A R2R3-MYB gene-based marker for the non-darkening seed coat trait in pinto and cranberry beans (*Phaseolus vulgaris* L.) derived from ‘Wit-rood boontje’

M. Erfatpour¹ · K. P. Pauls¹

Received: 26 November 2019 / Accepted: 21 February 2020 / Published online: 28 February 2020
© The Author(s) 2020

Abstract

Key message The gene *Phvul.010G130600* which codes for a MYB was shown to be tightly associated with seed coat darkening in *Phaseolus vulgaris* and a single nucleotide deletion in the allele in Wit-rood disrupts a transcription activation region that likely prevents its functioning in this non-darkening genotype.

Abstract The beige and white background colors of the seed coats of conventional pinto and cranberry beans turn brown through a process known as postharvest darkening (PHD). Seed coat PHD is attributed to proanthocyanidin accumulation and its subsequent oxidation in the seed coat. The *J* gene is an uncharacterized classical genetic locus known to be responsible for PHD in common bean (*P. vulgaris*) and individuals that are homozygous for its recessive allele have a non-darkening (ND) seed coat phenotype. A previous study identified a major colorimetrically determined QTL for seed coat color on chromosome 10 that was associated with the ND trait. The objectives of this study were to identify a gene associated with seed coat postharvest darkening in common bean and understand its function in promoting seed coat darkening. Amplicon sequencing of 21 candidate genes underlying the QTL associated with the ND trait revealed a single nucleotide deletion (c.703delG) in the candidate gene *Phvul.010G130600* in non-darkening recombinant inbred lines derived from crosses between ND ‘Wit-rood boontje’ and a regular darkening pinto genotype. In silico analysis indicated that *Phvul.010G130600* encodes a protein with strong amino acid sequence identity (70%) with a R2R3-MYB-type transcription factor *MtPAR*, which has been shown to regulate proanthocyanidin biosynthesis in *Medicago truncatula* seed coat tissue. The deletion in the ‘Wit-rood boontje’ allele of *Phvul.010G130600* likely causes a translational frame shift that disrupts the function of a transcriptional activation domain contained in the C-terminus of the R2R3-MYB. A gene-based dominant marker was developed for the dominant allele of *Phvul.010G130600* which can be used for marker-assisted selection of ND beans.

Introduction

Common bean (*Phaseolus vulgaris* L.) is the most important food legume crop. It is grown worldwide for its nutrient-rich green pods or mature seeds (shell beans and dry beans). Dry beans are rich sources of vitamins and minerals, as well as

plant phytochemicals. Different market classes of dry beans (including black, white navy, dark red kidney, light red kidney, pinto, cranberry, and many others) are recognized and bred that are based on the physical properties of their seeds, including seed coat color and pattern, seed size, and seed shape.

Flavonoids represent one of the largest classes of plant phenolic compounds. More than 9000 individual compounds, characterized by a 15-carbon skeleton, arranged in two aromatic rings connected by a three-carbon bridge (C6–C3–C6) have been identified (Buer et al. 2010). Flavonoids are further divided by the nature of the C3 element into major subclasses, including chalcones, flavones, flavonols, flavan 3-ols, isoflavonoids, anthocyanins, and proanthocyanidins (PAs) (Davies and Schwinin 2006).

These compounds have a variety of biological activities in humans and plants. As dietary components, flavonoids are

Communicated by Albrecht E. Melchinger.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00122-020-03571-7>) contains supplementary material, which is available to authorized users.

✉ K. P. Pauls
ppauls@uoguelph.ca

¹ Department of Plant Agriculture, University of Guelph, Guelph, ON N1G 2W1, Canada

believed to enhance human health due to their antioxidant, hepatoprotective, antiviral, antibacterial, anti-inflammatory, and anticancer activities (Mishra et al. 2013; van Dam et al. 2013; Cushnie and Lamb 2005; Pan et al. 2010; Koen et al. 2005). Flavonoids participate in a wide range of physiological activities in plants, including defense against herbivores, pests, and pathogens (Bennett and Wallsgrove 1994), UV-light protection (Cetin 2014; Zhao et al. 2017), attraction of insect pollinators and fruit-dispersing animals, signaling to rhizobial bacteria (Cipollini and Levey 1997; Wink 2008), and allelopathism (Li et al. 2010).

These phenolic secondary metabolites are synthesized as part of the larger phenylpropanoid pathway, the flavonoid biosynthesis pathway, by the activity of a multienzyme complex (flavonoid metabolon) which is weakly associated with the cytoplasmic face of the endoplasmic reticulum (ER) (Hrazdina and Wagner 1985). The early steps in the pathway begin with the condensation of three molecules of malonyl-CoA with one 4-coumaroyl-CoA molecule to yield chalcone by chalcone synthase (CHS) (Clegg and Durbin 2003). Chalcone then is converted to naringenin by chalcone isomerase (CHI) (Shimada et al. 2003). From this point on, the pathway diverges into several branches leading to the synthesis of different flavonoid compounds according to the plant species, developmental stage, and growth conditions (Debeaujon et al. 2001). Two main branches of the flavonoid pathway, leading to the production of proanthocyanidins (PAs) and anthocyanins, share two more steps, including hydroxylation of naringenin to dihydroflavonols (dihydrokaempferol, dihydroquercetin, and dihydromyricetin) by flavanone-3-hydroxylase (F3H) and flavonoid 3'-hydroxylase (F3'H), and reduction of dihydroflavonols to leucoanthocyanidins by dihydroflavonol reductase (DFR). From this step onward, the pathway diverges into the PA branch and the anthocyanin branch.

The PAs, also called condensed tannins, are oligomers or polymers of 2,3-*cis*-flavan-ols (e.g., (–)-epiafzelechin, (–)-epicatechin, and (–)-epigallocatechin) and/or 2,3-*trans*-flavan-3-ols (e.g., (+)-afzelechin, (+)-catechin, and (+)-gallocatechin) (Davies and Schwinning 2006). The two most common monomeric units of PAs are (–)-epicatechin and (+)-catechin (He et al. 2008). In addition to DFR, PA biosynthesis is dependent upon leucoanthocyanidin reductase (LAR), anthocyanidin synthase (ANS), and anthocyanidin reductase (ANR). LAR catalyzes the conversion of leucoanthocyanidins to 2,3-*trans*-flavan-3-ols through the PA branch, and ANR catalyzes the conversion of anthocyanidins (synthesized from the conversion of leucoanthocyanidins by ANS through the anthocyanin branch) to 2,3-*cis*-flavan-ols (Davies and Schwinning 2006).

Monomer composition and the degree of polymerization of PAs, as well as the positions and stereochemistries of the carbon–carbon interflavan bonds, vary in different plant

species and in different tissues and developmental stages of a single plant (Tanner et al. 2003). For example, PAs are predominantly composed of (–)-epicatechin units in the seeds of the model plants *Arabidopsis thaliana* and *Medicago truncatula* and in the leaves of the legume *Lotus corniculatus*, while PAs contain both catechin and epicatechin in *P. vulgaris* and *Pisum sativum* (pea) seeds (Beninger et al. 2005; Chen et al. 2015; Ferraro et al. 2014).

In seeds, PAs are synthesized in the inner integument or endothelium layer of the seed coat during development (Smýkal et al. 2014) and localize to the cell walls during maturation (Zhao et al. 2010). The presence of PA precursors, catechin and epicatechin, in the seed coat has been associated with seed coat postharvest darkening (PHD) phenomenon in pinto beans and cranberry beans (Beninger et al. 2005; Chen et al. 2015). Because a light seed coat color is considered by consumers to be an indicator of quality or freshness (Nasar-Abbas et al. 2009), darkened seeds have reduced value in the market. It is worth mentioning that flavan-3-ols and their polymers have been associated with sensory properties of bitterness and astringency (Robichaud and Noble 1990). As the degree of polymerization of flavan-3-ols increases, bitterness decreases, and astringency increases (Peleg et al. 1999).

The structural genes encoding PA biosynthetic enzymes and their regulatory factors have been well characterized in several model plant species including *Arabidopsis*, maize (*Zea mays*), snapdragon (*Antirrhinum majus*), petunia (*Petunia hybrida*), and *Medicago* (Mol et al. 1998; Quattrocchio et al. 2006; Schwinning et al. 2006; Pang et al. 2007). This information has been used to aid genetic elucidation of PA biosynthesis in other plant species. The genes underlying the *Arabidopsis* flavonoid pathway are mostly single-copy genes named *Transparent TESTA* (*TT*), making this species a unique organism to understand the major branch pathways of flavonoid biosynthesis and their subcellular localization (Winkel-Shirley 2001). However, the genetic basis of PA biosynthesis, leading to the seed coat PHD in two market classes of *P. vulgaris*, including pinto beans and cranberry beans, remains unelucidated.

PA precursors are transported from the cytosolic surface of the ER to the vacuole. At least three non-exclusive mechanisms have been suggested for the intracellular and/or extracellular transport of flavonoid compounds: (1) membrane-mediated transport, (2) vesicle trafficking, and (3) glutathione *S*-transferase (GST)-mediated transport (Zhao 2015). Two multi-drug and toxic compound extrusion (MATE) transporters, MATE and MATE1, are vacuolar membrane-localized transporters in *Arabidopsis TT12* and *Medicago*, respectively. They mediate vacuolar uptake of epicatechin 3-*O*-glucoside (glycosylated 2,3-*cis*-flavan-3-ol) for PA biosynthesis in the seed coat (Debeaujon et al. 2001; Zhao and Dixon 2009). The *Arabidopsis TT9* gene encodes a

Golgi-localized vesicle trafficking factor, GFS9, which contributes to deposition of PAs in vacuoles of the seed coat (Ichino et al. 2014). The *Arabidopsis* *TT19* encodes a GST protein with possible involvement in vacuolar transport of PA precursors (Kitamura et al. 2004).

During aging, vacuolar and apoplastic phenolic substrates such as the flavonols (kaempferol and quercetin) and the flavan-3-ols [(+)-catechin and (–)-epicatechin] are oxidatively transformed to brown components by the activity of peroxidase (POD) and polyphenol oxidase (PPO) enzymes, as well as autoxidation (Vaughn and Duke 1984; Jiménez and García-Carmona 1999; Laukkanen et al. 1999; Takahama and Hirota 2000; Takahama 2004). *AtTT10* and eight *BnTT10* genes have been proposed to be involved in these reactions in *Arabidopsis* and three *Brassica* species (three from *B. napus*, three from *B. rapa*, and three from *B. oleracea*), respectively (Pourcel et al. 2005; Zhang et al. 2013).

The genes encoding the flavonoid biosynthetic enzymes are classified as early biosynthetic genes (EBGs) (*CHS*, *CHI*, and *F3H*) and late biosynthetic genes (LBGs) (*DFR*, *ANS*, *ANR*, and *LAR*) (Nesi et al. 2001; Winkel-Shirley 2001; Lepiniec et al. 2006). The expression of LBGs involved in PA biosynthesis in the seed coat is mainly regulated by a ternary MYB-bHLH-WD40 (MBW) protein complex which is formed by highly conserved transcription factors, including a MYB interacting with MYC-like basic helix-loop-helix (bHLH) proteins and WD40-repeat proteins (Baudry et al. 2004). Generally, transcription factors have modular structures and contain two functional domains, a highly conserved DNA-binding domain which binds to the promoters of the target genes, and a highly variable C-terminal region which contains binding sites for interaction with other proteins such as coregulatory proteins (Liu et al. 1999). DNA-binding domains have been widely used to classify prokaryotic and eukaryotic transcription factors into families. MYB DNA-binding domains are generally composed of 1–4 imperfect amino acid sequence repeats (R) of 50–53 amino acids (Chen et al. 2006; Jiang et al. 2004a, b; Ogata et al. 1996; Stracke et al. 2001). Based on the number of repeats found in the DNA-binding domain, MYB proteins are classified into four subfamilies, including one repeat or MYB-like proteins (R1MYB); two repeats (R2R3-MYB); three repeats (R1R2R3-MYB); and four repeats (R0R1R2R3-MYB) (Du et al. 2012; Jiang et al. 2004a, b). Each repeat contains three conserved tryptophan (W) residues spaced 18 or 19 amino acids apart, as well as three α -helices with the second and the third helices forming a helix-turn-helix structure (Ogata et al. 1996). The first tryptophan of the R3 repeat can be replaced by hydrophobic amino acids such as isoleucine (I) and phenylalanine (F) (Ambawat et al. 2013; Du et al. 2012). The third helices of the R2 and R3 repeats serve as ‘recognition helices’ that make direct contact with the major groove of the promoter DNA (Jia et al. 2004).

