



Loss-of-Function Mutation of Soybean R2R3 MYB Transcription Factor Dilutes Tawny Pubescence Color

Fan Yan^{1*}, Stephen M. Githiri², Yajing Liu¹, Yu Sang¹, Qingyu Wang¹ and Ryoji Takahashi³

¹ College of Plant Science, Jilin University, Changchun, China, ² Horticulture Department, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya, ³ NARO Institute of Crop Science, Tsukuba, Japan

OPEN ACCESS

Edited by:

Roberto Papa,
Marche Polytechnic University, Italy

Reviewed by:

Francesca Sparvoli,
Istituto di Biologia E Biotecnologia
Agraria (IBBA), Italy
Frédéric Marsolais,
Agriculture and Agri-Food Canada
(AAFC), Canada

*Correspondence:

Fan Yan
yanfan@jlu.edu.cn

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 29 September 2019

Accepted: 24 December 2019

Published: 31 January 2020

Citation:

Yan F, Githiri SM, Liu Y, Sang Y,
Wang Q and Takahashi R (2020)
Loss-of-Function Mutation
of Soybean R2R3 MYB
Transcription Factor Dilutes
Tawny Pubescence Color.
Front. Plant Sci. 10:1809.
doi: 10.3389/fpls.2019.01809

Pubescence color of soybean is controlled by two genes, *T* and *Td*. In the presence of a dominant *T* allele, dominant and recessive alleles of the *Td* locus generate tawny and light tawny (or near-gray) pubescence, respectively. Flavones, responsible for pubescence color, are synthesized *via* two copies of flavone synthase II genes (*FNS II-1* and *FNS II-2*). This study was conducted to map and clone the *Td* gene. Genetic and linkage analysis using an F₂ population and F₃ families derived from a cross between a Clark near-isogenic line with light tawny pubescence (genotype: *TT tdt*) and a Harosoy near-isogenic line with tawny pubescence (*TT TdTd*) revealed a single gene for pubescence color around the end of chromosome 3. Genome sequence alignment of plant introductions revealed an association between premature stop codons in Glyma.03G258700 (R2R3 MYB transcription factor) and recessive *td* allele. Cultivars and lines having near-gray or light tawny pubescence and a gray pubescence cultivar with *td* allele had premature stop codons in the gene. These results suggest that Glyma.03G258700 corresponds to the *Td* gene. It was predominantly expressed in pubescence. Compared to a tawny pubescence line, a near-isogenic line with *td* allele produced extremely small amounts of transcripts of Glyma.03G258700, *FNS II-1*, and *FNS II-2* in pubescence. The promoter of *FNS II-1* and *FNS II-2* shared *cis*-acting regulatory elements for binding of MYB proteins. These results suggest that the wild type of Glyma.03G258700 protein may bind to the promoter of *FNS II* genes and upregulate their expression, resulting in increased flavone content and deeper pubescence color. In contrast, mutated Glyma.03G258700 protein may fail to upregulate the expression of *FNS II* genes, resulting in decreased flavone content and dilute pubescence color.

Keywords: soybean, MYB transcription factor, flavone synthase II, pubescence color, *Td* gene

INTRODUCTION

Pubescence color of soybean [*Glycine max* (L.) Merr.] is controlled by two genes (*T* and *Td*) (Palmer et al., 2004). The *T* gene has a major effect on pubescence color; dominant *T* and recessive *t* allele produce tawny and gray pubescence, respectively. Further, the *T* gene controls coloration of the seed coat and hypocotyl (Palmer and Payne, 1979; Palmer et al., 2004; Murai et al., 2016). The *T* gene encodes a flavonoid 3'-hydroxylase and hydroxylates flavonoids at the 3'-position of the B-ring to generate dihydroxylated flavonoids (Buttery and Buzzell, 1973; Toda et al., 2002). The entire coding region of *F3'H* was cloned from a pair of near-isogenic lines (NILs) for the *T* gene, To7B (*TT*), and To7G (*tt*). They differed by one base deletion in To7G, resulting in a short polypeptide lacking the consensus sequence GGEK and heme-binding domain (Toda et al., 2002).

Bernard (1975) reported another gene, *Td*, affecting pubescence color. In the presence of the dominant *T* allele, dominant and recessive alleles of the *Td* locus generate tawny and light tawny (or near-gray) pubescence, respectively. Recessive *t* allele generates gray pubescence color irrespective of the allele of the *Td* locus (*tt TdTd* or *tt tdt*) (Bernard, 1975). In contrast to the *T* gene, *Td* gene affects only pubescence color (Palmer and Payne, 1979; Palmer et al., 2004).

