Dissecting the genetic control of seed coat color in a RIL population of common bean (*Phaseolus vulgaris* L.)

Carmen García-Fernández1 · Ana Campa1 · Juan Jose Ferreira1

Received: 6 April 2021 / Accepted: 20 July 2021 / Published online: 30 July 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

**Key message** Three genes associated with the seed coat color in a TU/Musica RIL population were located on a genetic map, and two candidate genes proposed to control black seed coat in the TU genotype were characterized.

**Abstract** Seed coat color is an important characteristic of common bean (*Phaseolus vulgaris* L.) associated with the marketability of dry bean cultivars, quality and nutritional characteristics of seed, as well as response to pathogens. In this study, the genetic control of seed coat color in a recombinant inbred line population (175 lines) obtained from the cross ‘TU’ × ‘Musica’ was investigated. Phenotypic segregation fitted 1:1 for white vs. nonwhite, and 3:1 for brown versus black, indicating the involvement of three independent genes, one controlling white color and two (with epistatic interaction) controlling black color. Using a genetic map built with 842 SNPs, the gene responsible for the white seed coat was mapped on the linkage group Pv07, in the position previously described for the *P* gene. For the black seed coat phenotype, two genes were mapped to the beginning of chromosomes Pv06 and Pv08, in the positions estimated for the *V* gene and the complex *C* locus, respectively, by classical studies. The involvement of these two genomic regions was verified through two crosses between three selected RILs exhibiting complementary and dominant inheritance, in which the TU alleles for both genes resulted in a black phenotype. Two genes involved in the anthocyanin biosynthesis pathway were proposed as candidate genes: *Phvul.006G018800* encoding a flavonoid 3′,5′-hydroxylase and *Phvul.008G038400* encoding MYB113 transcription factor. These findings add knowledge to the complex network of genes controlling seed coat color in common bean as well as providing genetic markers to be used in future genetic analysis or plant breeding.

Introduction

Common bean (*Phaseolus vulgaris* L.) is one of the most important grain legumes and is cultivated widely throughout the world (FAOSTAT 2020). Depending on the genotype, beans can be consumed as immature pods (snap or green beans) or as mature seeds after re-hydration and cooking (dry beans). According to Food and Agriculture Organization of the United Nations (FAO) estimates, world production of dry beans for the year 2019 was 28.9 million tons (FAOSTAT 2020). Bean seeds exhibit wide phenotypic diversity, including variations in size, shape, and coat color, leading to many market classes. Within seed coat color variation there is a wide spectrum of solid colors (white, cream, yellow, brown, red, purple, and black) of different intensities, as well as patterns combining different colors (e.g., bicolor, mottled, and spotted) (Voysest 2000).

Classical inheritance studies in common bean have described a complex network of Mendelian genes controlling seed coat color with epistatic interactions among them, through which different allelic gene combinations can produce the same color (Prakken 1970; Bassett 2007). Within this network of genes, there is a primary or basic gene, the *P* gene (*pigment*), whose dominant genotype produces a colored seed coat while the recessive genotype prevents color expression and produces a white seed coat color (Emerson 1909; Kooiman 1931; McClean et al 2018).
The remaining genes included in this network are classically classified as follows.

1. Color-determining genes including both genes that produce color and modifier genes that influence (intensify or darken) the color produced by other genes. Briefly, among the 20 genes included by Basset (2007) in this group are genes such as the complex C locus [cluster of very tightly linked genes, among them, C (sulfur-white or primrose-yellow seed coat), R (dominant red color gene), Prp (produces blacks seed coats in the presence of alleles of other genes), and Gy (controlling the change from pale greenish-yellow to strong greenish-yellow seed coat color)]; the gene B (described as the ‘greenish-brown factor’ that changes chamois to gray-brown and also dark-violet to black seed coat); gene G (controlling the yellow–brown seed color); gene J (giving light yellow–brown or pale ochraceous seed coat); gene V (called the “violet factor” involved in the production of color ranges from pale violet to black depending upon other color genes present); gene Asp (controlling the control brightness of the seed coat); and gene Bic (conferring a dark olive-brown seed coat).

2. Color pattern genes such as the complex C locus and gene Z (controlling constant mottling of the seed coat) the genes T, Bip, and Fib (controlling partly colored seed coat).

Some of these genes, such as J, Z, and the complex C locus, are involved in the control of both the color and the pattern of the seed coat. Several genes controlling seed coat color or molecular markers linked to them have been positioned on bean genetic maps: genes B (Pv02), Z (Pv03), G (Pv04), V (Pv06), P (Pv07), complex C locus (genes C and Gy, Pv08), T (Pv09), J, and Bip (Pv10) (McClean et al. 2002, 2018; Hagerty et al. 2016). Genetic mapping helps to differentiate seed coat color genes, but to clarify this complex network composed of Mendelian genes with epistatic interactions and locating and characterization of the candidate genes in the genome is advisable.

Seed coat color is the result of the accumulation of pigments in the coat such as flavonoids, water-soluble polyphenolic compounds biosynthesized via the phenylpropanoid pathway (Beninger et al. 1999). Flavonoids can be clustered into three main subclasses according to their abundance: flavonols (colorless to pale-yellow pigments), anthocyanins (red to purple-black pigments), and proanthocyanidins, also called condensed tannins (colorless pigments that turn brown; associated with the seed coat post-harvest darkening) (Feenstra 1960; Nesi et al. 2001; Caldas and Blair 2009; Elsadr et al. 2011). Much of the health benefits attributed to the bean seed consumption (neuroprotective, anti-diabetic, anti-inflammatory, and anti-tumor activity, among others) derive from the antioxidant activity of these secondary metabolites (Qiong-Qiong et al. 2018).

Even though the flavonoid biosynthesis pathway has been widely elucidated in different species, such as Arabidopsis thaliana and Glycine max, information available in common bean is more limited (Caldas and Blair 2009; Reinprecht et al. 2013; Myers et al. 2019), since to date few candidate genes have been physically positioned in the genome and the implication of nucleotide allelic variations in determining seed coat color has been poorly evaluated (Reinprecht et al. 2013; McClean et al. 2018; Islam et al. 2020). Although certain species-specific differences have been described, the flavonoid biosynthesis pathway presents a high degree of conservation among species at both the structural and functional genetic levels (Chaves-Silva et al. 2018; Petroni and Tonelli 2011). At the molecular level, the flavonoid biosynthetic pathway is organized into (1) structural genes encoding enzymes that catalyze each step of the pathway and (2) regulatory genes responsible for controlling the function of structural genes at the transcriptional level. In turn, structural genes are divided into early (EBGs) and late (LBGs) biosynthetic genes (Dubos et al. 2010). Whereas EBGs encode enzymes that catalyze the biosynthesis of flavones, flavonols, and other flavonoid compounds, LBGs encode enzymes that specifically lead to anthocyanin and proanthocyanidin biosynthesis. The regulatory genes include three families of transcription factors: R2R3-MYBs, bHLH (basic helix–loop–helix), and WD40 repeat transcription factors. According to the regulation pattern described in A. thaliana, independent R2R3-MYBs are responsible for the activation of EBGs, while activation of LBGs requires the formation of ternary complexes through the coupling of R2R3-MYB, bHLH, and WD40 transcription factors (MBW complexes) (Petroni and Tonelli 2011).