In *Arabidopsis*, TT2, a R2R3-MYB transcription factor, TT8, a bHLH transcription factor, and TRANSPARENT TESTA GLABRA1 (TTG1), a WD40 protein, form an MBW regulatory complex which plays the main role in production of PA in the endothelium of wild-type seeds by regulating the expression of: *DFR*, *ANS*, *BAN*, *TT19*, *TT12*, and *AHA10* (Xu et al. 2014). The MYB5–TT8–TTG1 protein complex is another MBW which regulates the expression of *DFR*, *ANS*, and *TT12* for PA accumulation in *Arabidopsis* seed coat (Xu et al. 2014).

In the seed coat of *Medicago*, a MBW complex consisting of *MtPAR* (R2R3-MYB), *MtTT8* (bHLH), and *MtWD40-1* (WD40 protein) regulates the expression of two key structural genes, *ANR* and *LAR*, leading to PA biosynthesis (Pang et al. 2009; Verdier et al. 2012; Li et al. 2016). Moreover, a quaternary complex has been suggested for the regulation of PA biosynthesis in the seed coat of *Medicago* in which *MtMYB5* and *MtMYB14* activate the *ANR* and *LAR* genes synergistically in the presence of *MtTT8* and *MtWD40-1* proteins (Liu et al. 2014).

Three distinct phenotypes have been defined in beans based on the rapidity and intensity of seed coat PHD they undergo (Elsadr et al. 2011), namely: (1) regular darkening (RD) which darken during normal storage within a few weeks after harvest or within 24 h after exposure to UV light (Junk-Knievel et al. 2008), (2) slow darkening (SD) which require distinctly longer periods for the seed coat to exhibit PHD than RD beans, and (3) non-darkening (ND) which never darken with age. These phenotypes are determined by interactions between two loci identified by classical segregation studies, including the *J* locus which contributes to mature color development in RD seed coats (Bassett 1996a, b) and the *Sd* locus which determines how quickly a seed coat will darken in plants with a *J* genotype (Junk-Knievel et al. 2008). Plants that are homozygous recessive at *J* do not form brown pigmentation during seed maturation and after harvest (Bassett 2007; Elsadr et al. 2011; Erfatpour et al. 2018).

The STS marker OL4S₅₀₀ was linked to *J* on chromosome Pv10 (McClean et al. 2002). Using a recombinant inbred line (RIL) population derived from a cross between a ND cranberry-like bean ‘Wit-rood boontje’ and a SD pinto bean (1533-15), Erfatpour et al. (2018) mapped a QTL associated with the ND trait in an interval flanked by single nucleotide polymorphism (SNP) markers ss715646341 (Pv10:29.74 cM) and ss715647913 (42.27 cM) on chromosome Pv10 of the genetic map and between SNPs ss715646348 (Pv10:40 164,667 bp) and ss715647917 (Pv10:40 295,580 bp) on the physical assembly. In addition to the OL4₅₀₀ marker previously linked to *J*, forty candidate genes were identified in the ND QTL region including three genes that encode ‘MYB proteins’ (Erfatpour et al. 2018). The *Sd* locus was mapped on chromosome Pv07 in a different study

(Felicetti et al. 2012) and in the study based on the ‘Witrood boontje’ × 1533-15 RILs (Erfatpour et al. 2018) confirming that the genetic locus controlling the SD trait differs from the ND trait.

There is limited information about the biochemistry, structural genes, and regulatory genes contributing to the seed coat PHD in *P. vulgaris* seeds. The seed coat of a RD pinto bean, CDC Pintum, was found to contain higher levels of kaempferol and PAs as compared to SD 1533-15 (Beninger et al. 2005). Formation and subsequent oxidation of kaempferol-catechin dimers was the proposed mechanism for seed coat PHD in pinto beans (Beninger et al. 2005). Marles et al. (2008) reported a higher concentration of kaempferol and PPO activity in the seed coat of RD pinto beans compared to SD samples. It was found that the seed coat of ND cranberry beans do not contain catechin, epicatechin, or PAs (Chen et al. 2015). Transcript levels of flavonoid biosynthetic genes, *PvF3H1*, *PvDFR1*, *PvDFR2*, *PvLAR*, *PvANS*, *PvANR1*, *PvMATE1*, and *PvMATE2* were found to be significantly higher at early and intermediate stages of seed development in the seed coat of a RD cranberry bean compared to a ND cranberry bean and their expression levels were coordinately increased with transcript levels of *PvMYB6* (*Phvul.006G114800*), *PvMYB9* (*Phvul.011G105600*), and *PvMYB11* (*Phvul.003G222400*) (Freixas Coutin et al. 2017).

PvANR1 was reported to be more strongly associated with proanthocyanidin levels in the seed coat of RD cranberry bean than *PvF3H1* and *PvDFR1*, consistent with contribution of *ANR* genes to PA accumulation in seed coats of pea, soybean, *Medicago*, *Arabidopsis*, and *B. napus* (Debeaujon et al. 2003; Auger et al. 2009; Pang et al. 2007; Kovinich et al. 2012; Ferraro et al. 2014; Freixas Coutin et al. 2017). Although *PvLAR* and *MtLAR* are expressed in the seed coat tissue in *P. vulgaris* and *Medicago*, respectively, the roles of these genes’ products in seed coat PA biosynthesis remains unknown (Pang et al. 2007; Freixas Coutin et al. 2017).

This research aimed to (1) determine a variant or set of variants in candidate genes previously identified in the ND QTL that might identify it as the gene regulating seed coat PHD in common bean, (2) predict the functional effect of the genetic variant(s) on PA biosynthesis, and (3) develop a gene-based marker that can be utilized in marker-assisted selection (MAS) in early stages of a breeding program for ND beans.

Materials and methods

Plant material

Six different genotypes including ‘Wit-rood boontje,’ 1533-15, RIL29, RIL81, Othello, and Etna, were used for target

resequencing. ‘Wit-rood boontje’ is a plant introduction (PI439540) obtained from the USDA National Centre for Genetic Resources Preservation at Ft. Collins, CO, which has very pale pink, cranberry bean-like stripes on a white background (Erfatpour et al. 2018). 1533-15 is a slow darkening pinto bean variety from the Crop Development Center (CDC), of the University of Saskatchewan (Felicetti et al. 2012). RIL29 with ND seed coat phenotype and RIL81 with regular darkening seed coat phenotype are two F_2 -derived F_6 RILs from a ‘Wit-rood boontje’ × 1533-15 cross (Erfatpour et al. 2018). Othello and Etna are regular darkening pinto bean and regular darkening cranberry bean varieties, respectively. A population of 128 F_2 -derived F_6 RILs developed from the ‘Wit-rood boontje’ × 1533-15 cross plus 38 commercial varieties and breeding lines of pinto beans and cranberry beans were used to assess the efficiency of a SNP-based marker. The varieties and lines used for marker screening were: AC Island, Burdett, Centennial, Croissant, Jackpot, Kodiak, Kimberly, La Paz, Lariat, Marmot, ME70, ME98, ME105, ME128, ME145, ME182, ME233, Othello, P13HR088, PT11-9, Santa Fe, Sequoia, Shoshone, Stampepe, UI 114, Windbreaker, and 16-NDP1 from pinto beans, and CBX 1148, C13HR118, C15HR009, C15HR185, Etna, OAC Candycane, OAC Racer, P12HR1410, Red Rider, SVMTH, and 94CTCOOP-8184 from cranberry beans. The seed coat phenotypes were determined by visual screening of UV treated seeds (Junk-Knievel et al. 2008).

Targeted resequencing

Candidate genes in the QTL region were prioritized based on their function annotations obtained from the *P. vulgaris* v2.1 reference assembly available in Phytozome v12.1 (<https://phytozome.jgi.doe.gov>). Targeted resequencing of the most promising genes for the seed coat PHD trait was performed using Illumina Next Generation Sequencing (NGS) with MiSeq system. The instructions for Illumina targeted resequencing, including amplicon generation, library preparation, amplicon sequencing, and data analysis (www.illumina.com/technology/next-generation-sequencing.html) were followed as described at the Illumina website (<http://www.illumina.com>).

Amplicon generation

DNA samples from the genotypes were isolated using CTAB extraction protocol (Doyle and Doyle 1990) and quantified using a Qubit fluorometer. The Integrated DNA Technologies PrimerQuest Tool (<https://www.idtdna.com/PrimerQuest/Home/Index>) was used to design primers. Amplicons were generated from the promising genes in each genotype using polymerase chain reactions (PCRs) in volumes of 20 µl containing: 10 ng template DNA,

1 × PCR buffer, 1.5 mM magnesium chloride (MgCl_2), 0.2 mM dNTPs, 0.5 μM forward and 0.5 μM reverse primers, and 1 U of Taq polymerase. The PCRs were performed in a Bio-Rad Laboratories, Inc. (USA) thermal cycler with a program that included: an initial denaturation step for 2 min at 94 °C, followed by 35 cycles each of 1 min at 94 °C (denaturation), 1 min at annealing temperature of primer (annealing), 1 min at 72 °C (extension), and a final extension of 7 min at 72 °C. The amplicons were electrophoretically separated with 0.8% agarose gels at 90 V for 2 h, stained with ethidium bromide (0.5 $\mu\text{l/ml}$ of agarose gel), and visualized with UV transillumination. The DNA bands with expected sizes were cut from the agarose gels, purified using Zymoclean™ DNA Gel Recovery Kit (ZYMO RESEARCH), quantified by Qubit and normalized to 2.5 ng/ μl .

Library preparation

Amplicons from each genotype were pooled and 50 ng of DNA (25 μl of 2.5 ng/ μl) from each pool were used for library preparation, using the Illumina Nextera® DNA Library Prep Kit as described in the reference guide (Document # 15027987 v01). The library prep workflow included a ‘tagmentation’ reaction which simultaneously fragmented and tagged the amplicons with adapter sequences using the Nextera transposome. The tagmented fragments were purified from the Nextera transposome and PCR amplified to add Index 1 (i7) and Index 2 (i5) adapters required for cluster generation and sequencing to each fragment. AMPure XP beads were used to purify the libraries for each genotype and remove short library fragments. The size distributions of the fragments in the amplicon libraries were determined with an Agilent Technology 2100 Bioanalyzer. The optimal fragment sizes range from ~250 to 1000 bp. The libraries were normalized to 4 nM, and equal volumes were pooled.

Amplicon sequencing

The prepared library was denatured and diluted to a final volume of 600 μl using MiSeq Reagent kit v3, 600 cycles (Catalog # MS-102-3003) and following the MiSeq System Denature and Dilute Libraries Guide (Document # 15039740 v01). A PhiX library was prepared, denatured, and diluted for use as a sequencing control. The Library (594 μl) and PhiX control (6 μl) were combined and loaded onto the pre-filled reagent cartridge according to the Reagent Preparation Guide (Part # 15044983 Rev.). The MiSeq instrument was set up according to the MiSeq System Guide (Document # 15027617 v03 Material # 20024228) and paired-end 300-cycle per read sequencing was performed.

