MC03g0810, an Important Candidate Gene Controlling Black Seed Coat Color in Bitter Gourd (*Momordica* spp.)

Jian Zhong[1,2,3†], Jiaowen Cheng[1,2,3†], Junjie Cui[4], Fang Hu[5], Jichi Dong[1,2,3], Jia Liu[1,2,3], Yichao Zou[1,2,3] and Kailin Hu[1,2,3]*

[1] College of Horticulture, South China Agricultural University, Guangzhou, China, [2] Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (South China), Ministry of Agriculture and Rural Affairs, Guangzhou, China, [3] Guangdong Vegetables Engineering Research Center, Guangzhou, China, [4] Department of Horticulture, Foshan University, Foshan, China, [5] Henry Fok School of Biology and Agricultural, Shaoguan University, Shaoguan, China

Seed coat color is one of the most intuitive phenotypes in bitter gourd (*Momordica* spp.). Although the inheritance of the seed coat color has been reported, the gene responsible for it is still unknown. This study used two sets of parents, representing, respectively, the intersubspecific and intraspecific materials of bitter gourd, and their respective F_1 and F_2 progenies for genetic analysis and primary mapping of the seed coat color. A large $F_{2:3}$ population comprising 2,975 seedlings from intraspecific hybridization was used to fine-map the seed coat color gene. The results inferred that a single gene, named *McSC1*, controlled the seed coat color and that the black color was dominant over the yellow color. The *McSC1* locus was mapped to a region with a physical length of ~7.8 Mb and 42.7 kb on pseudochromosome 3 *via* bulked segregant analysis with whole-genome resequencing (BSA-seq) and linkage analysis, respectively. Subsequently, the *McSC1* locus was further fine-mapped to a 13.2-kb region containing only one candidate gene, *MC03g0810*, encoding a polyphenol oxidase (PPO). Additionally, the variations of *MC03g0810* in the 89 bitter gourd germplasms showed a complete correlation with the seed coat color. Expression and PPO activity analyses showed a positive correlation between the expression level of *MC03g0810* and its product PPO and the seed coat color. Therefore, *MC03g0810* was proposed as the causal gene of *McSC1*. Our results provide an important reference for molecular marker-assisted breeding based on the seed coat color and uncover molecular mechanisms of the seed coat color formation in bitter gourd.

Keywords: bitter gourd, seed coat color, bulked segregant analysis, linkage analysis, fine-mapping

INTRODUCTION

The seed coat color is one of the most important agronomic traits in many crops. Seeds with colored coats have higher nutritional and healthy values (Zhu, 2018), for instance, yellow coat seeds of Brassicaceae crops have a lower lignin concentration, thinner shells, less fiber, and higher amounts of oil and protein than the black ones (Marles and Gruber, 2004; Yu et al., 2007). The seed coat

pigmentation of soybean (*Glycine max*) affects the levels of isoflavones and fatty acids, with the black varieties exhibiting the highest antioxidant activity compared with the other varieties (Cho et al., 2013; Lee et al., 2017). The seed coat color is also closely related to the seed viability, weight, and germination rate (Mavi, 2010). Additionally, dark-colored seeds have higher resistance to various pathogens and herbivores than those with light color, probably because higher polyphenol oxidase activity in the former enhances their stress tolerance (Marles et al., 2008; Fuerst et al., 2014) and the dark color is more likely to form a natural camouflage to prevent feeding by birds (Zhu et al., 2011).

In Cucurbitaceae crops, the seed coat displayed a wide range of colors. For example, *Luffa acutangula* and *Luffa cylindrica* have black seed coats, *Bryonia alba*, *Diplocyclos palmatus,* and *Nothoalsomitra suberosa* have brown, while *Acanthosicyos horridus* and *Lagenaria spaerica* have whitish yellow, and other cucurbits mostly have yellow to brown seed coats (Heneidak and Khalik, 2015). Each cucurbit has distinct seed features, for example, watermelon (*Citrullus lanatus*) seed coats are flat black, dotted black, green, tan, clump, red, white with tan tip, or white with pink tip (McKay, 1936; Poole et al., 1941). Therefore, Cucurbitaceae crops are ideal for studying the inheritance of the seed coat color.

The seed coat color is genetically controlled in Cucurbitaceae. In watermelon, a four-gene model, including *R*, *T*, *W*, and *D* genes, was proposed to elucidate the inheritance of the seed coat color, of which gene *D* determines the phenotype only when the other three genes are in the dominant state (Poole et al., 1941). In addition, T^1, different from the *T* locus, was reported to produce dotted black (RT^1W, Rt^1W) and tan (rT^1W) seed coats (Paudel et al., 2019). Recently, the *R*, T^1, *W,* and *D* genes have been mapped on chromosomes 3, 5, 6, and 8, respectively, based on the QTL-seq analysis (Paudel et al., 2019). Furthermore, *Cla019481*, encoding a polyphenol oxidase (PPO) protein, was proposed as an important candidate for the *W* gene responsible for the black seed coat formation in watermelon (Li et al., 2018, 2020). The white seed coat color in pumpkin (*Cucurbita maxima*) is controlled by a single gene, *wsc,* which is recessive to the yellow seed coat color (Shi et al., 2021). A candidate gene (*CmaCh15G005270*) encoding an MYB transcription factor has been recently proposed for the *wsc* gene using the fine-mapping approach (Shi et al., 2021). In melon (*Cucumis melo*), a single dominant gene, *CmSC1*, controls the white seed coat color that is dominant over the yellow color (Ma et al., 2021). Conversely, the yellow seed coat color is controlled by a single dominant gene, *CmBS-1,* and is dominant over the brown color (Hu et al., 2021). Recently, *MELO3C014406*, a melon orthologous gene of the *Arabidopsis AtTT8* encoding a basic helix–loop–helix domain protein, has been identified as a candidate gene of the *CmSC1* (Ma et al., 2021). The *MELO3C019554*, encoding a homeobox protein, has been proposed as a candidate gene of *CmBS-1 via* the fine-mapping strategy (Hu et al., 2021).

Bitter gourd (*Momordica* spp.), an herbaceous climbing plant of Cucurbitaceae, originated from the tropical regions of Africa and is currently widely distributed in tropical and subtropical regions of Africa and Asia (Schaefer and Renner, 2010). Its seeds have various bioactive components such as ribosome-inactivating

proteins (Puri et al., 2009), trypsin inhibitors (Lo et al., 2014), cucurbitane triterpenoids (Chen et al., 2005; Shah et al., 2014), and hypoglycemic peptide (Khanna et al., 1981; Wang et al., 2011). Most cultivated bitter gourds have yellow seeds; however, black, brown patched, brownish tan, and whitish brown-colored seeds have also been reported in the wild and semi-domesticated varieties (Kole et al., 2020). Genetic analysis indicated that the black seed coat color of the bitter gourd is dominant over the yellow one, thus conforming to the genetic model of a single dominant gene (Liu, 2011; Tan et al., 2013). In addition, the black coloration was dominant over the creamy seed coat color but exhibited a digenic inheritance mode. Furthermore, one of two loci was mapped on LG3 between the amplified fragment length polymorphism (AFLP) markers, E12M47a and E11M48a, which were 52.5 cM apart (Kole et al., 2012). A region spanning ∼300 kb on pseudochromosome 3 was reported to be related to the seed coat color through a genome-wide association study (Cui et al., 2022). To our knowledge, there is no further study on the mining of the seed coat color genes in bitter gourd until now.

Thus, this study aimed at identifying the gene regulating the black seed coat color in bitter gourds. We used intersubspecific and intraspecific F_2 populations for the genetic analysis to ensure the inheritance reliability of the seed coat color in bitter gourd. Forward genetics strategies, including BSA-seq, linkage analysis, and fine-mapping, were used to identify genes regulating the seed coat color in the bitter gourd. Our study provides a basis for molecular marker-assisted breeding and understanding the molecular mechanisms of the seed coat color formation in bitter gourd.