In addition to its direct influence on the marketability of dry bean varieties and nutritional characteristics (Elias et al. 2006; Rodriguez-Madrera et al. 2020), seed coat color is also linked to other seed characters. In legume species, the presence of a colored seed coat is often correlated with a lower permeability to water (Dübbern and Marcos-Filho 2001). This is consistent with bean seeds having a white coat requiring less cooking time than those with a colored coat, especially a black coat (Cichy et al. 2015). In contrast, the lower permeability attributed to the black seed coat leads to higher longevity by reducing the risk of deterioration by imbibition (Dübbern and Marcos-Filho 2001). Associations between colored seed coat and emergence, seed yield, and weight have also been reported in common bean (Deakin 1974; McClean et al. 2018). Likewise, the coloration of the seed coat has been correlated with greater resistance to microorganism attacks. Specifically, bean seeds with colored coats are associated with resistance to pathogens such as
potyviruses (Kyle and Dickson 1988) and Pythium ultimum (Campa et al. 2010).

Understanding the inheritance of seed coat color is fundamental to breeding programs involving seed coat color or its pleiotropic effects on other characters. The reference genome for P. vulgaris (Schmutz et al. 2014) provides a framework for the fine location of major genes conferring important traits and for identifying candidate genes. The main aim of this work was to investigate the genetic and physical positions of genes controlling seed coat color in the genotype TU through a forward genetic analysis using a recombinant inbred line (RIL) population derived from the cross ‘TU’ × ‘Musica’.

Materials and methods

Genetic mapping population

A mapping population (TUM population) of 175 recombinant inbred lines (RILs, F_{6:7}) derived from the cross ‘TU’ (female parent) and ‘Musica’ (male parent), was obtained by single seed descent. The parent ‘TU’ is a dry cultivar well known for its resistance to anthracnose disease with small (30.0 g/100 seeds) and black seed coat color (Fig. 1). ‘Musica’ is a snap bean with a medium-sized seed (45.6 g/100 seeds) and a white seed coat color (Fig. 1). Both parents have indeterminate growth habits and are related to the Mesoamerican gene pool (Campa et al. 2018). Although both parental lines have a majority Mesoamerican genetic background, they present a different Andean introgression level (‘TU’ estimated Andean introgression level around 36%; ‘Musica’ estimated Andean introgression level less than 7%) (Campa et al. 2018).

The TUM population was phenotyped with the parental lines in a greenhouse using a randomized complete block with a replicate per line at the Regional Agrifood Research and Development Service (SERIDA), Villaviciosa, Asturias, Spain (43° 29′ 01″ N, 5° 26′ 11″ W; elevation 6.5 m) during two consecutive seasons (2018–2019). Each plot has a single 1-m row with 8–10 plants per recombinant inbred line. The seed coat color was qualitatively recorded by visual examination of the seeds derived from each line and classified considering the seed coat color of the two parents.

Genotyping and SNP calling

Fresh young trifoliate leaves from each one of the 175 F_{6} lines and the parental lines were collected, frozen, and homogenized into a fine powder for isolating Genomic DNA using the CTAB method (Doyle and Doyle 1990). DNA integrity was analyzed by electrophoresis on 1% agarose gel, stained with RedSafe (iNtRON Biotechnology, Gyunggi-Do, Korea) and visualized under ultraviolet light. Final DNA yield was quantified by photometric absorbance at 280/260 nm using a Biomate 3 ultraviolet–visible spectrophotometer (Thermo Scientific, Waltham, MA, USA). DNA restriction analysis was performed on 10% of samples selected randomly by digesting with HindIII and running on 1% agarose gel. DNA samples were preserved at –80 °C until use.

Genotyping-by-sequencing (GBS), as described by Elshire et al. (2011), was carried out at BGI-Tech (Copenhagen, Denmark) using the ApeKI restriction enzyme. A GBS sequencing library was prepared by ligating the digested DNA to unique nucleotide adapters (barcodes) followed by PCR with flow-cell attachment site tagged primers. Sequencing was performed using Illumina HiSeq4000 with 100× paired-ends. The sequencing reads from different genotypes were deconvoluted using the barcodes and aligned to the P. vulgaris L. v2.1 (genotype G19833) reference genome (available at https://phytozome.jgi.doe.gov/pz/portal.html), using the Burrow Wheelers Alignment tool (Li and Durbin 2009). Single nucleotide polymorphisms (SNPs) were
named according to physical position in the bean genome: chromosome and genomic position (bp). SNP markers were called using the GBS pipeline implemented in TASSEL v5.2.39 software (Bradbury et al. 2007).

Some quality control filters were applied to avoid including variants with poor quality data. SNPs with a missing rate higher than 10% and heterozygous genotypes were removed using Tassel v5.2.39 software (Bradbury et al. 2007).

**Construction of genetic linkage map**

SNPs were filtered considering the physical distance among them and the segregation in the RIL population. Those SNP markers with a physical distance less than 0.1 Mb from a neighboring marker were removed. Markers showing significant deviation from the expected Mendelian segregation ratio (1:1), as determined by Chi-square test ($p < 0.05$), were not considered. Linkage analysis was conducted with the R/OneMap mapping package (Margarido et al. 2007; R Core Team 2020). SNPs were clustered into linkage groups (LGs) using a log of likelihood ratio threshold (LOD) of 4.0 and a maximum recombination fraction (RF) of 0.35. Loci order was estimated based on rapid chain delineation algorithm (Doerge 1996) and ripple analyses. Distance between ordered loci (in cM) was estimated using the Kosambi mapping function. To avoid redundant information, SNP blocks without observed recombination events were simplified and only one tag SNP was used for map generation. Graphical visualization of the LGs was achieved by MapChart 2.32 software (Voorrips 2002). The linkage groups were named according to the method described by Pedrosa-Harand et al. (2008).