Data analysis

FASTQC software (Andrews 2010) was used to control the quality of the Illumina output and sequences were aligned to the *P. vulgaris* reference genome using CLC Genomics Workbench 11 (<http://www.qiagenbioinformatics.com>). Multiple DNA and protein sequence alignments were performed with Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). Protein–protein BLAST (blastp) analysis was performed in NCBI to query the non-redundant protein sequence (nr) database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). A phylogenetic tree was built using the neighbor-joining algorithm with Mega7.0 software (<http://www.megasoftware.net/>).

Results

Twenty-one candidate genes were selected out of the genes that are annotated in the QTL region identified in previous work to be associated with the PHD trait (Erfatpour et al. 2018) as the most likely genes to be control this trait. The information regarding the annotation of the selected genes and the primer sets used for amplicon generation is shown in Table 1 and supplementary Table S1, respectively.

The *Phvul.010G130300* gene was annotated as ‘autophagy-related gene 12 (*ATG12*)’ which is known for its roles in plant autophagic nutrient recycling during senescence and survival under N- and C-limiting conditions (Doelling et al. 2002; Thompson et al. 2005) through its attachment to *ATG5* by a ubiquitin-like conjugation system. Plants with enhanced autophagy show elevated levels of expression of genes involved in anthocyanin and flavonoid biosynthesis (Masclaux-Daubresse et al. 2014; Minina et al. 2018). The genes *Phvul.010G130500*, *Phvul.010G130600*, and *Phvul.010G131400* were annotated as ‘MYB proteins’ which are involved in transcriptional regulation of PA biosynthesis in a variety of plant species (Akagi et al. 2009; Koyama et al. 2014; Lu et al. 2017) and were the most interesting candidates in the present study. *Phvul.010G130700* and *Phvul.010G133300* were annotated as ‘transcription initiation factor (TFIID) subunit 12B’ and ‘nuclear transcription factor Y subunit A-3-related,’ respectively. Both genes are general (basal) transcription factors which play key roles in the regulation gene expression of a subset of genes (Robles et al. 2007; Kabe et al. 2005). *Phvul.010G131300*, *Phvul.010G133700*, *Phvul.010G133800*, *Phvul.010G133900*, and *Phvul.010G134000* are genes of unknown function. The *Phvul.010G133100* gene was annotated as a ‘developmentally regulated GTP-binding protein 2 (DRG2)’ (formerly known as DRG1) which is a class of

Table 1 Twenty-one candidate genes prioritized for sequencing

Gene ID	Annotation	Gene size (bp)	Number of amplicon(s) per gene	Length of genomic region (bp) covered by amplicon(s)
Phvul.010G130300	Ubiquitin-like ATG12 (ATG12)	2465	2	2380
Phvul.010G130500	MYB-like DNA-binding protein	1279	1	1314
Phvul.010G130600	MYB-like DNA-binding protein	1506	1	1453
Phvul.010G130700	Transcription initiation factor TFIID Subunit 12B	9823	8	11670
Phvul.010G131300	Uncharacterized protein	5563	4	5539
Phvul.010G131400	MYB Family transcription factor TRFL6-related	6607	5	6697
Phvul.010G133100	Developmentally regulated GTP-binding protein 2	684	1	712
Phvul.010G132300	Ring/U-box domain-containing protein 2	590	1	720
Phvul.010G132500	MDC16-related	1708	1	1544
Phvul.010G132600	Apoptosis inhibitor 5-related	8359	6	7653
Phvul.010G132700	Nonsense-mediated mRNA decay protein 3	2390	1	1564
Phvul.010G132800	Protein RCC2	4719	2	2351
Phvul.010G132900	Diglyceride acyltransferase	5781	4	4674
Phvul.010G133300	Nuclear transcription factor Y subunit A-3-related	4567	4	5170
Phvul.010G133400	PHD finger protein alfin-like 6-related	4027	2	3621
Phvul.010G133500	nitrogen metabolic regulation protein NMR-related	2999	2	2756
Phvul.010G133600	Solute carrier family 25	2370	1	1158
Phvul.010G133700	Domain of unknown function	3531	2	26,40
Phvul.010G133800	Domain of unknown function	1257	1	1154
Phvul.010G133900	Domain of unknown function	1323	1	1260
Phvul.010G134000	Domain of unknown function	552	1	1043
Total			51	67,073

The table includes their gene name (Phytozome v12.1), annotations according to *P. vulgaris* v2.1, sizes (bp), the number of PCR amplicons generated for each gene, and the length of genomic region (bp) covered by amplicon(s)

signal-transducing GTPases with possible roles in vesicle transport (Assmann 2002) and heat stress responses (Stafstrom 2008). The *Phvul.010G132300* gene possess ‘E3 ubiquitin ligase activity’ which can participate in a range of functions, from self-incompatibility (Stone et al. 2003) and hormone responses (Amador et al. 2001; Samuel et al. 2008) to biotic (Heise et al. 2002; Kirsch et al. 2001) and abiotic stress responses (Cho et al. 2006; Samuel et al. 2006). The proteins encoded by the *Phvul.010G132500* and *Phvul.010G132600* genes are uncharacterized. The *Phvul.010G132700* gene was annotated as a ‘Nonsense-mediated mRNA decay (NMD) protein’ that controls the quality of mRNA via regulation of gene expression and destruction of aberrant mRNA and may function in plant pathogen responses (Rayson et al. 2012). *Phvul.010G132800* was annotated as a ‘regulator of chromosome condensation 2 (RCC2),’ which plays a functional role in cells undergoing mitosis (Yenjerla et al. 2013). The *Phvul.010G132900* gene was annotated as a ‘diglyceride acyltransferase (DGAT)’ which catalyzes the synthesis of triacylglycerol (Ohlrogge and Browse 1995). *Phvul.010G133400* was annotated as a ‘PHD finger protein,’ which participates in transcriptional

regulation activity (Wilson et al. 2001), chromatin regulation (Aasland et al. 1995) and confers abiotic stress tolerance in *Arabidopsis* (Wei et al. 2015). *Phvul.010G133500* encodes a nitrogen metabolic regulation protein which may influence the C/N ratio and can be correlated with expression of the late flavonoid/anthocyanin biosynthetic genes, *ANS*, *UFGT* and *FLS* (Wan et al. 2015). *Phvul.010G133600* was annotated as a ‘solute carrier family 25 (SLC25)’ gene which is a mitochondrial transporter that regulates the transport of many types of substances into or out of mitochondria (Höglund et al. 2011).

In total, 51 PCR amplicons were successfully generated and sequenced, covering 67,073 bp of the QTL region and representing 21 candidate genes (Table 1). The FASTQC analysis indicated that the per base sequence quality score was high (> Q30) and the mean quality score per read was over 36 (Figure S1), both results indicating that the raw sequencing data were of good quality. DNA sequence analysis revealed single and multiple nucleotide polymorphisms distributed either within coding sequences (exons) or non-coding sequences (introns) of most of the genes across the six different genotypes, including non-darkening ‘Wit-rood

boontje' and RIL29; slow darkening 1533-15; and regular darkening RIL81, Othello, and Etna. The results showed that for most of the detected polymorphisms in the amplicons from the candidate genes across the genotypes there was no consistent association between the genetic variants and the seed coat PHD trait (Figure S2). However, there was a single nucleotide polymorphism that was located in the coding region of gene *Phvul.010G130600* that did differentiate the individuals with the ND seed coat phenotype from the ones with SD and RD seed coat phenotypes. The polymorphism represents a deletion of a guanine from the sequence of 'Wit-rood boontje' and RIL29 (which is ND) in c.703delG (Figs. 1, 2). *Phvul.010G130600* is annotated as 'MYB transcription factor' in the *P. vulgaris* v2.1 reference assembly (Phytozome V12.1).

As shown in Fig. 3, the gene *Phvul.010G130600* of the reference genome *P. vulgaris* G19833 (Schmutz et al. 2014) encodes a R2R3-MYB protein with conserved DNA-binding domains R2 [-W-(X19)-W-(X19)-W-] and R3 [-F/I-(X18)-W-(X18)-W-], as well as the amino acid motif [D/E]Lx₂[R/K]x₃Lx₆Lx₃R (Phytozome V12.1). The 3 tryptophan (W) residues of the R2 repeat are at positions 16, 36, and 56, and in the R3 repeat an isoleucine residue replaces the first tryptophan (W) at position 69 and 2 additional tryptophan residues are at positions 88 and 107, respectively. The amino acid motif [D/E]Lx₂[R/K]x₃Lx₆Lx₃R is conserved among MYB proteins and located on helices 1 and 2 of the R3 repeat that interact with R/B-like bHLH factors (Zimmermann et al. 2004). The sequence for gene *Phvul.010G130600* in 1533-15 was determined to be identical to the sequence in the reference genome (data not shown) and thus the protein would have an identical amino acid sequence to that of the reference genome.

According to a protein BLAST search, the MYB encoded by the *Phvul.010G130600* exhibited 91% sequence identity with a transcription factor WER-like (*LOC106773092*) in *Vigna radiata* (mung bean) across its entire query length, followed by 83%, 82%, and 82% sequence identity with transcription factor MYB8 (*LOC109802394*) in *Cajanus cajan* (pigeon pea), an anthocyanin regulatory C1 protein (*C1*) in *Mucuna pruriens*, and a transcription repressor MYB6 isoform X1 (*Glyma.16G007100*) in *Glycine max*, respectively. It also showed 80% identity with a transcription factor MYB8 (*Glyma.07G037700*) in *G. max* and 70% identity with a MYB transcription factor *MtPAR* (*Medtr8G20490*) in *M. truncatula* (Figs. 3, 4, Table 2).

The SNP in the sequence for gene *Phvul.010G130600* in 'Wit-rood boontje' is predicted to cause a frame shift change, with alanine235 as the first affected amino acid and the new reading frame remaining open for 36 amino acids (A235fsX271). Therefore, the *Phvul.010G130600* protein in 'Wit-rood boontje' would be expected to have identical DNA-binding domains to the wild-type protein, but the

amino acid residues 236 to 256, which show enrichment for negatively charged residues glutamic acid (E) and aspartic acid (D) and hydrophobic residues leucine (L), in the wild-type *Phvul.010G130600* protein, will be disrupted in the protein encoded by the ND allele because of the frame shift caused by the SNP (Fig. 5). In addition, the protein in the ND lines will have a longer C-terminal region (Fig. 5).

A sequence-tagged site (STS) marker was developed for the wild-type allele of the *Phvul.010G130600* gene. The primer sequences were designed to bind a site that ends with the G at c.703 that is missing in the ND genotype (Fig. 6, Table 3). A single DNA fragment with a size of 254 bp was amplified from genomic DNA from RD and SD genotypes, but not the ND genotypes (Fig. 6).

The occurrence of the marker was determined for the parents, 'Wit-rood boontje' and 1533-15, and a mapping population from a 'Wit-rood boontje' × 1533-15 cross, including 46 RD RILs, 31 SD RILs, and 51 ND RILs. The marker was also tested with 38 commercial varieties and breeding lines of pinto beans and cranberry beans with RD and ND seed coat phenotypes. No recombination event was observed for the marker among the test individuals; indicating that the polymorphism on which it is based is very close to or in the gene responsible for the ND trait in common bean.