MATERIALS AND METHODS

Plant Materials

Four inbred lines of bitter gourd were used in this study. These included two lines with yellow seed coats, FOLI112 (*Momordica charantia* ssp. *charantia*) and S156 (*M. charantia* ssp. *charantia*), and another two lines with black seed coats, THMC170 (*M. charantia* ssp. *macroloba*) and 8–201 (*M. charantia* ssp. *charantia*). Among them, FOLI112 is a gynoecious line, which is a near-isogenic line of Dali-11, whose whole genome sequence is available (Cui et al., 2020), while THMC170 belongs to a variety with elongated seeds and small fruits. All the four inbred lines have completed resequencing. An intersubspecific F_1 generation and F_2 segregating population consisting of 120 individuals were constructed through artificial hybridization and self-pollination using FOLI112 and THMC170 as female and male parents, respectively. Similarly, an intraspecific F_1 generation and F_2 segregating population comprising 147 individuals were constructed using S156 and 8–201 as female and male parents, respectively. The two sets of F_1 and F_2 and their respective parental lines were then used for inheritance analysis of the seed coat color. Subsequently, the intersubspecific F_2 population and its parental lines were used for BSA-seq analysis, and the intraspecific F_2 population was used for linkage analysis. A large $F_{2:3}$ population consisting of 2,975 seedlings self-crossed from several intraspecific F_2 individuals was used for screening

the recombinants *via* a fine-mapping strategy. Additionally, 89 bitter gourd materials, including 58 inbred lines and 31 F_1 hybrids (**Supplementary Table 1**), were selected for cloning and comparing the candidate gene controlling the seed coat color. All plant materials, including the recombinants, were cultivated in an experimental field at the South China Agriculture University Teaching and Research Base in Zengcheng District, Guangzhou (23.24N, 113.64E).

Bulked Segregant Analysis With Whole-Genome Resequencing

The DNA of all samples was extracted using the CTAB method (Porebski et al., 1997). For BSA-seq, yellow and black DNA pools were constructed by sampling equal amounts of young fresh leaf tissues from each of the 30 individuals with yellow and black seed coats from the intersubspecific F_2 population.

Sequencing libraries of the yellow and black DNA pools and their parental DNA pools were prepared using the AxyPrep Mag PCR clean-up kit, according to the manufacturer's instructions (Illumina). The generated sequences were further amplified using 150 bp paired-end sequencing on an Illumina X-ten system. Raw reads were filtered by removing low-quality reads containing adapters, unknown nucleotides (N) of more than 10%, and low-quality (Q-value ≤ 20) bases of more than 50%. Subsequently, clean reads were aligned to the Dali-11 reference genome (Cui et al., 2020) using Burrows–Wheeler Aligner (Li and Durbin, 2009), and the alignment files were converted to SAM/BAM files using SAMtools (Li H. et al., 2009). The Genome Analysis Toolkit was used for multi-sample variant calling, and its VariantFiltration tools with uniform standards (Windows 4, filter "QD < 4.0| | FS > 60.0 | | MQ < 40.0," G_filter "GQ < 20") were used to filter single nucleotide polymorphisms (SNPs) and insertions–deletions (InDels) (McKenna et al., 2010). We used ANNOVAR software (Wang et al., 2010) for aligning and annotating the SNPs and InDels to determine the physical positions of each variant.

Frequency distributions of each SNP (SNP-index) in the DNA pool were analyzed by the sliding window analysis using a window size of 1,000 kb and a step length of 100 kb. The SNPs with SNP-index values less than 0.3 or greater than 0.7 and a read depth of less than 7 were excluded. The ΔSNP-index was calculated by subtracting the SNP-indexes between the yellow and black DNA pools. Finally, the initial location of the genes regulating the seed coat color was identified using the positive or negative peak regions with a 95 and 99% confidence interval in 10,000 bootstrap replicates.

Linkage Analysis

InDel primer pairs in the 99% confidence interval of BSA-seq based on the intersubspecific materials were designed by Primer 3 software, based on alignment results of S156 and 8–201 to the Dali-11 reference genome (Cui et al., 2020) aligned by SOAP2 (Li R. et al., 2009). Moreover, the polymerase chain reaction (PCR) was performed in a 10 μL reaction mixture containing 50–100 ng of template DNA, 0.2 μL of forward and reverse primers (10 μmol/L), and 5 μL of Green Taq Mix (Vazyme,

Nanjing, China) for identifying the polymorphic InDel markers between the parental lines S156 and 8–201. The PCR conditions were as follows: an initial denaturation cycle of 3 min at 94°C; 34 cycles of denaturation at 94°C for 15 s, annealing at 55°C for 15 s, and extension at 72°C for 30 s; and a final extension cycle of 5 min at 72°C. The PCR products were then genotyped in 6% polyacrylamide gel electrophoresis (PAGE). Subsequently, linkage analysis was performed using the JoinMap 4.0 (Van Ooijen, 2006). Additionally, the seed coat color of each F_2 plant was converted into genotypic codes (A/C), and the linkage groups were established using the independence logarithm of the odds (LOD). The LOD parameter used had a LOD-score threshold of 3.0 and a recombination frequency smaller than 0.4. A regression algorithm was used for mapping, and recombination values were converted to genetic distances using Haldane's mapping function.

Fine-Mapping

A large $F_{2:3}$ population consisting of 2,975 seedlings self-crossed from several S156 × 8–201 F_2 individuals was used to screen recombinants. These S156 × 8–201 F_2 plants had sequential heterozygosity between SC_7 and SC_12, two flanking markers of co-segregation interval based on the linkage analysis. Thereafter, a precise interval was identified based on phenotypes and genotypes of recombinant plants.
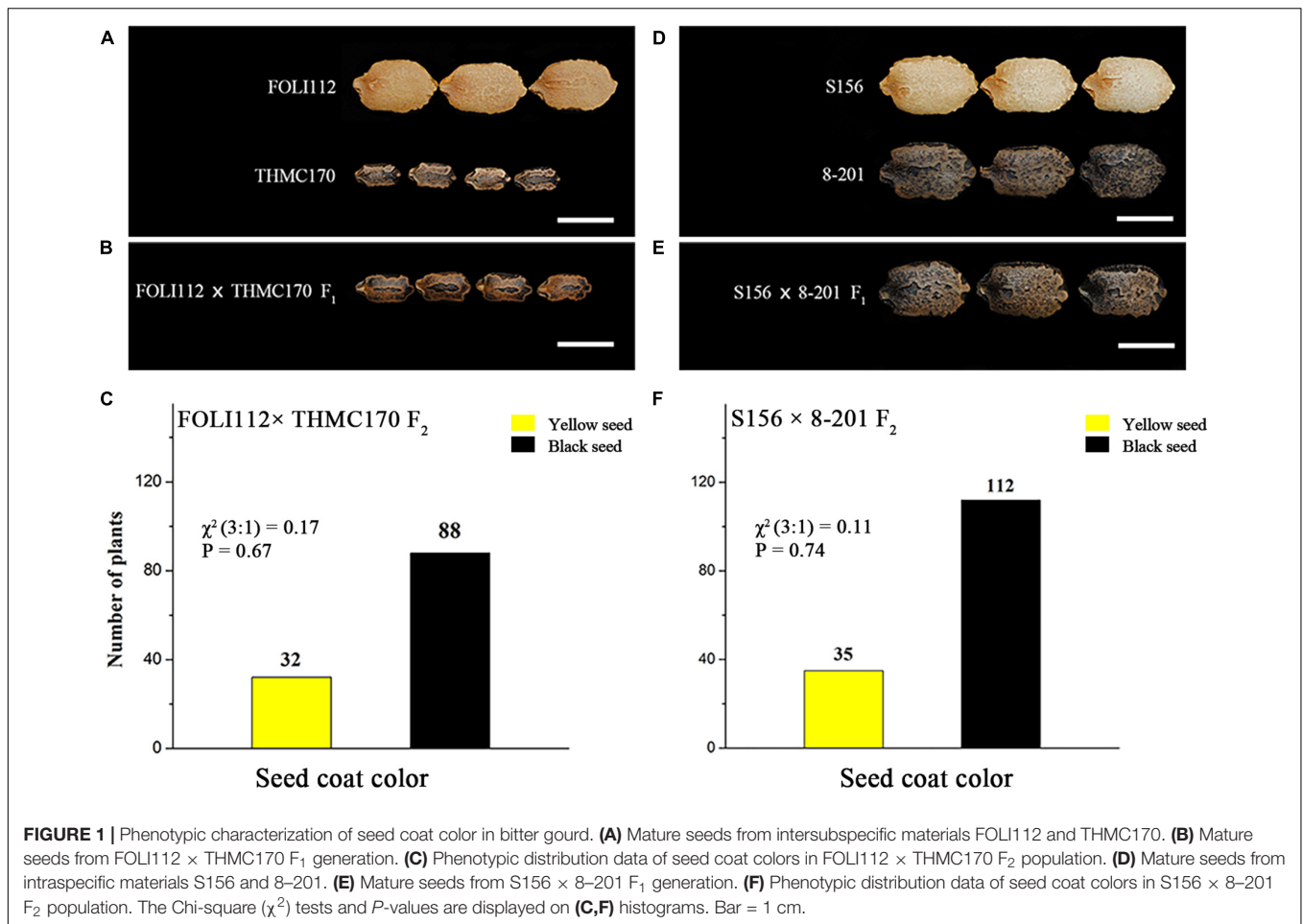
Gene Cloning and Sequence Analyses