Existence of different flavonoids is responsible for variation of tissue coloration in soybean. Iwashina et al. (2006) characterized flavonoids in soybean pubescence. A large amount of luteolin aglycone (flavone with 3,4'-dihydroxylation) was extracted together with small amounts of apigenin aglycone (flavone with 4'-hydroxylation) and luteolin derivatives from tawny pubescence. Luteolin aglycone comprised the largest portion (95–96%). From gray pubescence, a large amount of apigenin aglycone was extracted together with small amounts of luteolin aglycone, apigenin 7-*O*-glucoside, and apigenin derivatives. Apigenin aglycone was most abundant (90%). Thus, the hydroxylation pattern of the B-ring is associated with the dominance of the *T* gene. From light tawny pubescence of Clark-*td*, a Clark NIL with *td* allele, three compounds identical to those in tawny pubescence were extracted. However, compared to tawny pubescence, the amount of luteolin aglycone was reduced to half. In addition, two high performance liquid chromatography (HPLC) peaks corresponding to isoflavonoids were occasionally found in light tawny pubescence (Iwashina et al., 2006). These results suggest that the *Td* gene might be involved in the biosynthesis of flavones. Flavone glycosides exist in vacuoles in pubescence whereas flavone aglycones exist outside the cell surface, and they belong to so-called surface flavonoids (Iwashina et al., 2006). Biological roles of surface flavonoids remain to be investigated (Wollenweber, 1994). Most pigments in pubescence could not be extracted from pubescence, suggesting that flavone derivatives are highly polymerized (Iwashina et al., 2006). In soybean, flavones predominantly exist in pubescence and are quite rare in other tissues (Iwashina et al., 2006).

Plants have evolved two independent enzyme systems to synthesize flavone using the same substrates, flavone synthase

(FNS) I and II (Martens and Mithofer, 2006). Both enzymes never occur side by side in the same organism. FNS I is a soluble 2-oxoglutarate- and Fe²⁺-dependent dioxygenase, occurs in Apiaceae, whereas FNS II, a membrane-bound cytochrome P450 monooxygenase, is more widespread among plant species. Soybean has two functional copies of FNS II gene, *FNSII-1* (Glyma.12G067000) and *FNSII-2* (Glyma.12G067100) with 93% amino acid identity, with an approximately 8 kb distance on chromosome 12 (Fliegmann et al., 2010; Jiang et al., 2010).

Genome-wide association studies (GWASs) using a total of 1,402 soybean genotypes revealed a significant signal associated with pubescence color at 47,244,893 bp on chromosome 3, presumably corresponding to the *Td-td* locus (Wen et al., 2015). In addition, GWAS of a total of 12,360 accessions suggest a signal associated with the *Td-td* locus around the end of chromosome 3 (Bandillo et al., 2017). However, the identity and nature of the *Td* gene has not yet been clarified. This study was conducted to map and clone the *Td* gene responsible for flavone content and deepening/diluting of pubescence color in soybean.

MATERIALS AND METHODS

Plant Materials

Plant materials used in this study are listed in **Table 1**. Pedigree and pubescence color information was obtained from the USDA GRIN database (<https://www.ars-grin.gov/npgs/>). Soybean NIL of US cultivar Clark, L66-260 with light tawny pubescence (Clark-*td*; genotype *TT tdt*) was crossed with L66-707, NIL of Canadian cultivar Harosoy with tawny pubescence (Harosoy-*T*; *TT TdTd*) in 2004. Flowers of Clark-*td* were emasculated 1 day before opening and pollinated with pollen of Harosoy-*T*. Hybridity of F₁ plants was ascertained by the existence of tawny pubescence color. The NILs were developed by backcrossing the pubescence color traits five times into Clark or Harosoy backgrounds (Bernard et al., 1991). In addition, five US cultivars with light tawny or near-gray pubescence (Grant, Korean, Cloud, Kingwa, and Sooty) as well as a gray-pubescence US cultivar Seneca, based on pedigree information, presumably having the recessive *td* allele (Bernard, 1975) were used for genotyping. Seeds of the NILs and the cultivars were provided by the USDA Soybean Germplasm Collection.

Genetic Analysis

A total of 120 F₂ seeds together with 30 seeds each from the parents were planted in a field at the NARO Institute of Crop Science (36°06' N, 140°05' E) on 14 June 2005. N, P, and K were applied at 3.0, 4.4, and 8.3 g m⁻², respectively. Thirty seeds each from 98 F₃ families and parents were planted in the same field on 8 June 2007. Spaces between rows and plants were 70 and 10 cm, respectively. Genotype of F₂ plants was determined by pubescence color segregation of the F₃ families.