Discussion

The current results suggest that the gene *Phvul.010G130600*, which encodes a R2R3-MYB transcription factor located on chromosome Pv10, close to the STS marker OL4S₅₀₀ which was previously found to be linked to *J* (McClellan et al. 2002) and is located in the major QTL associated with the ND trait (Erfatpour et al. 2018), is responsible for the phenotypes conditioned by the *J* gene. Bassett 1996a, b described the effects of the recessive allele of the joker (*j*) gene, which include a decrease in the intensity of the background seed coat color it is introduced into and a reduction in seed coat after-darkening. The after-darkening observed in pinto and cranberry beans carrying the *J* allele has been associated with the accumulation of proanthocyanidin precursors (Beninger et al. 2005; Chen et al. 2015; Freixas Coutin et al. 2017). The after-darkening phenotype and its association with PA levels is consistent with the finding that the *Phvul.010G130600* gene encodes a protein with high sequence identity to the R2R3-MYB transcription factor encoded by *MtPAR*, which plays a significant role in PA biosynthesis in *Medicago* seed coat tissue (Pang et al. 2009; Verdier et al. 2012; Li et al. 2016).

Biosynthesis of PAs involves transcriptional regulation of late biosynthetic genes in the flavonoid pathway. Among the major groups of plant transcription factors, the R2R3-MYB proteins from the MYB family play important roles

Fig. 1 Sequence alignment of the gene *Phvul.010G130600* in ‘Wit-rood boontje’ [parent with the non-darkening (ND) seed coat phenotype] with 1533-15 [parent with the slow darkening (SD) seed coat phenotype]. Wit-rood boontje displayed a single nucleotide deletion (c.703delG, in red box) in the 3rd exon region (yellow shaded). The green-shaded segment identifies the 5' untranslated region (UTR), the blue-shaded segment identifies the 3' UTR, and the lowercase letters identify the introns (color figure online)

Wit-rood boontje ND 1533-15 SD	GTTAATTGAGTGCAATATGATAGTATAGGTAGTGTGTAGTGAGAAGAAGCAGTGGTA GTTAATTGAGTGCAATATGATAGTATAGGTAGTGTGTAGTGAGAAGAAGCAGTGGTA *****	60 60
Wit-rood boontje ND 1533-15 ND	AAGAAGATGGGAAGAAGACCCCTGCTGCCAAGAGATCAACAAGGTGCTTGGTCTCGA AAGAAGATGGGAAGAAGACCCCTGCTGCCAAGAGATCAACAAGGTGCTTGGTCTCGA *****	120 120
Wit-rood boontje ND 1533-15 SD	GAAGAAGATGAAACTCTCTCCAAATATGTTGCCATCCATGGAGAAGGAAATGGCAGAAG GAAGAAGATGAAACTCTCTCCAAATATGTTGCCATCCATGGAGAAGGAAATGGCAGAAG *****	180 180
Wit-rood boontje ND 1533-15 SD	GTTGCCCAAAATGCAGgtatataatacattttotgtgtgtactgtattacgtacataaa GTTGCCCAAAATGCAGgtatataatacattttotgtgtgtactgtattacgtacataaa *****	240 240
Wit-rood boontje ND 1533-15 SD	gtgtgcaacgtgtgttcatacacattcattgttaaatatgtgaagcactgtattgtgtgga gtgtgcaacgtgtgttcatacacattcattgttaaatatgtgaagcactgtattgtgtgga *****	300 300
Wit-rood boontje ND 1533-15 SD	gttttattgtgtgaatgaagtttgaagttggaagctgattttgtotattttgcatgtgtgtg gttttattgtgtgaatgaagtttgaagttggaagctgattttgtotattttgcatgtgtgtg *****	360 360
Wit-rood boontje ND 1533-15 SD	cctcacaacaaatcgtgtgttcattgacottaaatccaaatttaataactttgtttat cctcacaacaaatcgtgtgttcattgacottaaatccaaatttaataactttgtttat *****	420 420
Wit-rood boontje ND 1533-15 SD	tatagacattggcattttgtgtatacacacattacaatgatgtotgcagGTTTAAAGCG tatagacattggcattttgtgtatacacacattacaatgatgtotgcagGTTTAAAGCG *****	480 480
Wit-rood boontje ND 1533-15 SD	ATGTGAAAAAGCTGCAGACAAAGGTGGCTGAACATATCTCAAGCCTGGTATTAAAGAGG ATGTGAAAAAGCTGCAGACAAAGGTGGCTGAACATATCTCAAGCCTGGTATTAAAGAGG *****	540 540
Wit-rood boontje ND 1533-15 SD	TGACATCTCAGTGGATGAAGAAGATATGATTATCAGACTTCATCGTCTTCTTGGTAACAG TGACATCTCAGTGGATGAAGAAGATATGATTATCAGACTTCATCGTCTTCTTGGTAACAG *****	600 600
Wit-rood boontje ND 1533-15 SD	gtgatttttattatcaattgagtattaatagaccattttcaactcaattgaaataaa gtgatttttattatcaattgagtattaatagaccattttcaactcaattgaaataaa *****	660 660
Wit-rood boontje ND 1533-15 SD	aatttgaagtaaagaagaggtgtgttattttgtgcacacaaatgtaaaacatggatc aatttgaagtaaagaagaggtgtgttattttgtgcacacaaatgtaaaacatggatc *****	720 720
Wit-rood boontje ND 1533-15 SD	agaaattgtagtgaaggcaacatttagtacattacttactgtgtatgtgtataaaccatag agaaattgtagtgaaggcaacatttagtacattacttactgtgtatgtgtataaaccatag *****	780 780
Wit-rood boontje ND 1533-15 SD	tggttttttttatcatotttatgcagATGGGCATTGATAGCAAGAGGCTACCTGGACGA tggttttttttatcatotttatgcagATGGGCATTGATAGCAAGAGGCTACCTGGACGA *****	840 840
Wit-rood boontje ND 1533-15 SD	ACAGACAATGAAATCAAGAACTACTGGAACACCAATCTGTCAAGAAGTTACAGAAACAC ACAGACAATGAAATCAAGAACTACTGGAACACCAATCTGTCAAGAAGTTACAGAAACAC *****	900 900
Wit-rood boontje ND 1533-15 SD	CCTACATCATCATCCTCTGCTCTTCACTTCAACACAAACGTGGAGAAGAAGAGAAGATG CCTACATCATCATCCTCTGCTCTTCACTTCAACACAAACGTGGAGAAGAAGAGAAGATG *****	960 960
Wit-rood boontje ND 1533-15 SD	AGTAAATGTGTAGTACCACCTCCAGTCCTTCAAATAAAGTGGAGGGTAGTAGTGAA AGTAAATGTGTAGTACCACCTCCAGTCCTTCAAATAAAGTGGAGGGTAGTAGTGAA *****	1080 1080
Wit-rood boontje ND 1533-15 SD	GCTTGTTTTAGTGATTTCCTCATCGATATTGATCAAAACCAGTTGTTGATTGGTGATGAT GCTTGTTTTAGTGATTTCCTCATCGATATTGATCAAAACCAGTTGTTGATTGGTGATGAT *****	1140 1140
Wit-rood boontje ND 1533-15 SD	TCCAGCTCCAAGGTCCACAAACAGATGAAGAGCAGAACACAAACAGGTAGTCTTAAAC TCCAGCTCCAAGGTCCACAAACAGATGAAGAGCAGAACACAAACAGGTAGTCTTAAAC *****	1200 1200
Wit-rood boontje ND 1533-15 SD	AACAGTCTAGCTCATCATCACCTTCTCATCATTTGTCACAATCTCTTGACAGAAAATTC AACAGTCTAGCTCATCATCACCTTCTCATCATTTGTCACAATCTCTTGACAGAAAATTC *****	1259 1260
Wit-rood boontje ND 1533-15 SD	GACCCCTCTGGAGACCCCTCTTGGATGTGGAACCTCAAGAGGATGGCTTCCCTTCTTGGACTT GACCCCTCTGGAGACCCCTCTTGGATGTGGAACCTCAAGAGGATGGCTTCCCTTCTTGGACTT *****	1319 1320
Wit-rood boontje ND 1533-15 SD	GAAAATGATTGATCAAGGGATTATTATTATTCACTGAGATGAAAAAGGTGAAGTACTTA GAAAATGATTGATCAAGGGATTATTATTATTCACTGAGATGAAAAAGGTGAAGTACTTA *****	1379 1380
Wit-rood boontje ND 1533-15 SD	GCGATCAAGTGTCTAGATAAAATGCTATCTATATA TGGAACACCTTAACATAGCATCTA GCGATCAAGTGTCTAGATAAAATGCTATCTATATA TGGAACACCTTAACATAGCATCTA *****	1439 1440
Wit-rood boontje ND 1533-15 SD	GCTCTAGCTAGATGTACACACAAATCCAAGGTCTCTTCTCTTTTTCCTTTTTCCTTTTCCT GCTCTAGCTAGATGTACACACAAATCCAAGGTCTCTTCTCTTTTTCCTTTTTCCTTTTTCCT *****	1499 1500
Wit-rood boontje ND 1533-15 SD	TTCCCC 1505 TTCCCC 1506 *****	

5' UTR Exon 3' UTR

Fig. 2 Sequence alignments of genomic segments six different genotypes corresponding to the exon region (yellow shaded) of gene *Phvul.010G130600* containing the single nucleotide polymorphism (c.703delG, bordered by red lines) associated with the non-darkening trait in ‘Wit-rood boontje.’ The sequence in Wit-rood boontje [parent A with the non-darkening (ND) seed coat phenotype] is compared to: RIL29 (ND), 1533-15 [parent B with the slow darkening (SD) seed coat phenotype], RIL81 [regular darkening (RD)], Othello (RD pinto bean variety), Etna (RD cranberry variety), and G19833 (*P. vulgaris* reference genome G19833) (color figure online)

Wit-rood boontje ND	AACAGTCCTAGCTCATCATCACCTTCTCATCATTGTGCAATCTCTTG	CAGAGAAATTC	1259
RIL29 ND	AACAGTCCTAGCTCATCATCACCTTCTCATCATTGTGCAATCTCTTG	CAGAGAAATTC	1259
1533-15 SD	AACAGTCCTAGCTCATCATCACCTTCTCATCATTGTGCAATCTCTTG	CAGAGAAATTC	1260
RIL81 RD	AACAGTCCTAGCTCATCATCACCTTCTCATCATTGTGCAATCTCTTG	CAGAGAAATTC	1260
Othello RD	AACAGTCCTAGCTCATCATCACCTTCTCATCATTGTGCAATCTCTTG	CAGAGAAATTC	1260
Etna RD	AACAGTCCTAGCTCATCATCACCTTCTCATCATTGTGCAATCTCTTG	CAGAGAAATTC	1260
G19833	AACAGTCCTAGCTCATCATCACCTTCTCATCATTGTGCAATCTCTTG	CAGAGAAATTC	1260

Wit-rood boontje ND	GACCCCTCTGGAGACCCCTCTGGATGTGGAACCTCAAGAGGATGGCTTCCTTTCTTGGACTT		1319
RIL29 ND	GACCCCTCTGGAGACCCCTCTGGATGTGGAACCTCAAGAGGATGGCTTCCTTTCTTGGACTT		1319
1533-15 SD	GACCCCTCTGGAGACCCCTCTGGATGTGGAACCTCAAGAGGATGGCTTCCTTTCTTGGACTT		1320
RIL81 RD	GACCCCTCTGGAGACCCCTCTGGATGTGGAACCTCAAGAGGATGGCTTCCTTTCTTGGACTT		1320
Othello RD	GACCCCTCTGGAGACCCCTCTGGATGTGGAACCTCAAGAGGATGGCTTCCTTTCTTGGACTT		1320
Etna RD	GACCCCTCTGGAGACCCCTCTGGATGTGGAACCTCAAGAGGATGGCTTCCTTTCTTGGACTT		1320
G19833	GACCCCTCTGGAGACCCCTCTGGATGTGGAACCTCAAGAGGATGGCTTCCTTTCTTGGACTT		1320