The sequences of primer pairs used for cloning the gene encoding the seed coat color are listed in **Supplementary Table 2**. The PCR was performed in a 20 μL mixture using Phanta Max Super-Fidelity DNA Polymerase (Vazyme, Nanjing, China) according to the manufacturer's protocol. Subsequently, all PCR products were purified and cloned into the pMD[TM]19-T vector (Takara, Japan). Randomly selected positive colonies (at least three) for each amplicon were sequenced and assembled. Sequence alignment was performed using Clustal Omega,[1] and the Conserved Domain tool from NCBI[2] was used for the conserved domain analysis.

Expression Analyses

Tissues from S156 and 8–201 lines were excised and frozen in liquid nitrogen for total RNA extraction. These tissues included root, stem, leaf, ovary, and sarcocarp at 18 days after flowering (DAF) and seed coat and embryo at three different developmental stages (12, 18, and 24 DAF). The Eastep Super Total RNA Extraction Kit (Promega, Shanghai, China) and the Eastep RT Master Mix Kit (Promega, Shanghai, China) were used for total RNA isolation and synthesis of the first cDNA strand, respectively, according to the manufacturer's protocol. The real-time quantitative PCR (qRT-PCR) was carried out on a CFX384 Real-Time System (Bio-Rad, CA, United States) using TB Green® Premix Ex Taq[TM] II (Takara Japan). The relative expression levels with three biological and technical replicates for each sample were calculated using the delta–delta Ct method

---

[1]https://www.ebi.ac.uk/Tools/msa/clust alo/

[2]https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi

FIGURE 1 | Phenotypic characterization of seed coat color in bitter gourd. **(A)** Mature seeds from intersubspecific materials FOLI112 and THMC170. **(B)** Mature seeds from FOLI112 × THMC170 F_1 generation. **(C)** Phenotypic distribution data of seed coat colors in FOLI112 × THMC170 F_2 population. **(D)** Mature seeds from intraspecific materials S156 and 8–201. **(E)** Mature seeds from S156 × 8–201 F_1 generation. **(F)** Phenotypic distribution data of seed coat colors in S156 × 8–201 F_2 population. The Chi-square (χ^2) tests and P-values are displayed on **(C,F)** histograms. Bar = 1 cm.

($2^{-\triangle\triangle Ct}$) (Livak and Schmittgen, 2001). The primer sequences of expression analysis are provided in **Supplementary Table 2**.

Polyphenol Oxidase Activity Analyses

The seed coats were collected from the parental lines S156 and 8–201 at 18 and 24 DAF, respectively, and were immediately snap-frozen in liquid nitrogen. The bicinchoninic acid (BCA) kit (Elabscience, Wuhan, China) was used to extract crude enzymes and determine total protein concentration. Subsequently, PPO activity was evaluated using the PPO Colorimetric Assay Kit (Elabscience, Wuhan, China) at the OD value of 410 nm on an ultraviolet spectrophotometer. This was because PPO catalyzes the oxidation of phenolic compounds to quinones, which have a characteristic absorption peak at 410 nm (Steffens et al., 1994). The detection of each sample was repeated three times.

RESULTS

Inheritance of Seed Coat Color

The colors of dried seed coats from one intersubspecific hybridization, including 30 individuals each of female parent FOLI112 and male parent THMC170, 30 individuals of their F_1 plants, and 120 individuals of their F_2 plants, were evaluated. The

results showed that seed coat colors of FOLI112 and THMC170 plants were yellow and black, respectively (**Figure 1A**). Moreover, all F_1 plants exhibited a black seed coat coloration (**Figure 1B**). Seed coat colors of FOLI112 × THMC170 F_2 population segregated, with 88 showing black and 32 showing yellow coloration (**Figure 1C**). We also evaluated the seed coat color of four progenies from an intraspecific hybridization, including 30 individuals each of female parent S156 and male parent 8–201, 30 individuals of their F_1 plants, and 147 individuals of their F_2 plants. Similarly, seed coat colors of S156, 8–201, and F_1 plants showed yellow, black, and black, respectively (**Figures 1D,E**). The S156 × 8–201 F_2 population was segregated, with 112 plants showing black and 35 plants showing yellow coloration (**Figure 1F**). The Chi-square test indicated that the black/yellow ratios of both F_2 populations conformed to the expected ratio of 3:1 ($P > 0.05$). These results indicated that the seed coat color of bitter gourd is controlled by a single gene, named *McSC1*, making the black color dominant over the yellow color.

Identification of the *McSC1* Locus in Intersubspecific Materials

To primary map the *McSC1* locus, we conducted BSA-seq using a set of intersubspecific materials. We obtained 72.23

Gb of clean data for FOLI112, THMC170, black, and yellow DNA pools with an average of 18.06 Gb being equal to about 60 × depth per sample (**Supplementary Table 3**). The average