**Genetic analysis**

The goodness of fit of observed to expected ratios was tested using Chi-square tests ($\chi^2$) considering $\alpha \leq 0.05$. To map the gene(s) involved in the control of the seed coat color, a genetic analysis was performed as follows:

1. In cases in which the observed segregation fit to a single gene, it was included in the genetic map.
2. When the observed segregation suggested the implication of more than one gene, contingency Chi-square tests were conducted for the joint segregation of seed coat color and the markers included in the genetic map. A significant deviation from random segregation suggests the association of the SNPs tagging a specific region with the trait. Significance thresholds were determinate using Bonferroni correction considering $\alpha = 0.05$ (Bonferroni 1936). Finally, complementary crosses were designed to validate the implication of candidate chromosome regions in the control of seed coat color. Recombinant inbred lines (RILs) were selected according to seed coat color phenotype and genotype (TU or Musica) for underlying markers tagging the candidate chromosome regions. These lines were crossed to obtain F1 seeds, which were planted and self-pollinated to evaluate the seed coat color of the F1 genotypes. Given the maternal inheritance of this character, the flower color phenotype in the F1 plants was used as a control for successful manual cross-pollination.

**Candidate gene identification and sequencing**

Genes underlying significant genomic regions associated with seed coat color were identified using the Phytomine tool in Phytosome v12 (https://phytosome.jgi.doe.gov/phytmine/begin.do). Putative candidate genes were selected based on: (1) their involvement in the flavonoid biosynthesis pathway and (2) gene–trait associations previously described in the species itself (McClean et al. 2018) or other species (Herniter et al. 2018).

Complete genomic sequences of the selected putative candidate genes were obtained from the Phytosome v12 database. To resequence the candidate genes, sets of gene-specific primers pairs were designed throughout their genomic sequences, seeking the maximum coverage of the coding region. Primer design was carried out using Batch Primer3 (UC Davis Server) (You et al. 2008), and potential secondary structure and dimer formation were tested using Oligo Analyzer v3.1 (http://eu.idtdna.com/oligalyzer/Applications/OligoAnalyzer/). DNA from parental lines (‘TU’ and ‘Musica’), four inbred lines (TUM016, TUM051, TUM101, and TUM170), and line ‘G19833’ (positive control) were used for resequencing. PCRs were performed on a Verity Thermal Cycler (Applied Biosystems, Life Technologies, CA, USA) using a TaKaRa LA Taq ® kit (Clontech Laboratories, Inc., Mountain View, CA) according to the manufacturer’s protocol. For each reaction, 5 μL of genomic DNA (50 ng) was mixed with 5 μL of LA PCR buffer II (10x), 5 μL of MgCl2 (25 mM), 8 μL of dNTP mixture (2.5 mM each), 3 μL of each primer (10 μM), 0.5 μL of TaKaRa LA Taq (5 U/μL), and sterile purified water to a final volume of 50 μL. PCR amplification consisted of an initial denaturation step of 1 min at 94 °C, followed by 35 cycles of denaturation at 98 °C for 10 s and gene-specific annealing temperature for 5 min, and a final extension step at 72 °C for 10 min. PCR products were run on a 2% agarose gel stained with RedSafe (iNtRON Biotechnology, Gyunggi-Do, Korea) and visualized under ultraviolet light. Unique PCR bands were isolated and purified using Ilistra™ GFX™ PCR DNA and Gel Band Purification Kits (GE Healthcare Life Sciences, Buckinghamshire, UK). Purified PCR products were sent for sequencing using the same primers.
for sequencing using the Sanger method to the sequencing laboratory of the University of Oviedo (Asturias, Spain). The sequences obtained were analyzed by alignment with reference sequences using the BioEdit sequence alignment editor v7.0.5.3 (Hall 1999, 2011). Polymorphisms identified were named according to their corresponding physical position (chromosome and bp) on the reference sequence. Impacts of the polymorphisms on the biological function of the proteins were predicted using Protein Variation Effect Analyzer (PROVEAN) software (www.provean.jcvi.org).

Results

Genotyping and genetic map construction

Approximately 97.93 Gb of high-quality raw sequence data were obtained by GBS, and a total of 12,532 SNPs were called. SNPs supplied by GBS analysis were filtered based on data completeness (> 90%), and SNPs showing heterozygous genotypes were removed. A total of 8473 SNPs distributed across the eleven bean chromosomes were obtained from the TU/Musica (TUM) population. Among them, 635 showed distorted segregation ($\chi^2$, $p < 0.01$) and 6777 had a physical intermarker distance of less than 0.1 Mb ($\approx$ 80%), so they were removed. Finally, 1061 SNPs were used to construct the genetic linkage map. The final version of the TUM linkage map contained 842 highly informative SNPs mapped to 13 linkage groups (LGs) corresponding to the 11 bean chromosomes, with chromosomes Pv04 and Pv05 divided into two blocks (Table S1, Figure S1). A gap in LG Pv04 was located at the end of the chromosome affecting SNPs located in the physical position 37.8–46.1 Mb. In Pv05, a gap was located near the beginning of the chromosome between the SNPs with physical positions of 3.0 Mb and 7.1 Mb. The gaps found in chromosomes Pv04 and Pv05 were due to the limited number of SNPs detected in those regions. In addition, LGs Pv02, Pv07, and Pv09 each included one gap of more than 15 cM. The number of SNPs clustered per LG varied from 132 (Pv03) to 45 (Pv04). The total genetic length of the map was 2188.89 cM with an average intermarker distance of 3.45 cM. The length of individual LGs ranged from 364.60 cM (Pv03) to 34.51 cM (Pv04).

Seed coat color segregation

The lines include in the RIL TUM population were classified into three phenotypic groups according to seed coat color: white, like parental Musica ($N$ = 89); black, like parental TU ($N$ = 30); and brown ($N$ = 56) (no parental phenotype) (see Fig. 1). Although for genetic analysis all seeds with brown coat color were clustered into a single phenotypic class, variations in color intensity among them were nevertheless observed. Segregation of seed coat phenotypes was initially analyzed considering two phenotypic classes: white vs. nonwhite (including brown and black phenotypes). The observed segregation fitted an expected Mendelian ratio of 1 (89): 1 (86) ($\chi^2$ = 0.05; $p$ = 0.82), suggesting genetic control based on a single gene for the white seed coat trait. For the colored subpopulation (brown vs. black), a fit to a 3 (56):1 (30) ratio ($\chi^2$ = 4.48; $p$ = 0.03) was observed, suggesting that black seed coat color is determined by two independent genes with epistatic effects. According to the segregation patterns obtained, the most consistent hypothesis is that three independent genes control seed coat color in the TUM population.

