-Hernandez AM (2017) A mutation in the melon Vacuolar Protein Sorting 41 prevents systemic infection of *Cucumber mosaic virus*. *Sci Rep* 7:10471
12. Gur A, Tzuri G, Meir A, Sa'ar U, Portnoy V, Katzir N, Schaffer AA, Li L, Burger J, Tadmor Y (2017) Genome-wide linkage-disequilibrium mapping to the candidate gene level in melon (*Cucumis melo*). *Sci Rep* 7:9770
13. Hassan MZ, Rahim MA, Natarajan S, Robin AHK, Kim HT, Park JI, Nou IS (2018) Gummy stem blight resistance in melon: inheritance pattern and development of molecular markers. *Int J Mol Sci* 19
14. Herniter IA, Munoz-Amatriain M, Lo S, Guo YN, Close TJ (2018) Identification of candidate genes controlling black seed coat and pod tip color in cowpea (*Vigna unguiculata* [L.] Walp). *G3 (Bethesda)* 8:3347–3355
15. Hill JT, Demarest BL, Bisgrove BW, Gorski B, Su YC, Yost HJ (2013) MMAPP: mutation mapping analysis pipeline for pooled RNA-seq. *Genome Res* 23:687–697
16. Hong M, Hu K, Tian T, Li X, Chen L, Zhang Y, Yi B, Wen J, Ma C, Shen J, Fu T, Tu J (2017) Transcriptomic analysis of seed coats in yellow-seeded *Brassica napus* reveals novel genes that influence proanthocyanidin biosynthesis. *Front Plant Sci* 8:1674
17. Hou J, Zhou Y-F, Gao L-Y, Wang Y-L, Yang L-M, Zhu H-Y, Wang J-M, Zhao S-J, Ma C-S, Sun S-R, Hu J-B (2018) Dissecting the genetic architecture of melon chilling tolerance at the seedling stage by association mapping and identification of the elite alleles. *Frontiers in Plant Science* 9
18. Huang P, Jiang H, Zhu C, Barry K, Jenkins J, Sandor L, Schmutz J, Box MS, Kellogg EA, Brutnell TP (2017) Sparse panicle1 is required for inflorescence development in *Setaria viridis* and maize. *Nat Plants* 3:17054
19. Huang Z, Liu L, Lu H, Lang L, Zhao N, Ding J, Xu A (2016) Development of IP and SCAR markers linked to the yellow seed color gene in *Brassica juncea* L. *Breed Sci* 66:175–180
20. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874
21. Lee K, Hwang OJ, Reiter RJ, Back K (2018) Flavonoids inhibit both rice and sheep serotonin N-acetyltransferases and reduce melatonin levels in plants. *J Pineal Res* 65:e12512