Wit-rood boontje ND	GAAAAATGATTGATCAAGGGATTATTATTATTTCACTGAGATGAAAAAGGTGAAGTACTTA		1379
RIL29 ND	GAAAAATGATTGATCAAGGGATTATTATTATTTCACTGAGATGAAAAAGGTGAAGTACTTA		1379
1533-15 SD	GAAAAATGATTGATCAAGGGATTATTATTATTTCACTGAGATGAAAAAGGTGAAGTACTTA		1380
RIL81 RD	GAAAAATGATTGATCAAGGGATTATTATTATTTCACTGAGATGAAAAAGGTGAAGTACTTA		1380
Othello RD	GAAAAATGATTGATCAAGGGATTATTATTATTTCACTGAGATGAAAAAGGTGAAGTACTTA		1380
Etna RD	GAAAAATGATTGATCAAGGGATTATTATTATTTCACTGAGATGAAAAAGGTGAAGTACTTA		1380
G19833	GAAAAATGATTGATCAAGGGATTATTATTATTTCACTGAGATGAAAAAGGTGAAGTACTTA		1380

Wit-rood boontje ND	GGGATCAAGTGTCATAGATAAAATATGCTATCTATATATGGAACACCTAACTAAGCATCTA		1439
RIL29 ND	GGGATCAAGTGTCATAGATAAAATATGCTATCTATATATGGAACACCTAACTAAGCATCTA		1439
1533-15 SD	GGGATCAAGTGTCATAGATAAAATATGCTATCTATATATGGAACACCTAACTAAGCATCTA		1440
RIL81 RD	GGGATCAAGTGTCATAGATAAAATATGCTATCTATATATGGAACACCTAACTAAGCATCTA		1440
Othello RD	GGGATCAAGTGTCATAGATAAAATATGCTATCTATATATGGAACACCTAACTAAGCATCTA		1440
Etna RD	GGGATCAAGTGTCATAGATAAAATATGCTATCTATATATGGAACACCTAACTAAGCATCTA		1440
G19833	GGGATCAAGTGTCATAGATAAAATATGCTATCTATATATGGAACACCTAACTAAGCATCTA		1440

Wit-rood boontje ND	GCTCTAGCTAGATGTACACACAAATCCAAGGTCTCTTCTCTTTTTCCTTTTCCTTTTCCT		1499
RIL29 ND	GCTCTAGCTAGATGTACACACAAATCCAAGGTCTCTTCTCTTTTTCCTTTTCCTTTTCCT		1499
1533-15 SD	GCTCTAGCTAGATGTACACACAAATCCAAGGTCTCTTCTCTTTTTCCTTTTCCTTTTCCT		1500
RIL81 RD	GCTCTAGCTAGATGTACACACAAATCCAAGGTCTCTTCTCTTTTTCCTTTTCCTTTTCCT		1500
Othello RD	GCTCTAGCTAGATGTACACACAAATCCAAGGTCTCTTCTCTTTTTCCTTTTCCTTTTCCT		1500
Etna RD	GCTCTAGCTAGATGTACACACAAATCCAAGGTCTCTTCTCTTTTTCCTTTTCCTTTTCCT		1500
G19833	GCTCTAGCTAGATGTACACACAAATCCAAGGTCTCTTCTCTTTTTCCTTTTCCTTTTCCT		1500

Wit-rood boontje ND	TTCCCC		1505
RIL29 ND	TTCCCC		1505
1533-15 SD	TTCCCC		1506
RIL81 RD	TTCCCC		1506
Othello RD	TTCCCC		1506
Etna RD	TTCCCC		1506
G19833	TTCCCC		1506

in PA synthesis in the seed coat. R2R3-MYBs promote transcription of genes encoding PA synthesis enzymes when they interact with bHLH and WD40 proteins and form ternary protein complexes (Xu et al. 2015). *MtPAR* regulates ANR and LAR expression during seed development and mutants exhibit large reductions in the PA content in their mature seeds with no change in anthocyanin levels compared to the wild type, indicating that this MYB factor is specifically involved in regulating PA biosynthesis (Verdier et al. 2012). *MtPAR* is a homolog of *AtTT2*, the key regulator of PA biosynthesis in *Arabidopsis* seeds (Liu et al. 2014). Pang et al. (2007) reported that the highest level of *MtANR* expression occurred in *Medicago* seed coats, and that lower levels of expression were found in flowers and immature pods, and no expression occurred in other tissues. However, low levels of *MtLAR* expression

were detected in *Medicago* flowers, pods, and seed coats (Pang et al. 2007).

Gene transcription is mediated by the enzyme RNA polymerase, which is recruited to a promoter region by general transcription factors. Its activity is modified by interactions with a multi-protein complex that includes transcriptional activators that bind DNA through DNA-binding domains and interact with the transcription complex through transcriptional activation domains (Ptashne 1988). Therefore, the specific amino acid sequence of the transcriptional activator plays a crucial role in its functionality and mutations in critical interaction domains may render it incapable of functioning in the transcription initiation complex (Figure S3). Amino acid modifications in the C-terminal regions of transcriptional activators that change the net charge of this protein region have been linked to changes in their activation

Fig. 3 Multiple sequence alignment of R2R3-MYB proteins and putative R2R3-MYB proteins encoded by the gene homologs of *Phvul.010G130600* in *P. vulgaris* G19833 showing characteristic motifs. The alignment includes putative translations of homologs of *Phvul.010G130600*: in ‘Wit-rood-boontje’ [parent A with the non-darkening (ND) seed coat phenotype], 1533-15 [parent B with slow darkening (SD) seed coat phenotype], RIL81 [regular darkening (RD) line], RIL29 (ND line), *MtPAR* (*Medtr8G020490* in *M. truncatula*), *GmMYB6* (*Glyma.16G007100* in *Glycine max*), *GmMYB8* (*Glyma.07G037700* in *Glycine max*), *CcMYB8* (*LOC109802394* in *Cajanus cajan*), *MpC1* (*C1* in *Mucuna pruriens*), and *VrWER-Like* (*LOC106773092* in *Vigna radiata*). Motifs characteristic of MYB are identified, including the primary structure R2 and R3 DNA-binding domains [-W-(X19)-W-(X19)-W-] and [-F/I-(X18)-W-(X18)-W-]; in which W represents tryptophan and X represents any amino acid. [D/E]Lx₂[R/K]x₃Lx₆Lx₃R is the bHLH-interaction motif located on helices 1 and 2 of the R3 repeat; The EDLL motif and amino acid residues in the transactivation domain are shown, containing characteristic E (glutamic acid), D (aspartic acid), and L (leucine) amino acids

Fig. 4 Neighbor-joining cluster analysis of putative R2R3-MYB proteins encoded by gene homologs of *Phvul.010G130600* in *P. vulgaris* G19833. The cluster includes the *Phvul.010G130600* protein sequence: in 1533-15 [parent B with the slow darkening (SD) seed coat phenotype] and RIL81 (SD line), the *Phvul.010G130600* protein sequence: in ‘Wit-rood-boontje’ [parent A with the non-darkening (ND) seed coat phenotype] and RIL29 (ND line), LOC106773092

transcription factor WER-like in *Vigna radiata*, MpC1 anthocyanin regulatory C1 protein in *Mucuna pruriens*, LOC109802394 transcription factor MYB8 in *Cajanus cajan*, Glyma.16G007100 transcription repressor MYB6 isoform X1 and Glyma.07G037700 transcription factor MYB8 in *Glycine max*, and Medtr8G020490 MYB transcription factor *MtPAR* in *M. truncatula*

Table 2 Putative homologs for *Phvul.010G130600* MYB-like transcription factor in different species

Gene ID	Species	Description	Query cover (%)	Identity (%)
<i>LOC106773092</i>	<i>Vigna radiata</i>	Transcription factor WER-like	100	91
<i>LOC109802394</i>	<i>Cajanus cajan</i>	Transcription factor MYB8	100	83
<i>C1</i>	<i>Mucuna pruriens</i>	Anthocyanin regulatory C1 protein	100	82
<i>Glyma.16G007100</i>	<i>G. max</i>	Transcription repressor MYB6 isoform X1	100	82
<i>Glyma.07G037700</i>	<i>G. max</i>	Transcription factor MYB8	100	80
<i>Medtr8G020490</i>	<i>M. truncatula</i>	Transcription factor MYB	97	70

The table includes their annotations, BLAST %query coverage, and BLAST % identity

rates. Estruch et al. (1994) reported that a two-amino-acid substitution in the transcriptional activation domain of the P14 peptide changed the net charge from +4 to +2 and led to 38% reduction in its activity. Alternatively, a plant protein ethylene-responsive element binding factor (ERF2) with a modified transcriptional activation domain displayed enhanced activation, when fused to DNA-binding proteins, compared to the unmodified ERF2 sequence (Li et al. 2013).

Multiple transcriptional activation domains have been identified on the basis of their amino acid compositions, including negatively charged activation domains which are rich in D and E residues and represented by the GAL4, GCN4 regulatory proteins in yeast and the herpes simplex virus VP16 protein (Lowe et al. 1985; Sadowski et al. 1988); a glutamine-rich activation domain which contains multiple repetitions such as ‘QQQXXXQQQ’ and is present in the mammalian factor SP1 (Courey et al. 1989); proline-rich

activation domains which are composed of repetitions such as ‘PPPXXXPPP’ found in c-jun, AP2 and Oct-2 transcription factors of mammals (Mermod et al. 1989); an isoleucine-rich activation domain with repetitions ‘IIXXII,’ present in *Drosophila* transcription factor NTF-1 (Attardi and Tjian 1993); and metal-binding cysteine-rich activation domain present in adenovirus E1a protein (Martin et al. 1990).

Transcriptional activation and repression domains play key roles in a wide variety of biological functions by positively and negatively regulating genes involved in developmental and physiological processes in various plant species. Sainz et al. (1997), through extensive mutagenesis studies of ZmMYBC1, found that two amino acid residues, L 253 and E 262, in the C-terminus of this protein are necessary for transcriptional activity. *ZmMYBC1* encodes a R2R3-MYB protein that interacts with a bHLH protein (ZmR) to regulate

Fig. 5 Sequence alignment of an R2R3-MYB protein encoded by the gene *Phvul.010G130600* of the reference genome *P. vulgaris* G19833 reference genome, 'Wit-rood-boontje' (parent A with the non-darkening (ND) seed coat phenotype), and 1533-15 (parent B with slow darkening (SD) seed coat phenotype). [-W-(X19)-W-(X19)-W-] and [-F/I-(X18)-W-(X18)-W-] are the primary structure of R2 and R3 DNA-binding domains, in which W represents tryptophan and X represents any amino acid. [D/E]Lx₂[R/K]x₃Lx₆Lx₃R is the bHLH-interaction motif located on helices 1 and 2 of the R3 repeat. E (glutamic acid), D (aspartic acid), and L (leucine) amino acids in the transactivation domain represent the EDLL motif and negatively charged residues

anthocyanin synthesis in maize aleurone tissue (Cone et al. 1986; Goff et al. 1992; Paz-Ares et al. 1987; Rabinowicz et al. 1999). The C1-encoded protein had previously been shown to contain a negatively charged transcriptional activation domain in its C-terminal (Paz-Ares et al. 1987) which is important for the C1 function (Goff et al. 1991; Paz-Ares et al. 1990). Similarly, Docimo et al. (2016) proposed that regulatory function of *Solanum melongena* MYB1 in anthocyanin biosynthesis can be impaired by a deletion of the C-terminal region. Albert et al. (2015) reported that the negatively charged C-terminal domains of the R2R3-MYB proteins involved in anthocyanin leaf markings in *Trifolium repens* are necessary to activate transcription. Members of class II ERFs such as AtERF3, AtERF4, and AtERF7 to AtERF12 in *Arabidopsis*, NtERF3 in tobacco (*Nicotiana tabacum*), and GmERF4 in soybean were identified as transcriptional repressors with ERF-associated amphiphilic repression (EAR) motifs in their C-terminal region for active repression (Fujimoto et al. 2000; Ohta et al. 2001; McGrath et al. 2005; Song et al. 2005; Yang et al. 2005; Zhang et al. 2010). EAR motif contains two distinct conservation patterns: LxLxL and DLNxxP (Kagale et al. 2010). The R2R3-MYB, MYB27, was found to be an anthocyanin repressor that functions as part of the MBW complex and represses transcription through its EAR motif (Albert et al. 2014).