Mapping of seed coat color genes

The white seed coat trait color, which segregated according to control by a single gene, was mapped on chromosome Pv07 between SNPs S07_28535059 (291.34 cM) and S07_28968582 (298.58 cM); the recombination fraction with the first flanking marker was 0.10 (LOD = 35.21). These genetic positions correspond to the physical interval between 28,535,059 and 28,968,582 bp on chromosome Pv07 according to the P. vulgaris L. v2.1 (genotype G19833) reference genome.

The phenotypic segregation ratio observed for colored seed coats suggested the involvement of two independent genes. Contingency Chi-square tests revealed two genomic regions significantly associated with this trait, located on chromosomes Pv06 and Pv08. The region detected on Pv06 was tagged with nine SNPs located between positions 1.41 and 16.04 Mb (Table S2). Recombination events among the SNP genotypes and seed coat phenotype in the colored subpopulation allowed delimiting the control of seed coat color within the region flanked by markers S06_2197326 and S06_5174714 (see Fig. 2). Enrichment of the flanking region with six additional SNPs (filtered during the map construction) narrowed down the candidate region to 2.07 Mb flanked by markers S06_2478502 and S06_4566963 (see Fig. 2), corresponding to the segment of chromosome Pv06 located between the physical positions 2,478,502 and 4,566,963 bp. On chromosome Pv08, the region identified was tagged by a block of six SNPs located at the beginning of the chromosome between 3.18 and 4.55 Mb (Table S3). Recombination events detected in six individual RILs allowed to narrow the genetic region to an interval of 0.25 Mb between markers S08_3183715 and S08_3430092 (Fig. 2), whose physical position corresponds to 3,183,715 and 3,430,092 bp, respectively.

Finally, investigating the seed coat color phenotype and genotype in these two regions in the colored subpopulation, it was observed that only the presence of the TU genotype in both candidate chromosomal regions gives rise to a black
seed coat color, while a Musica genotype in either region or both leads to a brown seed coat color (Fig. 2). This finding is consistent with the inheritance observed, suggesting two epistatic and independent genes controlling black seed color in this RIL population (segregation 3 brown: 1 black).

Genetic complementation

To verify the presence of genes controlling seed coat color at the two candidate chromosomal regions identified, two complementary crosses were performed using three selected inbred lines (TUM051, TUM101, and TUM170). Selection of the lines was based on two criteria: (1) seed coat color phenotype (brown) and (2) marker genotype in the candidate regions (TU or Musica) (Fig. 2). Line TUM051 had white flowers, brown seed coat phenotype, and carried the Musica genotype for the markers of the candidate region on chromosome Pv06 (SNP S06_2478502, S06_4221741, and S06_4566963) and the TU genotype for markers of the candidate region on chromosome Pv08 (SNP S08_3183715 and S08_3430092). The recombinant lines TUM101 and TUM170 presented purple flowers, brown seed coat phenotype, and carried the TU genotype for the markers of the candidate chromosomal region on Pv06 (SNP S06_2478502, S06_4221741, and S06_4566963), and the Musica genotype for the markers of the candidate chromosomal region on Pv08 (SNP S08_3183715 and S08_3430092). Line TUM051 was used as the female parent and lines TUM101 and TUM170 were used as male parents in these complementary crosses. All F1 seeds obtained (heterozygous genotype) showed the brown seed coat phenotype because seed coat color is determined by the maternal genotype, meaning the phenotype linked to the heterozygous genotype is expressed in the next generation. A total of 14 F1 plants derived from the crosses TUM051 x TUM101 (9 plants) and TUM051 x TUM170 (5 plants) were self-pollinated. All F1 plants presented purple flowers, supporting their origin from a manual cross-pollination, and produced F2 seeds with a black seed coat phenotype (Figure S2). This result shows that the black allele has a dominant effect over the brown allele.

Therefore, the inheritance pattern observed was consistent with the hypothesis that two dominant and complementary genes control black seed coat color in the TU genotype since only the presence of the TU genotype in both chromosomal regions leads to the black color phenotype.

Candidate genes identification

The three chromosomal regions involved in the genetic control of seed coat color in the TUM population were scanned in silico for underlying genes in the P. vulgaris L. v2.1 (genotype G19833) reference genome. Lists of genes identified are provided in Tables S4, S5, and S6. The physical interval on Pv07 linked to the white seed coat color contained a total of 38 genes, eight of them without functional annotation. Among the set of putative candidate genes, it was found two tandem genes, Phvul.007G171333 and Phvul.007G171466, located at 0.20 Mb and 0.19 Mb from marker S07_28968582, respectively. Both genes are physical components of the dominant-acting P (pigment) gene whose implication in the white seed coat phenotype was previously described by McClean et al. (2018).

The two chromosomal regions associated with the presence of black seed coat color harbored a total of 75 annotated genes. For the region defined on Pv06, the most interesting candidate genes were Phvul.06G018800, located close to marker S06_4221741 at a physical distance of approximately 0.11 Mb. The gene Phvul.06G018800 encodes a flavonoid 3′5′ hydroxylase (F3′5′H), a cytochrome P450 enzyme belonging to the CY75A subfamily that plays a key role in the anthocyanin biosynthesis pathway (see Figure S3).
Within the delimited region on Pv08, very close to the flanking marker S08_3188683 (at 4968 bp), it was found a cluster of three genes (Phvul.008G038400, Phvul.008G038500, and Phvul.008G038600) encoding MYB-113 transcription factors with a functional annotation implicating them in the regulation of anthocyanin metabolism. Alignment of multiple amino acid sequences revealed a similarity of between 40 and 53% among these MYB genes. BlastP search in the NCBI database (http://www.ncbi.nlm.nih.gov/ncbi.nlm.nih.gov) based on the protein sequences of vigun05g039500, a candidate gene for black seed coat color in cowpea (Vigna unguiculata [L.] walp), assigned the maximum score (312; 5e-109 e-value) to Phvul.008g038400, with a query coverage of 97% and percentage of identity of 69.10%. Based on these results, Phvul.006G018800 and Phvul.008G038400 were initially selected as the most likely candidate genes for the control of black seed coat color in the TU genotype.

Exploring candidate gene sequences for black seed coat color

A partial genomic nucleotide sequence of the two proposed candidate genes was obtained and allelic variants were identified using a set of primers (Table S7) designed using G19833 as reference sequences.