22. Lepiniec L, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, Nesi N, Caboche M (2006) Genetics and biochemistry of seed flavonoids. *Annu Rev Plant Biol* 57:405–430
23. Lian Q, Tang D, Bai Z, Qi J, Lu F, Huang S, Zhang C (2019) Acquisition of deleterious mutations during potato polyploidization. *J Integr Plant Biol* 61(1):7–11
24. Li B, Lu X, Gebremeskel H, Zhao S, He N, Yuan P, Gong C, Mohammed U, Liu W (2019) Genetic mapping and discovery of the candidate gene for black seed coat color in watermelon (*Citrullus lanatus*). *Front Plant Sci* 10:1689
25. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760
26. Li X, Chen L, Hong M, Zhang Y, Zu F, Wen J, Yi B, Ma C, Shen J, Tu J, Fu T (2012) A large insertion in bHLH transcription factor *BrTT8* resulting in yellow seed coat in *Brassica rapa*. *PLoS One* 7:e44145
27. Li X, Westcott N, Links M, Gruber MY (2010) Seed coat phenolics and the developing silique transcriptome of *Brassica carinata*. *J Agric Food Chem* 58:10918–10928
28. Li X, Xue Z, Zhu W (2011) Antioxidant activities and contents of total flavonoids and phenols from different parts of *Salvia miltiorrhiza* bunge. *Food Sci.* 3
29. Lv J, Fu Q, Lai Y, Zhou M, Wang H (2018) Inheritance and gene mapping of spotted to non-spotted trait gene *CmSp-1* in melon (*Cucumis melo* L. var. *chinensis* Pangalo). *Molecular Breeding* 38
30. Malencic D, Cvejic J, Miladinovic J (2012) Polyphenol content and antioxidant properties of colored soybean seeds from central Europe. *J Med Food* 15:89–95
31. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
32. Molitor AM, Bu Z, Yu Y, Shen WH (2014) *Arabidopsis* AL PHD-PRC1 complexes promote seed germination through H3K4me3-to-H3K27me3 chromatin state switch in repression of seed developmental genes. *PLoS Genet* 10:e1004091
33. Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* 8:4321–4325
34. Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L (2000) The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in *Arabidopsis* siliques. *Plant Cell* 12:1863–1878
35. Nimmakayala P, Tomason YR, Abburi VL, Alvarado A, Saminathan T, Vajja VG, Salazar G, Panicker GK, Levi A, Wechter WP, McCreight JD, Korol AB, Ronin Y, Garcia-Mas J, Reddy UK (2016) Genome-wide differentiation of various melon horticultural groups for use in GWAS for fruit firmness and construction of a high resolution genetic map. *Front Plant Sci* 7:1437
36. Padmaja LK, Agarwal P, Gupta V, Mukhopadhyay A, Sodhi YS, Pental D, Pradhan AK (2014) Natural mutations in two homoeologous *TT8* genes control yellow seed coat trait in allotetraploid *Brassica juncea* (AABB). *Theor Appl Genet* 127:339–347

37. Qu C, Fu F, Lu K, Zhang K, Wang R, Xu X, Wang M, Lu J, Wan H, Zhanglin T, Li J (2013) Differential accumulation of phenolic compounds and expression of related genes in black- and yellow-seeded *Brassica napus*. J Exp Bot 64:2885–2898
38. Ren Y, He Q, Ma X, Zhang L (2017) Characteristics of color development in seeds of brown- and yellow-seeded heading chinese cabbage and molecular analysis of *Brsc*, the candidate gene controlling seed coat color. Front Plant Sci 8:1410
39. Sheng Y, Wang Y, Jiao S, Jin Y, Ji P, Luan F (2017) Mapping and preliminary analysis of ABORTED MICROSPORES (*AMS*) as the candidate gene underlying the male sterility (*MS-5*) mutant in melon (*Cucumis melo* L.). Front Plant Sci 8:902
40. Song J, Li Z, Liu Z, Guo Y, Qiu LJ (2017) Next-generation sequencing from bulked-segregant analysis accelerates the simultaneous identification of two qualitative genes in soybean. Front Plant Sci 8:919
41. Sweeney MT, Thomson MJ, Pfeil BE, McCouch S (2006) Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. Plant Cell 18:283–294
42. Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S, Innan H, Cano LM, Kamoun S, Terauchi R (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 74:174–183
43. Thu SW, Rai KM, Sandhu D, Rajangam A, Balasubramanian VK, Palmer RG, Mendu V (2019) Mutation in a PHD-finger protein MS4 causes male sterility in soybean. BMC Plant Biol 19:378
44. Tu X, Liu Z, Zhang Z (2018) Comparative transcriptomic analysis of resistant and susceptible alfalfa cultivars (*Medicago sativa* L.) after thrips infestation. BMC Genom 19:116
45. Valentino R, Alexiou KG, Jordi M, Jason A, Marta P, Ryoichi Y, Satoko N, Hiroshi E, David L, Adnane B (2018) An improved assembly and annotation of the melon (*Cucumis melo* L.) reference genome. Sci Rep 8:8088
46. Wang Y, Xiao L, Guo S, An F, Du D (2016) Fine mapping and whole-genome resequencing identify the seed coat color gene in *Brassica rapa*. PLoS One 11:e0166464
47. Wei W, Huang J, Hao YJ, Zou HF, Wang HW, Zhao JY, Liu XY, Zhang WK, Ma B, Zhang JS, Chen SY (2009) Soybean GmPHD-type transcription regulators improve stress tolerance in transgenic Arabidopsis plants. PLoS One 4:e7209
48. Wei X, Zhu X, Yu J, Wang L, Zhang Y, Li D, Zhou R, Zhang X (2016) Identification of sesame genomic variations from genome comparison of landrace and variety. Frontiers in Plant Science 7
49. Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol 126:485–493
50. Yang X, Xia X, Zhang Z, Nong B, Zeng Y, Xiong F, Wu Y, Gao J, Deng G, Li D (2017) QTL mapping by whole genome re-sequencing and analysis of candidate genes for nitrogen use efficiency in rice. Front Plant Sci 8:1634