Recently, 'EDLL' motifs, comprised of negatively charged residues, aspartic acid (D) and glutamic acid (D) and hydrophobic leucine (L), have been identified as short

but strong transactivation motifs in the members of group IX of the APETALA2/ethylene response (AP2/ERF) transcription factors including *Arabidopsis* Ethylene Response Factor 98 (*AtERF98*) (Tiwari et al. 2012) and *AtERF96* (Chen et al. 2018). It has been suggested that the EDLL motif plays an important role for *AtERF96* to interact with MEDIATOR25, a subunit of the eukaryotic Mediator complex (Çevik et al. 2012), that functions as a transcriptional coactivator in *Arabidopsis* (Chen et al. 2018).

Aspartic acid and glutamic acid are among the most frequently occurring amino acids in wild-type transactivation domains (Ravarani et al. 2018). It is possible that the EDLL motif in the C-terminal region of the R2R3-MYB encoded by *Phvul.010G130600* functions as a transactivating domain that is required to activate transcription of genes that encode enzymes in the biosynthetic pathway of proanthocyanidin synthesis in *P. vulgaris* seed coat.

Among the genes that exhibit high levels of protein sequence identity with the *Phvul010G130600* gene, the gene *MtPAR* is of particular interest because it has been shown to affect PA biosynthesis. *MtPAR* is a R2R3-MYB transcription factor that plays a significant role in PA biosynthesis in *Medicago* seed coat tissue by regulating *ANR* and *LAR* expression during seed development (Pang et al. 2009; Verdier et al. 2012; Li et al. 2016). *Mtpar* mutants exhibited large reductions in the PA content of their mature seeds with no change in anthocyanin levels compared to the wild type, indicating that this MYB factor is specifically involved in

Fig. 6 *Phvul.010G130600* SNP-based marker and the marker screen of regular, slow (SD) and non-darkening (ND) recombinant inbred lines (RIL), breeding lines and varieties of bean. **a** PCR binding sites for PCR primers (Table 3) in regular darkening (RD) and non-darkening (ND) genotypes. The SNP is identified with a red box. **b** PCRs with genomic DNA from a ND genotype like ‘Wit-rood boontje’ did not produce an amplicon. PCRs with genomic DNA from a slow darkening (SD) parent 1533-15 produced a single DNA fragment with a size of 254 bp. **c** RILs88-101 are recombinant inbred lines from a ND ‘Wit-rood-boontje’ x SD 1533-15 cross. **d** Burdett, Othello, Windbreaker, Sequoia, PT11-9, Kodiak, Kimberly, La Paz, 16-NDP1, UI 114, and ME105 are Pinto bean varieties and lines. Etna, CBX 1148, C15HR009, Red Rider, OAC Racer, C13HR185, and 8184 (94CTCOOP-8184) are cranberry bean varieties and lines. Labels that are high-lighted red, yellow, and blue represent seeds with (regular darkening) RD, SD, and ND phenotypes, respectively (color figure online)

regulating PA biosynthesis (Verdier et al. 2012). Pang et al. 2007 reported that the highest level of *MtANR* expression occurred in *Medicago* seed coats, and that lower levels of expression were found in flowers and immature pods, and no expression occurred in other tissues. However, low levels of *MtLAR* expression were detected in *Medicago* flowers, pods,

and seed coats (Pang et al. 2007). *MtPAR* is a homolog of *AtTT2*, the key regulator of PA biosynthesis in *Arabidopsis* seeds (Liu et al. 2014).

We suggest that the MYB encoded by *Phvul.010G130600* is a PA activator that functions as part of the MBW complex and regulates transcription through its C-terminal EDLL motif. Further, we suggest that the MBW complex targets the seed coat PA biosynthesis genes encoding biosynthetic enzymes in the pathway such as *PvANR1* and *PvLAR* and this determines PHD in *P. vulgaris*. The SNP in the mutant *Phvul.010G130600* gene found in ‘Wit-rood boontje’ causes a translational frameshift so that the C-terminus in the MYB it encodes lacks the EDLL motif and cannot interact with the coactivator complex to activate transcription of *PvANR1* and *PvLAR* which ultimately leads to a reduction or cessation of proanthocyanidin synthesis when it is in a homozygous condition.

This proposal is consistent with the higher levels of proanthocyanidins that were detected in seed coats of darkening pinto and cranberry beans compared to the slow

Table 3 PCR primer sequences for STS marker associated with the gene for the ND seed coat phenotype in common bean based on *Phvul.010G130600*

STS marker	Primer Sequence (5'-3')	Linkage distance (cM) from <i>J</i> locus	T _m (°C)
PvJ-254	Forward: TCATCATTG TCACAATCTCTT GGC	0.0	63
	Reverse: GAGAAG AGACCTTGGATT TGTGTG		63

T_m DNA melting temperature

darkening and non-darkening beans (Beninger et al. 2005; Chen et al. 2015; Freixas Coutin et al. 2017). Proanthocyanidins are abundant in the brown seed coats of many plants such as *A. thaliana* (Dixon et al. 2005) and *ANR* plays a key role in the biosynthesis of PAs in that species (Wang et al. 2018). *PvANR1* was found to be strongly associated with seed coat PA accumulation in darkening cranberry beans (Freixas Coutin et al. 2017). Loss of function of *ANR* gave large reductions in both soluble and insoluble PAs in *M. truncatula* seed coat (Liu et al. 2016). *LAR* along with *ANR* was found to contribute to PA polymerization in pea seed coats (Ferraro et al. 2014). However, the role of *LAR* in PA biosynthesis in *Phaseolus* remains unanswered.

The seed coat develops from the two integuments or outer layers of cells surrounding the ovule, which are derived from maternal tissue. The inner integument forms the segment, and the outer forms the testa. This means that, segregation for seed coat color always occurs one generation after seed formation (Erfatpour et al. 2018). As shown in Figure S4, the F_1 seed with a seed coat genotype that is *jj* at the ND locus, has the same seed coat color as the female parent ‘Wit-rood boontje.’ For the F_2 seed the seed coat tissue genotype is *Jj*. Therefore, F_2 seeds have the darkening (RD) seed coat phenotype, but the embryo genotypes occur in the expected 1:2:1 ratio for *jj/Jj/JJ* individuals. Segregation for PHD can be observed in the F_3 . The dominant gene-based marker that was developed in the current study would enable bean breeders to differentiate F_2 seeds or seedlings carrying the dominant allele (*JJ* and *Jj*) from those that are homozygous recessive (*jj*) genotypes in early generations (including F_2 populations). In addition, this determination could be done at any stage of plant development, thus allowing selection at a seedling stage, rather than waiting to screen the phenotypic segregation in the UV treated F_3 seed (Figure S4).

Acknowledgements The authors wish to acknowledge the financial support from the Ontario Bean Growers, Agriculture and Agri-Food Canada, the Natural Science and Engineering Research Council of Canada, the Ontario Ministry of Research and Innovation, the Ontario Ministry for Agriculture, Food and Rural affairs, and also the technical assistance of Tom Smith and Lindsay Schram.

Author Contribution statement ME and KPP designed the experiment. ME conducted the experiments. ME wrote the manuscript and both authors made edits.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aasland R, Gibson TJ, Stewart AF (1995) The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem Sci* 20:56–59
- Akagi T, Ikegami A, Tsujimoto T, Kobayashi S, Sato A, Kono A, Yonemori K (2009) DkMyb4 is a myb transcription factor involved in proanthocyanidin biosynthesis in persimmon fruit. *Plant Physiol* 151:2028–2045
- Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, Brendolise C (2014) A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *Plant Cell* 26:962–980
- Albert NW, Griffiths AG, Cousins GR, Verry IM, Williams WM (2015) Anthocyanin leaf markings are regulated by a family of *R2R3-MYB* genes in the genus *Trifolium*. *New Phytol* 205:882–893
- Amador V, Monte E, Garcia-Martinez JL, Prat S (2001) Gibberellins signal nuclear import of PHOR1, a photoperiod responsive protein with homology to *Drosophila* armadillo. *Cell* 106:343–354
- Ambawat S, Sharma P, Yadav NR, Yadav RC (2013) MYB transcription factor genes as regulators for plant responses: an overview. *Physiol Mol Biol Plants* 19:307–321
- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. DPF <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed 21 Nov 2018
- Assmann SM (2002) Heterotrimeric and unconventional GTP binding proteins in plant signalling. *Plant Cell* 14:S355–S373
- Attardi LD, Tjian R (1993) Drosophila tissue-specific transcription factor NTF-1 contains a novel isoleucine-rich activation motif. *Genes Dev* 7:1341–1353
- Auger B, Baron C, Lucas M, Vautrin S, Berges H, Chalhoub B et al (2009) Brassica orthologs from *BANYULS* belong to a small multigene family, which is involved in procyanidin accumulation in the seed. *Planta* 230:1167–1183
- Bassett MJ (1996a) New genes, *stp* and *stp^{hbw}*, for flower and seed coat pattern in common bean. *J Am Soc Hort Sci* 121:388–392
- Bassett MJ (1996b) The margo (*mar*) seed coat color gene is a synonym for the joker (*j*) locus in common bean. *J Am Soc Hort Sci* 121:1028–1031
- Bassett MJ (2007) Genetics of seed coat color and pattern in common bean. *Plant Breed Rev* 28:239–315
- Baudry A, Heim MA, Dubreucq B, Caboche M, Weisshaar B, Lepiniec L (2004) TT2, TT8, and TTG1 synergistically specify the expression of *BANYULS* and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J* 39:366–380
- Beninger CW, Gu L, Prior RL, Junk DC, Vandenberg A, Bett K (2005) Changes in polyphenols of the seed coat during the after-darkening process in pinto bean (*Phaseolus vulgaris* L.). *J Agric Food Chem* 53:7777–7782
- Bennett RN, Wallsgrove RM (1994) Secondary metabolites in plant defence mechanisms. *New Phytol* 127:617–633
- Buer CS, Imin N, Djordjevic MA (2010) Flavonoids: new roles for old molecules. *J Integr Plant Biol* 52:98–111
- Cetin ES (2014) Induction of secondary metabolite production by UV-C radiation in *Vitis vinifera* L. *Öküzgözü callus cultures*. *Biol Res* 47:37
- Çevik V, Kidd BN, Zhang P, Hill C, Kiddle S, Denby KJ et al (2012) MEDIATOR25 acts as an integrative hub for the regulation of jasmonate-responsive gene expression in *Arabidopsis*. *Plant Physiol* 160:541–555