Gene Phvul.006G018800 The genomic sequence of Phvul.006G018800 spans 6787 bp and is organized into three exons and two introns flanked by 5' and 3' UTR domains in the bean reference genome (Fig. 2, Table S8). PCR fragments (not overlapping) resequenced for Phvul.006G18800 for the genotypes TU and Musica had a length of 4312 bp and 3784 bp respectively (Fig. 2, Table S8). The shorter sequence obtained from Musica was due to failed amplification of two fragments (sequencing gaps), one located in intron 2 and the other in the 3' UTR domain, according to the reference gene sequence. Nevertheless, the resequenced fragments included a full-length coding sequence composed of three exons, similar to the reference gene, in both parental lines (Fig. 3; Table S8). Alignment of sequences obtained from the genotypes G19833, TU, and Musica revealed more than one hundred polymorphic positions, among which are included a single base pair deletion in exon 3 at position 6125 in the TU and Musica genotypes (Figure S4). Nevertheless, the majority of the polymorphisms detected resulted in variations between G19833 and the parental lines. TU and Musica showed a high level of genomic nucleotide sequence conservation (96.5% identity), with only eight nucleotide differences identified, out of which, based on the reference gene structure, six were located in intron regions, one in exon 3, and one in 3' UTR domain. The coding variant

![Diagrammatic representation of resequencing coverage of the genes Phvul.006G018800 and Phvul.008G038400 using G19833 as reference sequence. The position of primers used for resequencing and primers used as gene markers (BSC06 and BSC08) are also indicated.](image-url)
(SNP06_6032G/A) involved a single base pair substitution that leading to a homologous amino acid replacement from serine (TU) to asparagine (Musica) at position 327 aa of the predicted protein sequence. PROVEAN analysis suggested a neutral functional effect associated with this amino acid change (score = −1.087; cut off = −2.5). The polymorphism in the 3’ UTR involved a deletion of 14 bp (DEL06_6296-6309) located close to the beginning of the region, which affects the binding site of the primer Phvul.006G018800_F11.

Gene Phvul.008G038400 The genomic sequence for Phvul.008G038400 is 2,275 nucleotides in length and shares a similar structure with Phvul.006G018800, consisting of three exons and two introns flanked by the 5’ and 3’ UTR domains. To avoid the interference of non-specific amplification resulting from the presence of regions highly conserved among different MYB genes, resequencing of Phvul.008G038400 involved a single amplicon using the flanking primers (Phvul.008G038400_F1 and Phvul.008G038400_R4) designed from the intergenic regions (see Fig. 3). This strategy allowed isolating the complete sequence of Phvul.008G038400 for the lines G19833 (used as a positive control) and the TU genotype. The length of the partial sequence obtained for the TU genotype was 1,591 bp, covering the complete gene sequence except for a gap in the second intron (Fig. 2; Table S8). It was not possible to amplify the Musica genotype using the same flanking primers. The longest amplicon isolated for this line was extended from the 5’ intergenic (Phvul.008G038400_E1_F) region to exon 2 (Phvul.008G038400_E2_R), and the sequence obtained was only 392 bp in length (Fig. 2; Table S8). Multiple sequence alignment revealed high levels of polymorphism (Figure S5). Around 60% of the polymorphisms identified represented allelic variations between the two parental lines, distributed throughout the 5’UTR (6), exon 1 (16), intron 1 (29), and exon 2 (1) (partially resequenced) based on the reference gene structure (Figure S5).

Gene marker genotyping

The observed allelic variation was used to develop markers tagging both candidate genes. The differences in amplification [presence (TU)/absence (Musica)] detected for the 3’UTR of Phvul.006G188000 and for the amplicon that included the complete gene sequence of Phvul.008G038400 were selected as direct gene markers by tagging allelic variations existing between the two genotypes (Fig. 4, Table S7). The gene markers were called BSC06 and BSC08, respectively, in reference to the Black Seed Coat trait and their chromosomal location. Both markers co-segregated with the SNPs tagging in the respective regions (see Fig. 2). BSC06 co-segregated with the SNP06_4221741 except in one line, whereas BSC08 co-segregated with SNP07_3183715 except in two lines. All TUM lines with black seed coat phenotype showed TU genotype for both gene markers, whereas lines with a brown seed coat phenotype presented Musica genotype for at least one of the two markers (Fig. 4).

Discussion

A complex network of genes is implicated in the genetic control of seed coat color in common bean (Bassett 2007). In this work, the genetic control of the seed coat color, as well as the location in the genetic and physical map, was investigated in an F6:7 TUM RIL population obtained from the cross ‘TU’ × ‘Musica’. Genotyping of the TUM RIL population was based on 12,532 SNPs obtained from a GBS approach. Filtering criteria applied for strict selection of high-quality SNPs allowed the construction of a genetic linkage map composed of 842 SNPs markers that covering the 11 LGs. Two major gaps were detected on the linkage groups LGs Pv04 and Pv05, indicating that these regions were highly conserved between the two parental lines. This finding is in agreement with the common Mesoamerican origin of their genetic background reported by Campa et al (2018). ‘Musica’ is a snap bean derived from breeding
programs, which is included in the Mesoamerican gene pool, whereas ‘TU’ is a dry bean included in a recombinant group between Mesoamerican and Andean gene pools. Consequently, less polymorphism is expected for this population than for populations derived from crosses between lines belonging to different gene pools (Mesoamerican vs Andean).

In this work, the inheritance of the seed coat color was evaluated as a qualitative character since different phenotypic classes were clearly established in the recombinant lines. The results obtained for the TUM RIL population indicated the involvement of three independent genes with epistatic interaction in the control of the three seed coat color phenotypic classes observed: white, brown, and black. White seed coat color is regulated by a single gene located in a physical interval between 28,535,059 and 28,968,582 bp on chromosome Pv07. Among the genes in this interval are Phvul.007G171333 and Phvul.007G171466, two physical components of a single-gene model characterized by McClean et al. (2018) as the \( P \) gene in common bean. The \( P \) gene encodes a TT8 bHLH transcription factor, a key component of the MBW complex that regulates activation of the flavonoid biosynthetic pathway (Mclean et al. 2018).