Q20 value of the four samples was 95.19%, indicating a high quality of the generated resequencing data (**Supplementary Table 3**). Subsequently, a total of 0.02, 2.64, 2.73, and 2.74
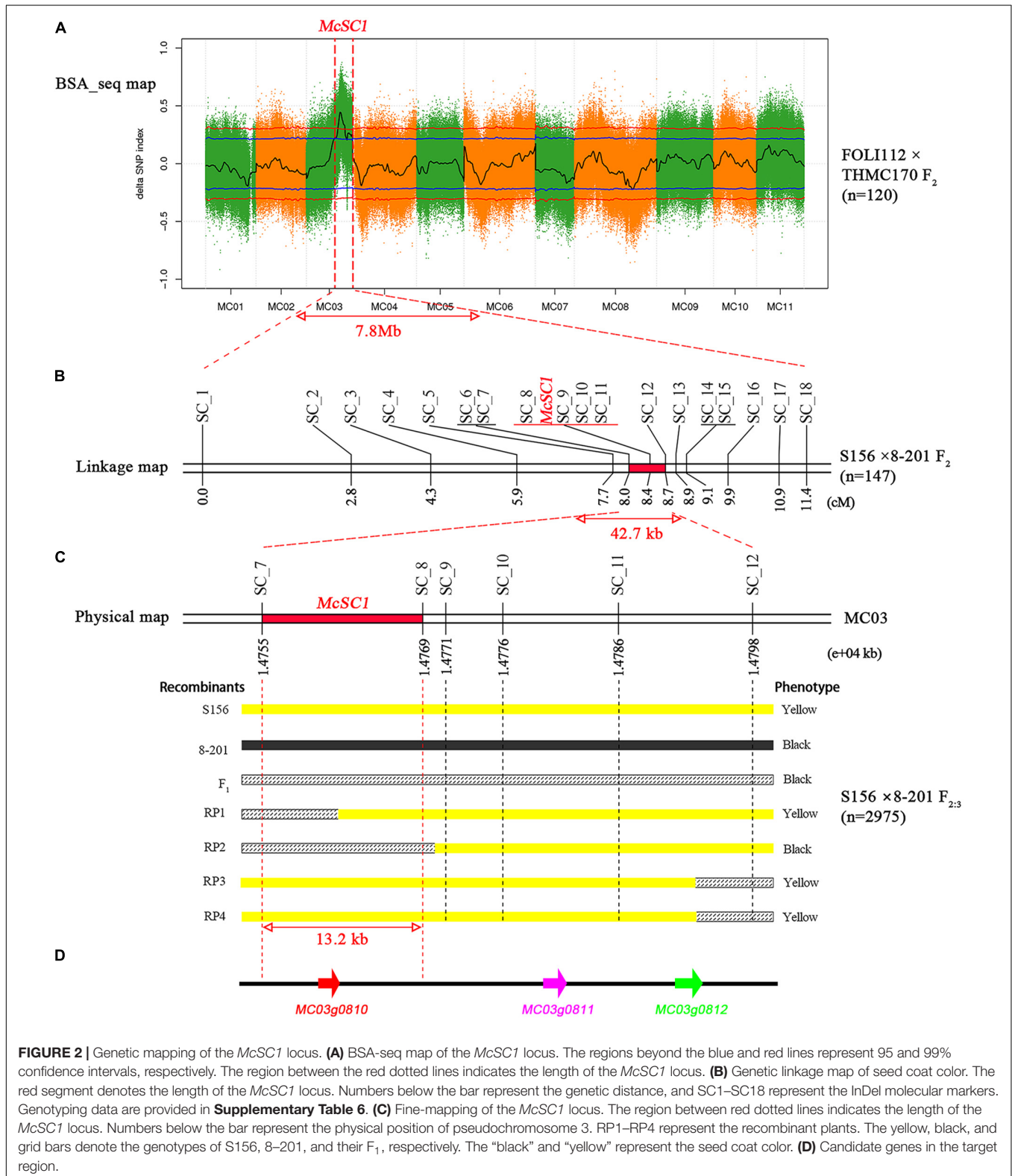

FIGURE 2 | Genetic mapping of the *McSC1* locus. **(A)** BSA-seq map of the *McSC1* locus. The regions beyond the blue and red lines represent 95 and 99% confidence intervals, respectively. The region between the red dotted lines indicates the length of the *McSC1* locus. **(B)** Genetic linkage map of seed coat color. The red segment denotes the length of the *McSC1* locus. Numbers below the bar represent the genetic distance, and SC1–SC18 represent the InDel molecular markers. Genotyping data are provided in **Supplementary Table 6**. **(C)** Fine-mapping of the *McSC1* locus. The region between red dotted lines indicates the length of the *McSC1* locus. Numbers below the bar represent the physical position of pseudochromosome 3. RP1–RP4 represent the recombinant plants. The yellow, black, and grid bars denote the genotypes of S156, 8–201, and their F_1, respectively. The "black" and "yellow" represent the seed coat color. **(D)** Candidate genes in the target region.

million SNPs were identified for FOLI112, THMC170, black, and yellow DNA pools, respectively, following sequence alignment (**Supplementary Table 4**). After filtering, 2.43 million SNPs were concurrently obtained from the two parental lines, and the two offspring pools were selected for calculating the SNP- and ΔSNP-indexes (**Supplementary Table 5**). Finally, a single 95% confidence interval from 11,600,001 to 19,355,601 bp on the pseudochromosome 3, covering about 7.8 Mb in physical length, was identified based on the ΔSNP-index statistics (**Figure 2A** and

Supplementary Table 5), and this region was inferred to be the *McSC1* locus.