- Chen Y, Yang X, He K, Liu M, Li J, Gao Z et al (2006) The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol Biol* 60:107–124
- Chen PX, Tang Y, Marcone MF, Pauls PK, Zhang B, Liu R, Tsao R (2015) Characterization of free, conjugated and bound phenolics and lipophilic antioxidants in regular- and non-darkening cranberry beans (*Phaseolus vulgaris* L.). *Food Chem* 185:298–308
- Chen C-Y, Lin P-H, Chen K-H, Cheng Y-S (2018) Structural insights into DNA sequence recognition by *Arabidopsis* ETHYLENE RESPONSE FACTOR 96. *bioRxiv* 302299. Doi: <https://doi.org/10.1101/302299>
- Cho SK, Chung HS, Ryu MY, Park MJ, Lee MM, Bahk YY et al (2006) Heterologous expression and molecular and cellular characterization of *CaPUB1* encoding a hot pepper U-box E3 ubiquitin ligase homolog. *Plant Physiol* 142:1664–1682
- Cipollini ML, Levey DJ (1997) Secondary metabolites of fleshy vertebrate-dispersed fruits: adaptive hypotheses and implications for seed dispersal. *Am Nat* 150:346–372
- Clegg MT, Durbin ML (2003) Tracing floral adaptations from ecology to molecules. *Nat Rev Genet* 4:215–260
- Cone KC, Burr FA, Burr B (1986) Molecular analysis of the maize anthocyanin regulatory locus *C1*. *PNAS* 83:9631–9635
- Courey AJ, Holtzman DA, Jackson SP, Tjian R (1989) Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. *Cell* 59:827–836
- Cushnie TP, Lamb AJ (2005) Antimicrobial activity of flavonoids. *Int J Antimicrob Agents* 26:343–356
- Davies KM, Schwinn KE (2006) Molecular biology and biotechnology of flavonoid biosynthesis. In: Andersen ØM, Markham KR (eds) *Flavonoids: chemistry, biochemistry and applications*. CRC Press, Boca Raton, pp 143–218
- Debeaujon I, Peeters AJM, Léon-Kloosterziel KM, Koornneef M (2001) The *TRANSPARENT TESTA12* gene of *Arabidopsis* encodes a multidrug secondary transporter-like protein required for flavonoid sequestration in vacuoles of the seed coat endothelium. *Plant Cell* 13:853–872
- Debeaujon I, Nesi N, Perez P, Devic M, Grandjean O, Caboche M, Lepiniec L (2003) Proanthocyanidin-accumulating cells in *Arabidopsis* testa: regulation of differentiation and role in seed development. *Plant Cell* 15:2514–2531
- Dixon RA, Xie DY, Sharma SB (2005) Proanthocyanidins—a final frontier in flavonoid research? *New Phytol* 165:9–28
- Docimo T, Francese G, Ruggiero A, Batelli G, De Palma M, Basolino L et al (2016) Phenylpropanoids accumulation in eggplant fruit: characterization of biosynthetic genes and regulation by a MYB transcription factor. *Front Plant Sci* 6:1233
- Doelling JH, Walker JM, Friedman EM, Thompson AR, Vierstra RD (2002) The APG8/12-activating enzyme APG7 is required for proper nutrient recycling and senescence in *Arabidopsis thaliana*. *J Biol Chem* 277:33105–33114
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Du H, Yang SS, Liang Z, Feng BR, Liu L, Huang YB et al (2012) Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biol* 12:106
- Elsadr HT, Wright LC, Pauls KP, Bett KE (2011) Characterization of seed coat post harvest darkening in common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet* 123:1267–1472
- Erfatpour M, Navabi A, Pauls KP (2018) Mapping the non-darkening trait from ‘Wit-rood boontje’ in bean (*Phaseolus vulgaris*). *Theor Appl Genet* 131:1331–1343
- Estruch JJ, Kadwell S, Merlin E, Crossland L (1994) Cloning and characterization of a maize pollen-specific calcium-dependent calmodulin-independent protein kinase. *PNAS* 91:8837–8841
- Felicetti E, Song Q, Jia G, Cregan P, Bett KE, Miklas PN (2012) Simple sequence repeats linked with slow darkening trait in pinto bean discovered by single nucleotide polymorphism assay and whole genome sequencing. *Crop Sci* 52:1600–1608
- Ferraro K, Jin AL, Nguyen T-D, Reinecke DM, Ozga JA, Ro D-K (2014) Characterization of proanthocyanidin metabolism in pea (*Pisum sativum*) seeds. *BMC Plant Biol* 14:238
- Freixas Coutin JA, Munholland S, Silva A, Subedi S, Lukens L, Crosby WL et al (2017) Proanthocyanidin accumulation and transcriptional responses in the seed coat of cranberry beans (*Phaseolus vulgaris* L.) with different susceptibility to post-harvest darkening. *BMC Plant Biol* 17:89
- Fujimoto SY, Ohta M, Usui A, Shinshi H, Ohme-Takagi M (2000) *Arabidopsis* ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. *Plant Cell* 12:393–404
- Goff SA, Cone KC, Fromm ME (1991) Identification of functional domains in the maize transcriptional activator C1: comparison of wild-type and dominant inhibitor proteins. *Genes Dev* 5:298–309
- Goff SA, Cone KC, Chandler VL (1992) Functional analysis of the transcriptional activator encoded by the maize *B* gene: evidence for a direct functional interaction between two classes of regulatory proteins. *Genes Dev* 6:864–875
- He X-Z, Li W-S, Blount JW, Dixon RA (2008) Regioselective synthesis of plant (iso)flavone glycosides in *Escherichia coli*. *Appl Microbiol Biotechnol* 80:253–260
- Heise A, Lippok B, Kirsch C, Hahlbrock K (2002) Two immediately pathogen-responsive members of the *AtCMPG* gene family in *Arabidopsis thaliana* and the W-box-containing elicitor-response element of *AtCMPG1*. *PNAS* 13:9049–9054
- Höglund PJ, Nordström KJV, Schiöth HB, Fredriksson R (2011) The solute carrier families have a remarkably long evolutionary history with the majority of the human families present before divergence of Bilateral species. *Mol Biol Evol* 28:1531–1541
- Hrazdina G, Wagner GJ (1985) Metabolic pathways as enzyme complexes: evidence for the synthesis of phenylpropanoids and flavonoids on membrane associated enzyme complexes. *Arch Biochem Biophys* 237:88–100
- Ichino T, Fuji K, Ueda H, Takahashi H, Koumoto Y, Takagi J et al (2014) GFS9/TT9 contributes to intracellular membrane trafficking and flavonoid accumulation in *Arabidopsis thaliana*. *Plant J* 80:410–423
- Jia L, Clegg MT, Jiang T (2004) Evolutionary dynamics of the DNA binding domains in putative R2R3-MYB genes identified from rice subspecies *indica* and *japonica* genomes. *Plant Physiol* 134:575–585
- Jiang C, Gu J, Chopra S, Gu X, Peterson T (2004a) Ordered origin of the typical two- and three-repeat Myb genes. *Gene* 326:13–22
- Jiang C, Gu X, Peterson T (2004b) Identification of conserved gene structures and carboxy-terminal motifs in the Myb gene family of *Arabidopsis* and *Oryza sativa* L. ssp. *indica*. *Genome Biol* 5:R46
- Jiménez M, García-Carmona F (1999) Oxidation of the flavonol quercetin by polyphenol oxidase. *J Agric Food Chem* 47:56–60
- Junk-Knievel DC, Vandenberg A, Bett KE (2008) Slow darkening in pinto bean (*Phaseolus vulgaris* L.) seed coats is controlled by a single major gene. *Crop Sci* 48:189–193
- Kabe Y, Yamada J, Uga H, Yamaguchi Y, Wada T, Handa H (2005) NF-Y is essential for the recruitment of RNA polymerase II and inducible transcription of several CCAAT box containing genes. *Mol Cell Biol* 25:512–522
- Kagale S, Links MG, Rozwadowski K (2010) Genome-wide analysis of ethylene responsive element binding factor-associated amphiphilic repression motif-containing transcriptional regulators in *Arabidopsis*. *Plant Physiol* 152:1109–1134

- Kirsch C, Logemann E, Lippok B, Schmelzer E, Hahlbrock K (2001) A highly specific pathogen-responsive promoter element from the immediate-early activated *CMPG1* gene in *Petroselinum crispum*. *Plant J* 26:217–227
- Kitamura S, Shikazono N, Tanaka A (2004) *TRANSPARENT TESTA 19* is involved in the accumulation of both anthocyanins and proanthocyanidins in *Arabidopsis*. *Plant J* 37:104–114
- Koen B, Ruth V, Guido V, Johannes VS (2005) Induction of cancer cell apoptosis by flavonoids is associated with their ability to inhibit fatty acid synthase activity. *J Biol Chem* 280:5636–5645
- Kovinich N, Saleem A, Arnason JT, Miki B (2012) Identification of two anthocyanidin reductase genes and three red-brown soybean accessions with reduced anthocyanidin reductase 1 mRNA, activity, and seed coat proanthocyanidin amounts. *J Agric Food Chem* 60:574–584
- Koyama K, Numata M, Nakajima I, Goto-Yamamoto N, Matsumura H, Tanaka N (2014) Functional characterization of a new grapevine MYB transcription factor and regulation of proanthocyanidin biosynthesis in grapes. *J Exp Bot* 65:4433–4449
- Laukkanen H, Haggman H, Soppela SK, Hohtola A (1999) Tissue browning of in vitro cultures of scots pine: role of peroxidase and polyphenol oxidase. *Physiol Plant* 106:337–343
- Lepiniec L, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, Nesi N, Caboche M (2006) Genetics and biochemistry of seed flavonoids. *Annu Rev Plant Biol* 57:405–430
- Li ZH, Wang Q, Ruan X, Pan CD, Jiang DA (2010) Phenolics and plant allelopathy. *Molecules* 15:8933–8952
- Li J, Blue R, Zeitler B, Strange TL, Pearl JR, Huizinga DH et al (2013) Activation domains for controlling plant gene expression using designed transcription factors. *Plant Biotechnol J* 11:671–680
- Li P, Chen B, Zhang G, Chen L, Dong Q, Wen J et al (2016) Regulation of anthocyanin and proanthocyanidin biosynthesis by *Medicago truncatula* bHLH transcription factor MtTT8. *New Phytol* 210:905–921
- Liu L, White MJ, MacRae TH (1999) Transcription factors and their genes in higher plants functional domains, evolution and regulation. *Eur J Biochem* 262:247–257
- Liu C, Jun JH, Dixon RA (2014) MYB5 and MYB14 play pivotal roles in seed coat polymer biosynthesis in *Medicago truncatula*. *Plant Physiol* 165:1424–1439
- Liu C, Wang X, Shulaev V, Dixon RA (2016) A role for leucoanthocyanidin reductase in the extension of proanthocyanidins. *Nat Plants* 2:16182
- Lowe KS, Taylor DB, Ryan PL, Paterson KP (1985) Plant regeneration via organogenesis and embryogenesis in the maize inbred line B73. *Plant Sci* 41:125–132
- Lu N, Roldan M, Dixon RA (2017) Characterization of two TT2-type MYB transcription factors regulating proanthocyanidin biosynthesis in tetraploid cotton, *Gossypium hirsutum*. *Planta* 246:323–335
- Marles MAS, Vandenberg A, Bett KE (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
- Martin KJ, Lillie JW, Green MR (1990) Evidence for interaction of different eukaryotic transcriptional activators with distinct cellular targets. *Nature* 346:147–152
- Masclaux-Daubresse C, Clément G, Anne P, Routaboul JM, Guiboileau A, Soulay F et al (2014) Stitching together the multiple dimensions of autophagy using metabolomics and transcriptomics reveals impacts on metabolism, development, and plant responses to the environment in *Arabidopsis*. *Plant Cell* 26:1857–1877
- McClellan PE, Lee RK, Otto C, Gepts G, Bassett MJ (2002) Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean (*Phaseolus vulgaris* L.). *J Hered* 93:148–152
- McGrath KC, Dombrecht B, Manners JM, Schenk PM, Edgar CI, Maclean DJ et al (2005) Repressor- and activator-type ethylene response factors functioning in jasmonate signaling and disease resistance identified via a genome-wide screen of *Arabidopsis* transcription factor gene expression. *Plant Physiol* 139:949–959
- Mermod N, O'Neill EA, Kelly TJ, Tjian R (1989) The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. *Cell* 58:741–753
- Minina EA, Moschou PN, Vetukuri RR, Sanchez-Vera V, Cardoso C, Liu Q et al (2018) Transcriptional stimulation of rate-limiting components of the autophagic pathway improves plant fitness. *J Exp Bot* 69:1415–1432
- Mishra D, Singh RK, Srivastava RK, Dubey SR (2013) Ethnomedicinal plants used to cure the gynaecological disorders by ethnic populace of Sitapur district, Uttar Pradesh, India. *Med Plants Int J Phytomed Relate Ind* 5:238–245
- Mol J, Grotewold E, Koes R (1998) How genes paint flowers and seeds. *Trends Plant Sci* 3:212–217
- Nasar-Abbas SM, Siddique KHM, Plummer JA, White PF, Harris D, Dods K, D'Antuono M (2009) Faba bean (*Vicia faba* L.) seeds darken rapidly and phenolic content falls when stored at higher temperature, moisture and light intensity. *LWT-Food Sci Technol* 42:1703–1711
- Nesi N, Jond C, Debeaujon I, Caboche M, Lepiniec 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
- Ogata K, Kaneishii C, Sasaki M, Hatanaka H, Nagadoi A, Enari M et al (1996) The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nat Struct Mol Biol* 3:178–187
- Ohlrogge J, Browse J (1995) Lipid biosynthesis. *Plant Cell* 7:957–970
- Ohta M, Matsui K, Hiratsu K, Shinshi H, Ohme-Takagi M (2001) Repression domains of class II ERF transcriptional repressors share an essential motif for active repression. *Plant Cell* 13:1959–1968
- Pan MH, Lai CS, Ho CT (2010) Anti-inflammatory activity of natural dietary flavonoids. *Food Funct* 1:15–31
- Pang Y, Peel GJ, Wright E, Wang Z, Dixon RA (2007) Early steps in proanthocyanidin biosynthesis in the model legume *Medicago truncatula*. *Plant Physiol* 145:601–615
- Pang Y, Wenger JP, Saathoff K, Peel GJ, Wen J, Huhman D et al (2009) A WD40 repeat protein from *Medicago truncatula* is necessary for tissue-specific anthocyanin and proanthocyanidin biosynthesis but not for trichome development. *Plant Physiol* 151:1114–1129
- Paz-Ares J, Ghosal D, Wienand U, Peterson PA, Saedler H (1987) The regulatory *c1* locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J* 6:3553–3558
- Paz-Ares J, Ghosal D, Saedler H (1990) Molecular analysis of the *C1-I* allele from *Zea mays*: a dominant mutant of the regulatory *C1* locus. *EMBO J* 9:315–321
- Peleg H, Gacon K, Schlich P, Noble AC (1999) Bitterness and astringency of flavan-3-ol monomers, dimers and trimers. *J Sci Food Agric* 79:1123–1128
- Pourcel L, Routaboul JM, Kerhoas L, Caboche M, Lepiniec L, Debeaujon I (2005) *TRANSPARENT TESTA10* encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in *Arabidopsis* seed coat. *Plant Cell* 17:2966–2980
- Ptashne M (1988) How eukaryotic transcriptional activators work. *Nature* 335:683–689