This annotation is consistent with the classical concept that defined \( P \) as a dominant basic gene that does not produce any color by itself, but whose presence is necessary to produce seeds with colored coats (Prakken 1970, 1972). Recessive alleles of the \( P \) gene \( p \) shut down the flavonoid biosynthesis pathway at a primary level; consequently, a biochemical profile characterized by the absence of flavonols and anthocyanins is responsible for the completely white seed coat phenotype (Rodríguez-Madrera et al. 2020). McClean et al. (2018) identified 10 race-specific \( P \) alleles involved in the expression of the white seed coat phenotype. These alleles are characterized by nucleotide variations resulting in deletions located in the functional protein domain, deletions that truncates the translation product prematurely, and non-synonymous coding SNPs responsible for substitution of essential amino acid residues that compromise the protein function.

Concerning to black seed coat phenotype, the results indicated that is controlled by two independent genes located at the beginning of the chromosomes Pv06 and Pv08, with epistatic interaction. Physical delimitation within the bean genome allowed to locate the first gene between positions 2,478,502 and 4,566,963 bp of chromosome Pv06, and the second gene between positions 3,183,715 and 3,430,092 bp of chromosome Pv08. This analysis also revealed that the black seed coat phenotype results from the presence of a \( P \) genotype in both chromosomal regions. This complementarity of the chromosomal regions identified was evidenced by the results of crosses between three inbred lines showing a brown seed coat color phenotype and complementary genotypes. The complementarity test also determined that black seed coat color is controlled by dominant alleles at both loci (black > brown). McClean et al. (2002) positioned the marker OD12,000 (physical position, 9.4 Mbp) at the beginning of chromosome Pv06 by genetic linkage to the \( V \) gene, also known as the violet factor (Prakken 1970). Classical genetic studies described the \( V \) gene as a color enhancer or modifier gene which has multiple alleles with pleiotropic effects on color in flowers and seed coats (Prakken 1970, 1974; Bassett 1997; Bassett and Miklas 2007). Specifically, the \([V]\) allele produces seed coats in which color ranges from pale violet to black depending upon other color genes present (Lamprecht 1932; Prakken 1934, 1972), whereas the \([v]\) allele (in the presence of other color genes like T P C J G B) produces mineral brown seed coat (Lamprecht 1935; Bassett and Miklas 2007). Beninger et al. (1999, 2000) affirmed that anthocyanin production is dependent on the \( V \) gene, since the presence of dominant genotypes \([V]\) leads to anthocyanin accumulation, whereas recessive genotypes \([v]\) lead to flavonol accumulation. McClean et al. (2002) also mapped the complex \( C \) locus to the beginning of chromosome Pv08 through the markers OW17S500 (physical position, 3.2 Mbp) and OAP2,700 (physical position, 9.6 Mbp) linked to the \( C \) and \( GY \) genes, respectively. Although most studies correlate the function of the complex \( C \) locus with the control of a wide diversity of patterns in common bean seed coats (Prakken 1974), Bassett (1994) established a tight linkage of the black seed coat color trait and the complex \( C \) locus in the presence of other dominant alleles of color genes, including \( P \) and \( V \). Thus, the genetic location of the \( V \) gene and the complex \( C \) locus, allows to speculate on their possible connection with the Pv06 and Pv08 chromosome regions identified in the present study, respectively.

In the tagged region of chromosome Pv06, it was found the annotated gene Phvul0006G18800, one EBG gene encoding an F3′5′H enzyme, which plays a crucial role in the branch of the flavonoid biosynthesis pathway leading to delphinidin-based production of anthocyanins (blue/purple pigments) (Figure S3). This finding supports black seed coat phenotypes resulting from a biochemical composition profile characterized by high anthocyanin accumulation (Takeoka et al. 1997; Díaz et al. 2010; Rodríguez-Madrera et al. 2020). The biosynthetic pathway leading to delphinidin-based anthocyanins appears to be the main way of producing black seed coat color in common bean (Takeoka et al. 1997; Díaz et al. 2010; Rodríguez-Madrera et al. 2020), although black color could also be produced by the accumulation of different anthocyanin classes (Rodríguez-Madrera et al. 2020). The resequencing of Phvul.006G18800 revealed eight polymorphic positions between the two parental genotypes, highlighting a non-synonymous coding SNP (SNP_6032G/A) and a deletion of 14 bp (DEL06_6296-6309), located in exon 3 and 3′UTR, respectively, according to the gene structure.
established for G19833. Although expression and functional studies are required to determine the possible implications of these polymorphic positions, everything seems to indicate that the presence of recessive Musica alleles in \textit{Phvul.006G18800} \textit{(musica)} would inhibit expression of the black seed coat phenotype by blocking the main anthocyanin biosynthesis pathway in the TUM lines (Figure S3).

Therefore, the TU and Musica genotype are consistent with genotypic implications described for the dominant \( [V = TU] \) and recessive \( [v = musica] \) alleles of the \( V \) gene (Beininger et al. 1999, 2000). The involvement of \( F3′5′H \) in the control of black seed coat coloration has also been described in soybean (\textit{Glycine max} and \textit{Glycine soja}) by means of the \( WI \) gene (Zabala and Vodkin 2007; Guo and Qiu 2013). Furthermore, the comparison of the genomic maps of soybean and common bean showed that the genomic position to which the \( WI \) gene was mapped (\textit{Gm13}) is syntenic to the region of common bean where the \( F3′5′H \) gene is located at the top of chromosome Pv06 (Reinprecht et al. 2013).

Within the \( P\nu08 \) chromosomal region identified, it was found a cluster of three genes members of the \textit{MYB} family of transcription factors (\textit{MYB113}) involved in the regulation of anthocyanin biosynthesis (Gonzalez et al. 2008). Within this \textit{MYB113} cluster is included \textit{Phvul.008G038400}, an ortholog gene of \textit{Vigun05G039500}, proposed as a candidate gene for black seed coat color in cowpea (Herniter et al. 2018). In \textit{A. thaliana}, the \textit{MYB113} gene is included within the subgroup of \textit{R2R3-MYBs} in charge of regulating the expression of enzymes involved in the last steps of the anthocyanin biosynthesis pathway in vegetative tissues (\textit{LBG} gene) (Gonzalez et al. 2008). Thus, \textit{Phvul.008G038400} would be acting on the final key step leading to anthocyanin accumulation and, consequently, towards the black seed coat phenotype in the TUM population. In addition, this late regulation by \textit{Phvul.008G038400} is consistent with the observation that within the brown phenotypic class, darker brown seed coats corresponded to TUM lines carrying the TU genotype for \textit{Phvul.006G18800} and the Musica genotype for \textit{Phvul.008G038400}. On the other hand, the identification of a cluster of \textit{MYB} genes involved in anthocyanin biosynthesis agrees with the classical interpretation that describes complex \( C \) locus as a group of closed linked genes involved in the determination of seed coat color phenotype (Prakken 1974).