Consistent Function of *McSC1* in Intersub- and Intraspecific Materials

To confirm whether *McSC1* plays the same role in intersubspecific and intraspecific materials, we used an intraspecific S156 × 8–201 F$_2$ population for linkage
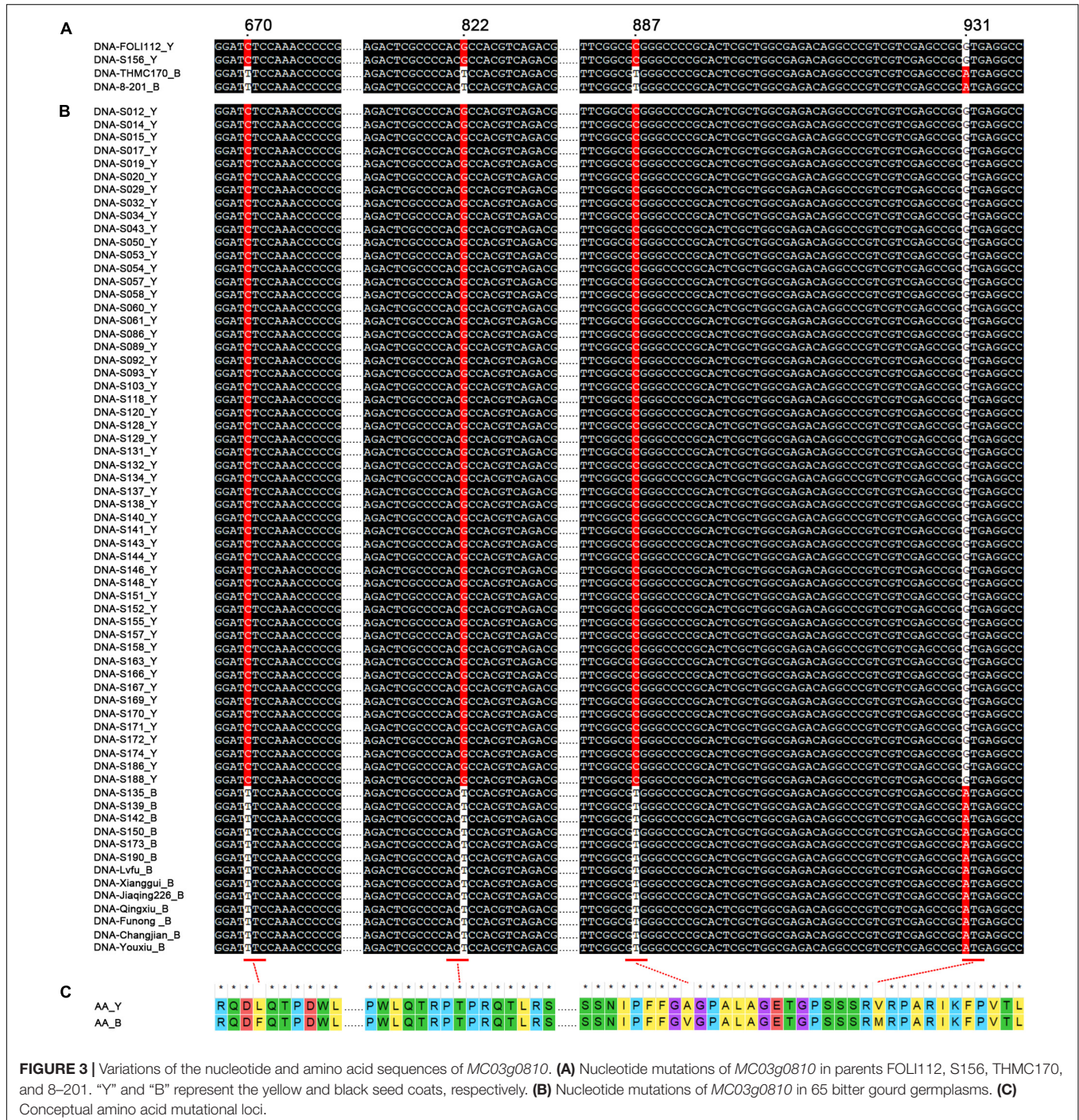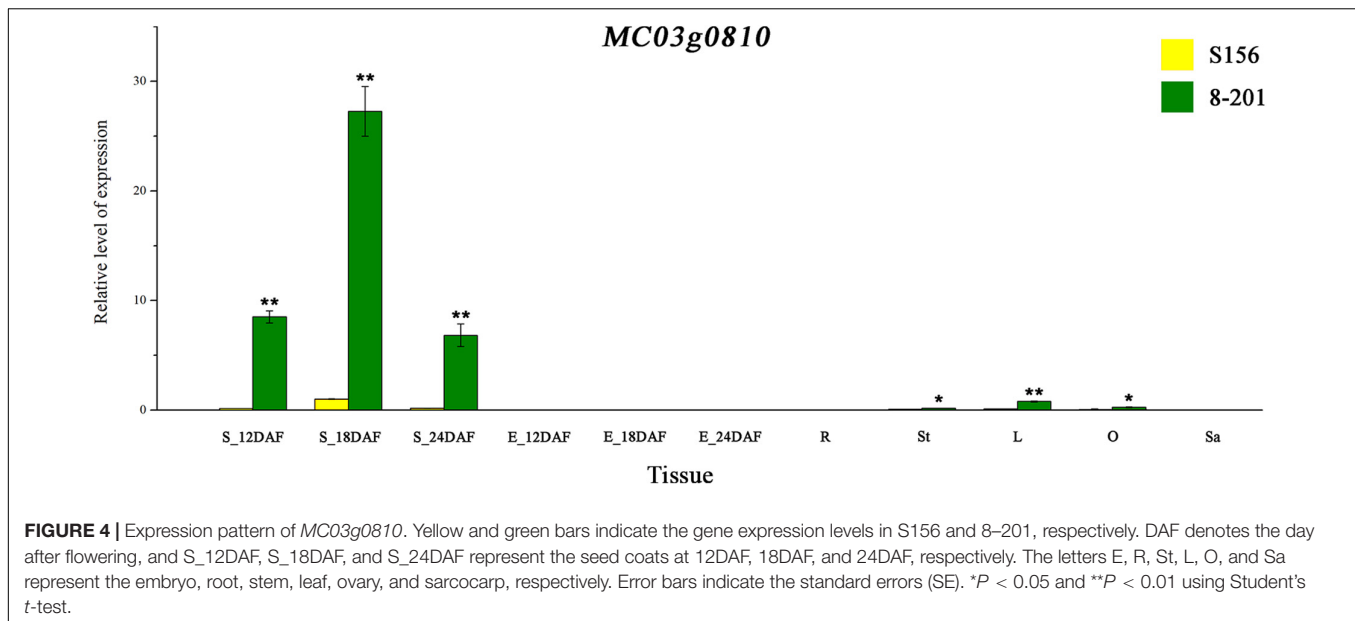

FIGURE 3 | Variations of the nucleotide and amino acid sequences of *MC03g0810*. **(A)** Nucleotide mutations of *MC03g0810* in parents FOLI112, S156, THMC170, and 8–201. "Y" and "B" represent the yellow and black seed coats, respectively. **(B)** Nucleotide mutations of *MC03g0810* in 65 bitter gourd germplasms. **(C)** Conceptual amino acid mutational loci.

FIGURE 4 | Expression pattern of *MC03g0810*. Yellow and green bars indicate the gene expression levels in S156 and 8–201, respectively. DAF denotes the day after flowering, and S_12DAF, S_18DAF, and S_24DAF represent the seed coats at 12DAF, 18DAF, and 24DAF, respectively. The letters E, R, St, L, O, and Sa represent the embryo, root, stem, leaf, ovary, and sarcocarp, respectively. Error bars indicate the standard errors (SE). *$P < 0.05$ and **$P < 0.01$ using Student's *t*-test.

analysis between the seed coat color and *McSC1*. A total of 18 polymorphic InDel markers between S156 and 8–201 (**Supplementary Table 2**) located in a more rigorous (99%) confidence interval of the BSA-seq analysis using the intersubspecific population were developed and used to genotype the 147 individuals of the S156 × 8–201 F_2 population. Linkage analysis indicated that the 18 markers exhibited different linkage relationships with the seed coat color (*McSC1*), ranging from 0 to 8.4 cM (**Figure 2B**). Among them, four markers, namely, SC_8, SC_9, SC_10, and SC_11, co-were segregated with the seed coat color in S156 × 8–201 F_2 population (**Figure 2B**). Thus, based on the results of BSA-seq and linkage analysis, we inferred that the causal gene controlling the seed coat color variation in intersubspecific and intraspecific materials of the bitter gourd would be the same. Furthermore, the *McSC1* locus was delimited into a candidate interval between markers SC_7 and SC_12 with a physical length of ∼42.7 kb, ranging from 14,755,455 to 14,798,136 bp on pseudochromosome 3 (**Figures 2A,B**).
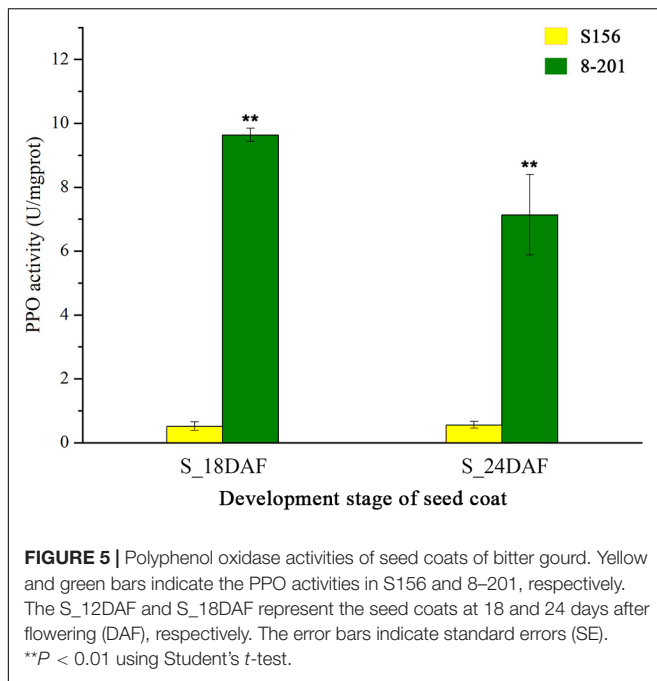
Fine-Mapping of the *McSC1* Locus

A large S156 × 8–201 $F_{2:3}$ population comprised of 2,975 seedlings was used to identify recombinant plants using two flanking markers, namely, SC_7 and SC_12. As a result, four recombinant plants (RP1–RP4) were identified. We inferred that the *McSC1* locus was delimited in a 13.2-kb region between the markers SC_7 and SC_8, following seed color observation and genotyping of the four recombinants with four co-segregating markers (**Figure 2C**). Based on the Dali-11 reference genome (Cui et al., 2020), we found only one protein-coding gene in this fine-mapping region, namely, *MC03g0810*, which encodes a PPO and has one intron measuring 1,570 bp and two exons measuring 482 and 862 bp in length (**Figure 2D**). The conserved domain analysis showed that there are three conserved domains in *MC03g0810*, including tyrosinase, PPO_DWL, and PPO_KFDV families (**Supplementary Figure 1**).