51. Ye Y, Zhao Y (2016) The pleiotropic effects of the seed germination inhibitor germostatin. *Plant Signal Behav* 11:e1144000
52. Yu CY (2013) Molecular mechanism of manipulating seed coat coloration in oilseed Brassica species. *J Appl Genet* 54:135–145
53. Zabala G, Vodkin L (2003) Cloning of the pleiotropic *T* locus in soybean and two recessive alleles that differentially affect structure and expression of the encoded flavonoid 3' hydroxylase. *Genetics* 163:295–309
54. Zhang H, Yi H, Wu M, Zhang Y, Zhang X, Li M, Wang G (2016) Mapping the flavor contributing traits on "Fengwei Melon" (*Cucumis melo* L.) chromosomes using parent resequencing and super bulked-segregant analysis. *PLoS One* 11:e0148150
55. Zhao X, Chen S, Wang S, Shan W, Wang X, Lin Y, Su F, Yang Z, Yu X (2019) Defensive responses of tea plants (*Camellia sinensis*) against tea green leafhopper attack: A multi-omics study. *Front Plant Sci* 10:1705
56. Zhu J, Chen J, Gao F, Xu C, Wu H, Chen K, Si Z, Yan H, Zhang T (2017) Rapid mapping and cloning of the virescent-1 gene in cotton by bulked segregant analysis-next generation sequencing and virus-induced gene silencing strategies. *J Exp Bot* 68:4125–4135

Figures

Figure 1

Phenotype of seed coat color in the parents. a Mature IC2508 (P1) seed with a yellow seed coat. b Mature IC2518 (P2) seed with a brown seed coat

Figure 2

The flavonoid content of the two parental lines

Figure 3

Mapping of CmBS-1 based on the application of the ED/ Δ SNP index to SNP. a ED method of SNPs between IC2508, IC2518, the yellow bulk, and the brown bulk. The black line is the fitted ED value, and the red dotted line is the significance association threshold (0.29). b Δ SNP index analysis of SNPs between IC2508, IC2518, yellow bulk, and brown bulk. The black line is the fitted Δ SNP index, and the red, blue, and green lines represent the thresholds of 99%, 95%, and 90%, respectively

Figure 4

Mapping of CmBS-1 based on the application of ED/ Δ InDel index to InDel. a ED method of InDels between IC2508, IC2518, yellow bulk, and brown bulk. The black line is the fitted ED value, and the red dotted line is the significance association threshold (0.30). b Δ InDel index analysis of InDels between IC2508, IC2518, yellow bulk, and brown bulk. The black line is the fitted Δ InDel index, and the red, blue, and green lines represent thresholds of 99%, 95%, and 90%, respectively

Figure 5

Genetic mapping of the seed color gene and analysis of the candidate gene in melon. a The genetic map constructed using the F2 population. b The 12 genes in the candidate region. c The structure of MELO3C019554. d The SNP mutation (G→A) resulted in the conversion of the amino acid (A→T) at residue 534

Figure 6

qRT-PCR analysis of MELO3C019554. a The expression level of MELO3C019554 between IC2508, IC2518, and F₁ at different developmental stages (5, 10, 15, 20, 25 and 30 days after pollination). b The relative expression level of MELO3C019554 between the yellow and brown seed coat accessions. No. 207, 209, 211, 213, and 214 are brown seed coat accessions, and No. 216, 223, 224, 226, and 228 are yellow seed coat accessions

a

b

Figure 7

Phylogenetic and alignment analysis of the candidate gene. a The phylogenetic tree of the protein encoded by MELO3C019554. The protein is indicated by a red quadrante. b The sequences alignment analysis of the candidate gene and its homologs. The mutate location is highlight by a red box

Supplementary Files

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

- [SupplementarymaterialFiguresandTables.pdf](#)