The nucleotide sequences obtained from the resequencing of \textit{Phvul.008G038400} showed a low conservation degree between the TU and Musica genotypes, and also concerning the reference sequence of G19833. The main difference was that while for the TU genotype a complete gene sequence was amplified, only a partial amplification (from \( 5′ \) intergenic region up to the second exon) was achieved for the Musica genotype. This is similar to results obtained by Herniter et al. (2018) for \textit{Vigun05G039500} in cowpea, who described consistent, successful amplification of a fragment of the third exon in all lines with black seed coats and a similarly consistent failure to amplify this fragment in lines with brown seed coats. These authors described a deletion of approximately 41 kb that involving \textit{Vigun05G039500} as well as other nearby \textit{MYB} genes and whose origin could be due to non-allelic homologous recombination, an unequal crossover between highly similar DNA sequences. Similar to \textit{Vigun05G039500}, \textit{Phvul.008G038400} belongs to a gene cluster of five \textit{MYB113}-related genes; two genes are located downstream and included among the genes within the region identified (\textit{Phvul.008G038500} and \textit{Phvul.008G038600}), and two genes are located upstream, outside of the associated chromosomal region (\textit{Phvul.008G038200} and \textit{Phvul.008G038200}). According to the results obtained and the high similarity between common bean and cowpea, the absence of amplification in the Musica genotype could correspond to the presence of a deletion similar to that described in cowpea. In addition, since \textit{Phvul.008G038400} is part of a cluster of genes, the detected polymorphism could act not only as a specific marker of the gene but also of the \textit{MYB} cluster.

On the other hand, from the polymorphism detected between the TU and Musica genotypes for the candidate genes \textit{Phvul.006G011880} and \textit{Phvul.008G038400}, two user-friendly markers (BSC06 and BSC08, respectively) were developed. BSC06 and BSC08, constitute gene-specific markers, which are advantageous over linked markers since possible recombination processes between marker and gene are avoided, and thus the selection of false positives in marker-assisted selection in breeding programs.

Thus, in the TU \( \times \) Musica genomic background, three genes, \textit{Phvul.007G171333-Phvul.007G171466} (components of the same gene), \textit{Phvul.006G018800}, and \textit{Phvul.008G038400}, which appear to be related to the classical genes \( P \), \( V \), and the complex \( C \) locus, respectively, would be regulated the expression of the different seed coat phenotypes observed. The presence of the recessive genotype of the \( P \) gene \( [p^{musica}] \) causes any genotypic combination to confer a white seed coat phenotype. However, in the presence of the \( P \) dominant allele \( [P^{TU}] \), the interaction between the \( V \) \( [V^{TU}] \) and the complex \( C \) \( [C^{TU}] \) alleles, leads to a black seed coat phenotype \( [p^{TU} \nu^{TU} C^{TU}] \), whereas the remaining gene combinations give rise to a brown seed coat phenotype \( [P^{TU} \nu^{TU} C^{musica}] \) or \( [P^{TU} \nu^{TUC^{musica}}] \).

**Conclusions**

This work verified the genetic and genomic position of the \( P \) gene in common bean. Likewise, it was determined the genetic and genomic position of two dominant and complementary genes controlling the black seed color in the
TU genotype at the beginning of chromosomes Pv06 and Pv08. Finally, this work provides two user-friendly markers (BSC06 and BSC08) to assist future genetic analysis or plant breeding.

This finding adds knowledge for elucidating the complex network of genes involved in the control of seed coat color determination in common bean. Nevertheless, additional forward genetic analysis should be carried out to locate the genomic position of the remaining seed coat color loci. Likewise, analysis of gene expression in the developing seed can be another alternative way to identify/validate new candidate genes controlling seed coat color in common bean.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-021-03922-y.

Acknowledgements The authors thank M. Bueno, J.A. Poladura, and F. Diaz for their technical assistance.

Author contributions CGF performed the phenotyping, genomic, and genetic analysis and prepared the manuscript. AC performed the genotyping and genetic analysis. JJF conceived the work and developed the mapping population. All authors have read and approved the last version.

Funding This work was supported in part by the Grant AGL2017-87050-R of the AEI-Spanish Government. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 774244.

Compliance with ethical standards

Conflict of interest The authors declare that they have not conflict of interest.

Ethical approval The authors declare that the work was original research that has not been published previously, and is not under consideration for publication elsewhere.

References

Bassett MJ (1994) Tight linkage of purple pod character and the complex C locus in common bean. J Hered 85(4):288–290. https://doi.org/10.1093/oxfordjournals.jhered.a111458

Bassett MJ (1997) A New Allele (Vwf) at the V Locus for Flower and Seedcoat Color in Common Bean. J Am Soc Hortic Sci 122(4):519–521. https://doi.org/10.21273/JASHS.122.4.519

Bassett MJ (2007) Genetics of seed coat color and pattern in common bean. Plant Breed Rev 28:239–315. https://doi.org/10.1002/978047168028.ech

Bassett MJ, Miklas PN (2007) A new gene, bic, with pleiotropic effects with (T P V) for bicolor flowers and dark olive-brown seed coat in common bean. J Am Soc Hortic Sci 132:352–356

Beninger CW, Hosfield GL, Bassett MJ (1999) Flavonoid composition of three genotypes of dry bean (Phaseolus vulgaris L.) differing in seed coat color. J Am Soc Hortic Sci 124:514–518

Beninger CW, Hosfield GL, Bassett MJ, Owens S (2000) Chemical and morphological expression of the B and Asp seedcoat genes in P. vulgaris L. J Am Soc Hortic Sci 125:52–58

Bonferroni CE (1936) Teoria statistica delle classi e calcolo delle probabilita. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 8: 3-62

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. https://doi.org/10.1093/bioinformatics/btm30

Caldas GV, Blair MW (2009) Inheritance of seed conditioned tannins and their relationship with seed-coat color and pattern genes in common bean (P. vulgaris L.). Theor Appl Genet 119:137–142. https://doi.org/10.1007/s00122-009-1023-4

Campa A, Murube E, Ferreira JJ (2018) Genetic diversity, population structure and linkage disequilibrium in a Spanish common bean diversity panel revealed through genotyping-by-sequencing. Genes 9(11):518. https://doi.org/10.3390/genes9110518