Sequence Variation of *MC03g0810* in Different Bitter Gourd Germplasms

We obtained four SNPs, namely SNP_0810:670, SNP_0810:822, SNP_0810:887, and SNP_0810:931, within the *MC03g0810* coding region in the FOLI112, S156, THMC170, and 8–201 parental lines (**Figure 3A**). These four SNPs formed two haplotypes embodying C–G–C–G and T–T–T–A (in the order: SNP_0810:670, SNP_0810:822, SNP_0810:887, and SNP_0810:931). The C–G–C–G haplotype (FOLI112 and S156) had yellow seed coats, while the T–T–T–A haplotype (THMC170 and 8–201) displayed black seed coats. Additionally, we sequenced and compared *MC03g0810* among 58 + 7 inbred lines (seven F_1 hybrids with homozygous *MC03g0810*) exhibiting distinct genetic backgrounds and 24 F_1 hybrids containing heterozygous *MC03g0810* to evaluate the relationship between the haplotype and the seed coat color. The results showed that all the 52 inbred lines with yellow seed coats were C–G–C–G haplotypes, and 13 inbred lines with black seed coats were T–T–T–A haplotypes, which was consistent with previous results (**Figure 3B** and **Supplementary Table 1**). Meanwhile, the 24 F_1 hybrids were heterozygous for the C–G–C–G and T–T–T–A haplotypes (**Supplementary Table 1**). Generally, the variation of the seed coat color in different bitter gourd germplasms correlates with the *MC03g0810* haplotype. Moreover, the four SNPs of *MC03g0810* generated two types of protein sequences corresponding to the two haplotypes (**Figure 3C**).

Characterization of *MC03g0810* Expression and Polyphenol Oxidase Activity

The results of tissue-specific expression analysis showed that *MC03g0810* had different expression levels in the seed coat, leaves, stems, and ovaries; however, it was barely expressed in

FIGURE 5 | Polyphenol oxidase activities of seed coats of bitter gourd. Yellow and green bars indicate the PPO activities in S156 and 8–201, respectively. The S_12DAF and S_18DAF represent the seed coats at 18 and 24 days after flowering (DAF), respectively. The error bars indicate standard errors (SE). **$P < 0.01$ using Student's t-test.

embryos, roots, and sarcocarps (**Figure 4**). Furthermore, the *MC03g0810* expression levels in the seed coat were significantly higher in three developmental stages (12 DAF, 18 DAF, and 24 DAF) of 8–201 than in S156 (**Figure 4**). These results suggest that *MC03g0810* has a relatively higher tissue specificity and shows the highest expression level at the 18 DAF stage of black seed coats, indicating that *MC03g0810* is associated with the black seed coat.

The PPO-induced browning is a major path darkening color in various plant tissues (Yoruk and Marshall, 2003). Therefore, considering that *MC03g0810* is a member of the PPO gene family, we compared the PPO activity of seed coats at 18 and 24 DAF between the S156 and 8–201 lines. Results showed that the PPO activity of the 8–201 seed coat was approximately 19.2 and 12.8 times higher than that of the S156 seed coat at 18 and 24 DAF, respectively (**Figure 5**). This was consistent with the differential expression of *MC03g0810* in the seed coats of S156 and 8–201. We speculate that the formation of the black seed coat in 8–201 may be correlated with higher PPO activity, resulting from the higher expression of *MC03g0810*.

DISCUSSION

The color variation of the yellow and black seed coats was observed in both cultivated and wild bitter gourd germplasms (Kole et al., 2020). Previous studies have established that the seed coat color is controlled by a single gene in the bitter gourd and that the black color is dominant over the yellow color in intersubspecific and intraspecific materials (Liu, 2011; Tan et al., 2013). Given this fact, we simultaneously used two F$_2$ segregating populations from two respective

intersubspecific and intraspecific hybridizations for the genetic analysis of the seed coat color. The results were consistent with those of previous studies (**Figure 1**). However, it is not clear whether these two genes from different genetic backgrounds are the same.

With the rapid development and remarkably decreasing cost of sequencing technologies, whole genome sequencing and resequencing of bitter gourd had been completed (Urasaki et al., 2017; Cui et al., 2020; Matsumura et al., 2020), which laid a solid foundation for gene mapping. Accordingly, we identified the *McSC1* locus related to the seed coat color in an intersubspecific F$_2$ population using BSA-seq (**Figure 2A**). Moreover, linkage analysis verified that this *McSC1* locus is also related to the seed coat color in an intraspecific F$_2$ population (**Figure 2B**). These studies provide evidence that the causal gene responsible for the seed coat color is the same in intersubspecific and intraspecific bitter gourd germplasms.

In addition, studies have suggested that dominance of the black seed coat coloration of bitter gourd over the creamy color conforms to the digenic inheritance mode, of which only one of two genes was mapped on LG3 between the AFLP markers, namely, E12M47a and E11M48a (Kole et al., 2012). Due to the dearth of information on E12M47a and E11M48a sequences, we failed to compare the locus of these sequences with the *McSC1* locus. However, we detected the same digenic inheritance pattern in some other bitter gourd materials, indicating that there may be other genes regulating the formation of black seed coats in the bitter gourd (not published).

To scale down the probable size of the *McSC1* locus, we conducted fine-mapping in a large intraspecific F$_{2:3}$ population. As a result, the *McSC1* locus was further fine-mapped to a 13.2-kb region (**Figure 2C**), much narrower than the ~300 kb determined by Cui et al. (2022) *via* a genome-wide association study. By comparing it to the Dali-11 reference genome (Cui et al., 2020), we found that the 13.2-kb region contained only one candidate gene, *MC03g0810*, encoding a PPO protein. Therefore, it is evident that the forward genetics strategy provides the most direct evidence, proving that *MC03g0810* is the causal gene of *McSC1* in bitter gourd.

We cloned the *MC03g0810* genes of 89 germplasms from different genetic backgrounds, including inbred lines and commercial F$_1$ hybrids, to further understand the relationship between the variation of *MC03g0810* and the seed coat color in different bitter gourd materials. As a result, we detected two haplotypes, namely, C–G–C–G and T–T–T–A, whereby C–G–C–G was contained in germplasms with yellow seed coats, while T–T–T–A was only found in the germplasms with black seed coats (**Figure 3B** and **Supplementary Table 1**). A haplotype is a combination of a series of genetic variation loci coexisting on the same chromosome and is an essential aspect of genetic research (Fan et al., 2011; Snyder et al., 2015). Thus, identifying *MC03g0810* haplotypes will help understand the relationship between different SNP combinations and the seed coat color. It is also crucial in determining a suitable marker-assisted selection for the *MC03g0810*-mediated seed coat color in bitter gourd breeding programs.

Among the four *MC03g0810* SNPs, three (SNP_0810:670, SNP_0810:887, and SNP_0810:931) are non-synonymous, while the remaining one (SNP_0810:822) exhibits synonymous mutations (**Figure 3C**). Furthermore, the conserved domain analysis suggested that SNP_0810:822, SNP_0810:887, and SNP_0810:931 are located in the non-conservative domain region, and only SNP_0810:670 belonged to the conserved DWL domain of the PPO protein (**Supplementary Figure 1**). Therefore, the SNP_0810:670 mutation, involving the substitution of leucine (L) with phenylalanine (F), should be focused on, since it may be the causal mutation that downregulates the expression of *MC03g0810* and thus reduces the PPO activity, resulting in the seed coat color changes (**Figure 3**).