- Quattrocchio F, Verweij W, Kroon A, Spelt C, Mol J, Koes R (2006) PH4 of Petunia is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell* 18:1274–1291
- Rabinowicz P, Braun E, Wolfe A, Bowen B, Grotewold E (1999) Maize *R2R3 Myb* genes: sequence analysis reveals amplification in the higher plants. *Genetics* 153:427–444
- Ravarani CNJ, Erkina TY, De Baets G, Dudman DC, Erkin AM, Babu MM (2018) High-throughput discovery of functional disordered regions: investigation of transactivation domains. *Mol Syst Biol* 14:e8190
- Rayson S, Arciga-Reyes L, Wootton L, Zabala MD, Truman W, Graham N et al (2012) A role for nonsense-mediated mRNA decay in plants: pathogen responses are induced in *Arabidopsis thaliana* NMD mutants. *PLoS ONE* 7:e31917
- Robichaud JL, Noble AC (1990) Astringency and bitterness of selected phenolics in wine. *J Sci Food Agric* 53:343–353
- Robles LM, Wampole JS, Christians MJ, Larsen PB (2007) *Arabidopsis enhanced ethylene response 4* encodes an EIN3-interacting TFIIID transcription factor required for proper ethylene response, including *ERF1* induction. *J Exp Bot* 58:2627–2639
- Sadowski I, Ma J, Triezenberg S, Ptashne M (1988) GAL4-VP16 is an unusually potent transcriptional activator. *Nature* 335:563–564
- Sainz MB, Goff SA, Chandler VL (1997) Extensive mutagenesis of a transcriptional activation domain identifies single hydrophobic and acidic amino acids important for activation *in vivo*. *Mol Cell Biol* 17:115–122
- Samuel MA, Salt JN, Shiu SH, Goring DR (2006) Multifunctional arm repeat domains in plants. *Int Rev Cytol* 253:1–26
- Samuel MA, Mudgil Y, Salt JN, Delmas F, Ramachandran S, Chil-elli A, Goring DR (2008) Interactions between the S-domain receptor kinases and AtPUB-ARM E3 ubiquitin ligases suggest a conserved signaling pathway in *Arabidopsis*. *Plant Physiol* 147:2084–2095
- Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J et al (2014) A reference genome for common bean and genome-wide analysis of dual domestications. *Nat Genet* 46:707–713
- Schwinn K, Venail J, Shang Y, Mackay S, Alm V, Butelli E et al (2006) A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell* 18:831–851
- Shimada N, Aoki T, Sato S, Nakamura Y, Tabata S, Ayabe S (2003) A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy(iso)flavonoids in *Lotus japonicus*. *Plant Physiol* 31:941–951
- Smýkal P, Vernoud V, Blair MW, Soukup A, Thompson RD (2014) The role of the testa during development and in establishment of dormancy of the legume seed. *Front Plant Sci* 5:351
- Song CP, Agarwal M, Ohta M, Guo Y, Halfter U, Wang P, Zhu JK (2005) Role of an *Arabidopsis* AP2/EREBP-type transcriptional repressor in abscisic acid and drought stress responses. *Plant Cell* 17:2384–2396
- Stafstrom JP (2008) Expression patterns of *Arabidopsis* *DRG* genes: promoter:GUS fusions, quantitative RT-PCR and patterns of protein accumulation in response to environmental stresses. *Int J Plant Sci* 169:1046–1056
- Stone SL, Anderson EM, Mullen RT, Goring DR (2003) ARC1 is an E3 ubiquitin ligase and promotes the ubiquitination of proteins during the rejection of self-incompatible Brassica pollen. *Plant Cell* 15:885–898
- Stracke R, Werber M, Weissshaar B (2001) The *R2R3-MYB* gene family in *Arabidopsis thaliana*. *Curr Opin Plant Biol* 4:447–456
- Takahama U (2004) Oxidation of vacuolar and apoplastic phenolics substrates by peroxidase: physiological significance of the oxidation reactions. *Phytochem Rev* 3:207–219
- Takahama U, Hirota S (2000) Deglucosidation of quercetin glucosides to the aglycone and formation of antifungal agents by peroxidase-dependent oxidation of quercetin on browning of onion scales. *Plant Cell Physiol* 41:1021–1029
- Tanner GJ, Francki KT, Abrahams S, Watson JM, Larkin PJ, Ashton AR (2003) Proanthocyanidin biosynthesis in plants: purification of legume leucoanthocyanidin reductase and molecular cloning of its cDNA. *J Biol Chem* 278:31647–31656
- Thompson AR, Doelling JH, Suttangkakul A, Vierstra RD (2005) Autophagic nutrient recycling in *Arabidopsis* directed by the ATG8 and ATG12 conjugation pathways. *Plant Physiol* 138:2097–2110
- Tiwari SB, Belachew A, Ma SF, Young M, Ade J, Shen Y et al (2012) The EDLL motif: a potent plant transcriptional activation domain from AP2/ERF transcription factors. *Plant J* 70:855–865
- van Dam RM, Naidoo N, Landberg R (2013) Dietary flavonoids and the development of type 2 diabetes and cardiovascular diseases: review of recent findings. *Curr Opin Lipidol* 24:25–33
- Vaughn KC, Duke SO (1984) Function of polyphenol oxidases in higher plants. *Physiol Plant* 60:106–112
- Verdier J, Zhao J, Torres-Jerez I, Ge S, Liu C, He X et al (2012) MtPAR MYB transcription factor acts as an on switch for proanthocyanidin biosynthesis in *Medicago truncatula*. *PNAS* 109:1766–1771
- Wan H, Zhang J, Song T, Tian J, Yao Y (2015) Promotion of flavonoid biosynthesis in leaves and calli of ornamental crabapple (*Malus sp.*) by high carbon to nitrogen ratios. *Front Plant Sci* 6:673
- Wang P, Zhang L, Jiang X, Dai X, Xu L, Li T et al (2018) Evolutionary and functional characterization of leucoanthocyanidin reductases from *Camellia sinensis*. *Planta* 247:139–154
- Wei W, Zhang Y-Q, Tao J-J, Chen H-W, Li Q-T, Zhang W-K et al (2015) The Alfin-like homeodomain finger protein AL5 suppresses multiple negative factors to confer abiotic stress tolerance in *Arabidopsis*. *Plant J* 81:871–883
- Wilson ZA, Morroll SM, Dawson J, Swarup R, Tighe PJ (2001) The *Arabidopsis* MALE STERILITY1 (MS1) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. *Plant J* 28:27–39
- Wink M (2008) Plant secondary metabolism: diversity, function and its evolution. *Front Physiol Nat Prod Commun* 3:1205–1216
- Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol* 126:485–493
- Xu W, Grain D, Bobet S, Le Gourrierc J, Thévenin J, Kelemen Z et al (2014) Complexity and robustness of the flavonoid transcriptional regulatory network revealed by comprehensive analyses of MYB-bHLH-WDR complexes and their targets in *Arabidopsis* seed. *New Phytol* 202:132–144
- Xu W, Dubos C, Lepiniec L (2015) Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci* 20:176–185
- Yang Z, Tian L, Latoszek-Green M, Brown D, Wu K (2005) *Arabidopsis* ERF4 is a transcriptional repressor capable of modulating ethylene and abscisic acid responses. *Plant Mol Biol* 58:585–596
- Yenjerla M, Panopoulos A, Reynaud C, Fotedar R, Margolis RL (2013) TD-60 is required for interphase cell cycle progression. *Cell Cycle* 12:837–841
- Zhang G, Chen M, Chen X, Xu Z, Li L, Guo J, Ma Y (2010) Isolation and characterization of a novel EAR-motif-containing gene GmERF4 from soybean (*Glycine max* L.). *Mol Biol Rep* 37:809–818
- Zhang K, Lu K, Qu C, Liang Y, Wang R, Chai Y et al (2013) Gene silencing of *BnTT10* family genes causes retarded pigmentation

- and lignin reduction in the seed coat of *Brassica napus*. PLoS ONE 8:e61247
- Zhao J (2015) Flavonoid transport mechanisms: how to go, and with whom. Trends Plant Sci 20:576–585
- Zhao J, Dixon R (2009) MATE transporters facilitate vacuolar uptake of epicatechin 3'-O-glucoside for proanthocyanidin biosynthesis in *Medicago truncatula* and *Arabidopsis*. Plant Cell 21:2323–2340
- Zhao J, Pang Y, Dixon RA (2010) The mysteries of proanthocyanidin transport and polymerization. Plant Physiol 153:437–443
- Zhao L, Zhang Q, Ma W, Tian F, Shen H, Zhou M (2017) A combination of quercetin and resveratrol reduces obesity in high-fat diet-fed rats by modulation of gut microbiota. Food Funct 8:4644–4656
- Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. Plant Physiol 136:2621–2632

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.