Campa A, Perez-Vega E, Pasucal A, Ferreira JJ (2010) Genetic analysis and molecular mapping of a quantitative trait loci in common bean against Pythium ultimum. Phytopathology 100(12):1315–1320. https://doi.org/10.1094/PHYTO-06-10-0161

Chaves-Silva S, dos Santos AL, Chanfun-Júnior A, Zhao J, Peres LE, Benedito VA (2018) Understanding the genetic regulation of anthocyanin biosynthesis in plants—Tools for breeding purple varieties of fruits and vegetables. Phytochemistry 153:11–27. https://doi.org/10.1016/j.phytochem.2018.05.013

Cichy KA, Wiesinger JA, Mendoza FA (2015) Genetic diversity and genome-wide association analysis of cooking time in dry bean (P. vulgaris L.). Theor Appl Genet 128(8):1555–1567. https://doi.org/10.1007/s00122-015-2531-z

Deakin JR (1974) Association of seed color with emergence and seed yield of snap beans. J Am Soc Hortic Sci 99(2):110–114

Diaz AM, Caldas GV, Blair MW (2010) Concentrations of condensed tannins and anthocyanins in common bean seed coats. Food Res Int 43:595–601. https://doi.org/10.1016/j.foodres.2009.07.014

Doerge RW (1996) Constructing genetic maps by rapid chain delineation. J Quant Trait Loci 2:121–131

Doyle JJ, Doyle LH (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Dübben FH, Marcos-Filho J (2001) The seed coat as a modulator of seed-environment relationships in Fabaceae. Braz J Bot 24(4):365–375. https://doi.org/10.1590/S0100-8404200100000002

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) MYB transcription factors in Arabidopsis. Trends Plant Sci 15:573–581. https://doi.org/10.1016/j.tplants.2010.06.005

Elias LG, De Fernandez DG, Bressani RJ (2006) Possible effects of seed coat polyphenolics on the nutritional quality of bean protein. J Food Sci 44(2):524–527. https://doi.org/10.1111/j.1365-2621.1979.tb03827.x

Elsadr HT, Wright LC, Pauls KP, Bett KE (2011) Characterization of seeds coat post-harvest darkening in common bean (P. vulgaris L.). Theor Appl Genet 123(8):1467–1472. https://doi.org/10.1007/s00122-011-1683-8

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. https://doi.org/10.1371/journal.pone.0019379

Emerson R (1909) Factors for molting in beans. J Hered 5(1):368–375. https://doi.org/10.1093/jhered/5s.1.368

FAO (2020) http://www.fao.org/faoestat/en/#data/QC. Last accessed March 2021
Feenstra WJ (1960) Biochemical aspect of seedcoat colour inheritance in *P. vulgaris* L. Meded Landbouwhogeschool Wageningen (60):1–53. Retrieved from https://edepot.wur.nl/183363

Gonzalez A, Zhao M, Leavitt JM, Lloyd AM (2008) Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/MYB transcriptional complex in *Arabidopsis* seedlings. Plant J 53:814–827. https://doi.org/10.1111/j.1365-313X.2007.03373.x

Guo Y, Qiu LJ (2013) Allele-specific marker development and selection efficiencies for both flavonoid 3′-hydroxylase and flavonoid 3′,5′-hydroxylase genes in soybean subgenus soja. Theor Appl Genet 126:1445–1455. https://doi.org/10.1007/s00122-013-2063-3

Hagerty C, Cuesta-Marcos A, Cregan P et al (2016) Mapping snap bean pod and color traits, in a dry bean × snap bean recombinant inbred population. J Am Soc Hortic Sci 141(2):131–138

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium, 41: 95–98. https://doi.org/10.14601/Phytopath_Mediterr-14998ul.29

Hall TA (2011) BioEdit: an important software for molecular biology. GERF Bull Biosci 2:60–61

Herniter IA, Muñoz-Amatriain M, Lo S, Go Y-N, Close TJ (2018) Identification of candidate genes controlling black seed coat and pod tip color in cowpea (*Vigna unguiculata* [L.] walp). G3 Genes Genomes Genet 10:3337–3355. https://doi.org/10.1534/g3.118.200521

Islam N, Bett K, Pauls K, Marsolais F, Dhaubhadel S (2020) Postharvest seed coat darkening in pinto bean (*P. vulgaris*) is regulated by Psd, an allele of the basic helix-loop-helix transcription factor P. New Phytol Found 2(6):663–677. https://doi.org/10.1002/ppp.10132

Kooiman H (1931) Monograph on the genetics of *Phaseolus* (especially Ph. vulgaris and Ph. multiflorus). Bibliographica Genetica 8:295–413

Kyle MM, Dickson MH (1988) Linkage of hyperversitivity to five viruses with the locus in *P. vulgaris* L. J Hered 79(4):308–311. https://doi.org/10.1093/oxfordjournals.jhered.a10156

Lamprecht H (1932). Beiträge zur Genetik von *P. vulgaris* L. XII. Zur Vererbung der Blüten-und Stammfarbe. Hereditas 21:129-166

Lamprecht H (1935) Zur Genetik von *Phaseolus vulgaris*. Zur Vererbung der Testafarbe. Hereditas 16:169-211

Lamprecht H (1935) Zur Genetik von *Phaseolus vulgaris*. XII. Über die Vererbung der Blütenund Stammfarbe. Hereditas 21:129-166

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp123

Margarido GR, Souza AP, Garcia AA (2007) OneMap: software for genetic mapping in outcrossing species. Hereditas 144:78–79. https://doi.org/10.1016/j.plantsci.2007.04.006-0661.20000.x

McCleal PE, Lee RK, Otto C, Gepts P, Bassett MJ (2002) Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean (*P. vulgaris* L.). J Hered 93:148–152. https://doi.org/10.1093/jhered/93.2.148

McCleal PE, Pett KE, Stonehouse R, Lee R, Pfieger S et al (2018) White seed color in common bean (*P. vulgaris*) results from convergent evolution in the P (pigment) gene. New Phytol 219:1112–1123. https://doi.org/10.1111/nph.15259

Myers J, Wallace L, Moghaddam S et al (2019) Improving the health benefits of snap bean: genome-wide association studies of total phenolic content. Nutrients 11:2509. https://doi.org/10.3390/nu11050558

Nesi N, Joncl C, Debeaujon K, Caboche M, Lepetitc L (2001) The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. Plant Cell 13:2099–2114. https://doi.org/10.1105/TPC.010098

Pedrosa-Harand A, Porch T, Gepts P (2008) Standard nomenclature for common bean chromosomes and linkage groups. Ann Re Bean Improv Coop