Polyphenol oxidases are copper metalloenzymes widely existing in plants and are involved in the browning of plant tissues (Yoruk and Marshall, 2003; Zhang and Sun, 2021). Previous studies have shown that mutations of the *PPO* gene, which darken the seed coat color, exist in rice (Yu et al., 2008), wheat (Beecher and Skinner, 2011; Beecher et al., 2012), and barley (Taketa et al., 2010), among other crops. Recently, a similar PPO-mediated mechanism has been reported in Cucurbitaceae; for example, *Cla019481* encoding a PPO protein was considered a candidate gene of *ClCS1* for the black seed coat color in watermelon (Li et al., 2020). Spatial–temporal expression patterns of *Cla019481* and *MC03g0810* (**Figure 4**) and variations in the PPO activities of seed coats (**Figure 5**) exhibited high similarities between watermelon and bitter gourd. These results suggest that the regulatory mechanisms of seed coat color blackening may be consistent in watermelon and bitter gourd, thus providing a reference for studies on the seed coat color in other cucurbits. Notably, according to the Dali-11 reference genome of bitter gourd (Cui et al., 2020), *MC03g0810* and the other three paralogs of *PPO* genes, including *MC03g0811*, *MC03g0812,* and *MC03g0813,* are adjacent in their physical positions, forming a *PPO* cluster. The *PPO* clusters have been reported in rice, tomato, and red clover (Newman et al., 1993; Winters et al., 2009); however, there is no report on *PPO* clusters in Cucurbitaceae so far. Our expression analyses suggested that the *PPO* genes have similar expression patterns with *MC03g0810* in seed coats (not published). Therefore, more efforts are needed to determine whether there exists a relationship between the other three *PPO* paralogs and the seed coat color in bitter gourd.

CONCLUSION

The seed coat color of bitter gourd is controlled by a single gene named *McSC1*. The *McSC1* locus was fine-mapped to a 13.2-kb region on pseudochromosome 3, which contains only one candidate gene, *MC03g0810*, encoding a polyphenol oxidase protein. Based on the sequence, expression, and PPO activity analyses, *MC03g0810* was considered the causal gene of *McSC1*.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The name of the repository and accession number can be found below: National Center for Biotechnology Information (NCBI) BioProject, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA808187.

AUTHOR CONTRIBUTIONS

JZ, JWC, and KH conceived, designed all experiments, and wrote the manuscript. JZ, JD, YZ, and JL performed the experiments. JZ, JJC, and FH analyzed the data. All authors read and approved the final manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.875631/full#supplementary-material

REFERENCES

Beecher, B., and Skinner, D. Z. (2011). Molecular cloning and expression analysis of multiple polyphenol oxidase genes in developing wheat (*Triticum aestivum*) kernels. *J. Cereal Sci.* 53, 371–378. doi: 10.1016/j.jcs.2011.01.015

Beecher, B. S., Carter, A. H., and See, D. R. (2012). Genetic mapping of new seed-expressed polyphenol oxidase genes in wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 124, 1463–1473. doi: 10.1007/s00122-012-1801-2

Chen, J. C., Chiu, M. H., Nie, R. L., Cordell, G. A., and Qiu, S. X. (2005). Cucurbitacins and cucurbitane glycosides: structures and biological activities. *Nat. Prod. Rep.* 22, 386–399. doi: 10.1039/b418841c

Cho, K. M., Ha, T. J., Lee, Y. B., Seo, W. D., Kim, J. Y., Ryu, H. W., et al. (2013). Soluble phenolics and antioxidant properties of soybean (*Glycine max*

L.) cultivars with varying seed coat colours. *J. Funct. Foods* 5, 1065–1076. doi: 10.1016/j.jff.2013.03.002

Cui, J., Yang, Y., Luo, S., Wang, L., Huang, R., Wen, Q., et al. (2020). Whole-genome sequencing provides insights into the genetic diversity and domestication of bitter gourd (*Momordica* spp.). *Hortic. Res.* 7:85. doi: 10.1038/s41438-020-0305-5

Cui, J., Zhou, Y., Zhong, J., Feng, C., Hong, Y., Hu, K., et al. (2022). Genetic diversity among a collection of bitter gourd (*Momordica charantia* L.) cultivars. *Genet. Resour. Crop. Evol.* 69, 729–735. doi: 10.1007/s10722-021-01258-6

Fan, H. C., Wang, J., Potanina, A., and Quake, S. R. (2011). Whole-genome molecular haplotyping of single cells. *Nat. Biotechnol.* 29, 51–57. doi: 10.1038/nbt.1739

Fuerst, E. P., Okubara, P. A., Anderson, J. V., and Morris, C. F. (2014). Polyphenol oxidase as a biochemical seed defense mechanism. *Front. Plant Sci.* 5:689. doi: 10.3389/fpls.2014.00689

Heneidak, S., and Khalik, K. A. (2015). Seed coat diversity in some tribes of Cucurbitaceae and its implications for taxonomy and species identification. *Acta Bot. Bras.* 29, 129–142. doi: 10.1590/0102-33062014abb3705

Hu, Z., Shi, X., Chen, X., Zheng, J., Zhang, A., Wang, H., et al. (2021). Fine-mapping and identification of a candidate gene controlling seed coat color in melon (*Cucumis melo* L. var. *chinensis* Pangalo). *Theor. Appl. Genet.* 135, 803–815. doi: 10.1007/s00122-021-03999-5

Khanna, P., Jain, S. C., Panagariya, A., and Dixit, V. P. (1981). Hypoglycemic activity of polypeptide-p from a plant source. *J. Nat. Prod.* 44, 648–655. doi: 10.1021/np50018a002

Kole, C., Matsumura, H., and Behera, T. K. (2020). "The bitter gourd genome," in *Botanical Description of Bitter Gourd*, eds A. C. Asna, J. Joseph, and K. J. John (Berlin: Springer Nature), 7–31. doi: 10.1007/978-3-030-15062-4

Kole, C., Olukolu, B., Kole, P., Rao, V., Bajpai, A., Suthanthiram, B., et al. (2012). The first genetic map and positions of major fruit trait loci of bitter melon (*Momordica charantia*). *J. Plant Sci. Mol. Breed.* 1:1. doi: 10.7243/2050-2389-1-1

Lee, J., Hwang, Y. S., Kim, S. T., Yoon, W. B., Han, W. Y., Kang, I. K., et al. (2017). Seed coat color and seed weight contribute differential responses of targeted metabolites in soybean seeds. *Food Chem.* 214, 248–258. doi: 10.1016/j.foodchem.2016.07.066

Li, B., Lu, X., Dou, J., Aslam, A., Gao, L., Zhao, S., et al. (2018). Construction of a high-density genetic map and mapping of fruit traits in watermelon (*Citrullus Lanatus* L.) based on whole-genome resequencing. *Int. J. Mol. Sci.* 19:3268. doi: 10.3390/ijms19103268

Li, B., Lu, X., Gebremeskel, H., Zhao, S., He, N., Yuan, P., et al. (2020). Genetic mapping and discovery of the candidate gene for black seed coat color in watermelon (*Citrullus lanatus*). *Front. Plant Sci.* 10:1689. doi: 10.3389/fpls.2019.01689

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., Kristiansen, K., et al. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25, 1966–1967. doi: 10.1093/bioinformatics/btp336

Liu, Z. G. (2011). Study on genetic character of bitter gourd (*Momordica charantia* L.) seed colour. *Seed* 30:109. doi: 10.16590/j.cnki.1001-4705.2011.04.075

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lo, H. Y., Ho, T. Y., Li, C. C., Chen, J. C., Liu, J. J., and Hsiang, C. Y. (2014). A novel insulin receptor-binding protein from *Momordica charantia* enhances glucose uptake and glucose clearance in vitro and in vivo through triggering insulin receptor signaling pathway. *J. Agric. Food Chem.* 62, 8952–8961. doi: 10.1021/jf5002099

Ma, J., Li, C. C., Huang, Y. T., Xie, Y. L., Cheng, L. L., and Wang, J. S. (2021). Fine mapping and candidate gene analysis of seed coat color gene CmSC1 in melon. *Sci. Agric. Sin.* 54, 2167–2178. doi: 10.3864/j.issn.0578-1752.2021.10.012

Marles, M. A., Vandenberg, A., and Bett, K. E. (2008). Polyphenol oxidase activity and differential accumulation of polyphenolics in seed coats of pinto bean (*Phaseolus vulgaris* L.) characterize postharvest color changes. *J. Agric. Food Chem.* 56, 7049–7056. doi: 10.1021/jf8004367

Marles, M. A. S., and Gruber, M. Y. (2004). Histochemical characterisation of unextractable seed coat pigments and quantification of extractable lignin in the Brassicaceae. *J. Sci. Food Agric.* 84, 251–262. doi: 10.1002/jsfa.1621

Matsumura, H., Hsiao, M. C., Lin, Y. P., Toyoda, A., Taniai, N., Tarora, K., et al. (2020). Long-read bitter gourd (*Momordica charantia*) genome and the genomic architecture of nonclassic domestication. *Proc. Natl. Acad. Sci. U.S.A.* 117, 14543–14551. doi: 10.1073/pnas.1921016117

Mavi, K. (2010). The relationship between seed coat color and seed quality in watermelon Crimson sweet. *Hort. Sci.* 37, 62–69. doi: 10.17221/53/2009-hortsci

McKay, J. (1936). Factor interaction in citrullus: seed-coat color, fruit shape and markings show evidence of mendelian inheritance in watermelon crosses. *J. Hered.* 27, 110–112. doi: 10.1093/oxfordjournals.jhered.a104182

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110

Newman, S. M., Eannetta, N. T., Yu, H., Prince, J. P., de Vicente, M. C., Tanksley, S. D., et al. (1993). Organisation of the tomato polyphenol oxidase gene family. *Plant Mol. Biol.* 21, 1035–1051. doi: 10.1007/bf00023601

Paudel, L., Clevenger, J., and McGregor, C. (2019). Chromosomal locations and interactions of four loci associated with seed coat color in watermelon. *Front. Plant Sci.* 10:788. doi: 10.3389/fpls.2019.00788

Poole, C. F., Grimball, P. C., and Porter, D. R. (1941). Inheritance of seed characters in watermelon. *J. Agric. Res.* 63, 433–456.

Porebski, S., Bailey, L. G., and Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* 15, 8–15. doi: 10.1007/BF02772108

Puri, M., Kaur, I., Kanwar, R. K., Gupta, R. C., Chauhan, A., and Kanwar, J. R. (2009). Ribosome inactivating proteins (RIPs) from *Momordica charantia* for anti viral therapy. *Curr. Mol. Med.* 9, 1080–1094. doi: 10.2174/156652409789839071

Schaefer, H., and Renner, S. S. (2010). A three-genome phylogeny of Momordica (Cucurbitaceae) suggests seven returns from dioecy to monoecy and recent long-distance dispersal to Asia. *Mol. Phylogenet. Evol.* 54, 553–560. doi: 10.1016/j.ympev.2009.08.006

Shah, S. S., Hussain, M. I., Aslam, M. K., and Rivera, G. (2014). Natural products; pharmacological importance of family Cucurbitaceae: a brief review. *Mini. Rev. Med. Chem.* 14, 694–705. doi: 10.2174/1389557514666140820113055

Shi, Y., Zhang, M., Shu, Q., Ma, W., Sun, T., Xiang, C., et al. (2021). Genetic mapping and identification of the candidate gene for white seed coat in *Cucurbita maxima*. *Int. J. Mol. Sci.* 22:2972. doi: 10.3390/ijms22062972

Snyder, M. W., Adey, A., Kitzman, J. O., and Shendure, J. (2015). Haplotype-resolved genome sequencing: experimental methods and applications. *Nat. Rev. Genet.* 16, 344–358. doi: 10.1038/nrg3903

Steffens, J. C., Harel, E., and Hunt, M. D. (1994). "Polyphenol oxidase," in *Genetic Engineering of Plant Secondary Metabolism*, eds B. E. Ellis, G. W. Kuroki, and H. A. Stafford (Boston, FL: Springer), 275–312. doi: 10.1007/978-1-4615-2544-8_11

Taketa, S., Matsuki, K., Amano, S., Saisho, D., Himi, E., Shitsukawa, N., et al. (2010). Duplicate polyphenol oxidase genes on barley chromosome 2H and their functional differentiation in the phenol reaction of spikes and grains. *J. Exp. Bot.* 61, 3983–3993. doi: 10.1093/jxb/erq211

Tan, S., Cheng, J., Cui, J., Li, W., and Hu, K. (2013). Genetic analysis on single fruit seed numbers and seed coat color of bitter melon. *China Veg.* 18, 48–52.

Urasaki, N., Takagi, H., Natsume, S., Uemura, A., Taniai, N., Miyagi, N., et al. (2017). Draft genome sequence of bitter gourd (*Momordica charantia*), a vegetable and medicinal plant in tropical and subtropical regions. *DNA Res.* 24, 51–58. doi: 10.1093/dnares/dsw047

Van Ooijen, J. W. (2006). *Joinmap4, Software for Calculation of Genetic Linkage Maps in Experimental Populations*. Wageningen: Kyazma B.V.

Wang, B., Zhang, W., Zhao, J., Wang, F., Fan, L., Wu, Y., et al. (2011). Gene cloning and expression of a novel hypoglycaemic peptide from *Momordica charantia*. *J. Sci. Food Agric.* 91, 2443–2448. doi: 10.1002/jsfa.4485

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164. doi: 10.1093/nar/gkq603

Winters, A., Heywood, S., Farrar, K., Donnison, I., Thomas, A., and Webb, K. J. (2009). Identification of an extensive gene cluster among a family of PPOs in *Trifolium pratense* L. (red clover) using a large insert BAC library. *BMC Plant Biol.* 9:94. doi: 10.1186/1471-2229-9-94

Yoruk, R., and Marshall, M. R. (2003). Physicochemical properties and function of plant polyphenol oxidase: a review. *J. Food Biochem.* 27, 361–422. doi: 10.1111/j.1745-4514.2003.tb00289.x

Yu, C., Svobodova, L., Kucera, V., Vyvadilová, M., Ovesna, J., Dotlacil, L., et al. (2007). Assessment of genetic diversity of yellow-seeded rapeseed (*Brassica*

napus L.) Accessions by AFLP Markers. *Czech J. Genet. Plant Breed.* 43, 105–112. doi: 10.17221/2071-CJGPB

Yu, Y., Tang, T., Qian, Q., Wang, Y., Yan, M., Zeng, D., et al. (2008). Independent losses of function in a polyphenol oxidase in rice: differentiation in grain discoloration between subspecies and the role of positive selection under domestication. *Plant Cell* 20, 2946–2959. doi: 10.1105/tpc.108.06 0426

Zhang, J., and Sun, X. (2021). Recent advances in polyphenol oxidase-mediated plant stress responses. *Phytochemistry* 181:112588. doi: 10.1016/j.phytochem. 2020.112588

Zhu, B. F., Si, L., Wang, Z., Zhou, Y., Zhu, J., Shangguan, Y., et al. (2011). Genetic control of a transition from black to straw-white seed hull in rice domestication. *Plant Physiol.* 155, 1301–1311. doi: 10.1104/pp.110.16 8500

Zhu, F. (2018). Anthocyanins in cereals: composition and health effects. *Food Res. Int.* 109, 232–249. doi: 10.1016/j.foodres.2018. 04.015