Linkage Mapping

A total of 94 F₂ plants were randomly selected and used for linkage mapping because the PCR plates and electrophoresis

apparatus are designed for multiples of 96 samples (2 parents and 94 F₂ plants). Total DNA was extracted from trifoliolate leaves of the parents and each of the F₂ plants by the CTAB method (Murray and Thompson, 1980). Pubescence color was scored in each of the F₂ plants and F₃ families. Based on pubescence color in F₂ and F₃ generations, 24 F₂ plants fixed for tawny pubescence and 24 F₂ plants fixed for light tawny pubescence were identified. Two bulked DNA samples (one for tawny pubescence and another for light tawny pubescence) were obtained by mixing 10 µl from each of the selected plants and were subjected to bulked segregant simple sequence repeat (SSR) analysis. Polymorphic SSR markers in Harosoy x Clark (Nebraska) population (Cregan et al., 1999) were used for analyzing variation between the two bulked DNA samples. PCR conditions were similar to those in a previous report (Rojas Rodas et al., 2014). Variation between the two bulked samples was used to determine which chromosome harbored the pubescence color gene. Based on the results, SSR markers on chromosome 3 (Song et al., 2004) were then tested for variation among the 94 F₂ plants. The linkage map was constructed using MAPMAKER/EXP version 3.0 (Lander et al., 1987) with the Kosambi function and the threshold logarithm of odds (LOD) score of 3.0. To fill the gap in the linkage group, SSRs were screened and an SSR marker designated as Gm03:45226162-45226406 was constructed using the Simple Sequence Repeat Identification Tool (<http://archive.gramene.org/db/markers/ssrtool>).

Next-Generation Sequencer Data Analysis

Based on information provided by Zhou et al. (2015), whole genome sequences of three USDA plant introductions with near-gray pubescence (PI 157421: SRR1533292; PI 84631: SRR1533280; PI 549046: SRR1533204) were downloaded using the DRA Search of the DNA Data Bank of Japan (<http://ddbj.nig.ac.jp/DRAsearch/query?keyword=PI+196166&show=20>) (Table 1). The genome sequences together with our next-generation sequencer (NGS) data of US cultivar Bay with gray pubescence were uploaded to the Integrative Genomics Viewer (<http://software.broadinstitute.org/software/igv/>) (Robinson et al., 2011) and were compared with those of Williams 82 having tawny pubescence.

Molecular Cloning

Total RNA was isolated from trifoliolate leaves (100 mg) of Harosoy-*T*, Clark-*td*, Cloud, and Williams 82 using the Spin Column Plant Total RNA Purification Kit (Sangon Biotech, Shanghai, China) according to the manufacturer's instructions. cDNA was synthesized by reverse transcription polymerase chain reaction of 5 µg of total RNA using the PrimeScript II 1st Strand cDNA Synthesis Kit (Takara, Dalian, China) and an oligo (dT) primer following the manufacturer's instructions. cDNA was amplified by end-to-end PCR using primers (Table 2) that were based on the genome sequence of US cultivar Williams 82 deposited in the soybean genome database (Phytozome, <https://phytozome.jgi.doe.gov/pz/portal.html#>). The total volume of 50 µl PCR mixture contained 10 pmol of each primer, 1 × PrimeSTAR Max Premix (Takara), and 200 ng cDNA. The PCR program was 35 cycles of 10 s denaturation at 98°C, 15 s annealing at 56°C and 1 min extension at 72°C. PCR was performed in an Applied Biosystems 9700 thermal cycler (Thermo Fisher Scientific, Waltham, USA). Each target band was TA cloned into the plasmid vector pMD18-T (Takara). The plasmids were sent to the Sangon Biotech for sequencing.

Genomic DNA was extracted from trifoliolate leaves of Grant, Korean, Cloud, Kingwa, Sooty, Harosoy-*T*, Clark-*td*, and Williams 82 by the CTAB method. Entire genomic fragment was amplified from Williams 82 whereas the second exon was amplified from the other cultivars using the PCR primers in Table 2. PCR mixture contained 10 pmol of each primer, 1 × PrimeSTAR Max Premix, and 50 ng genomic DNA in a total volume of 50 µl. The PCR program was 35 cycles of 10 s denaturation at 98°C, 5 s annealing at 56°C, and 10 s extension at 72°C. The PCR products were sent to Sangon Biotech for direct sequencing.

Sequencing Analysis

Nucleotide sequences of both strands were determined with the BigDye terminator cycle method using an ABI 3730XL (Thermo Fisher Scientific). Nucleotide sequences and the putative amino acid translations were analyzed with the BLAST program (Altschul et al., 1997). Nucleotide and amino acid sequences were aligned using ClustalW (<http://clustalw.ddbj.nig.ac.jp/top-j.html>) with default settings. The amino acid alignment was used

TABLE 1 | Plant materials used in this study.

Cultivar or line name	Line designation	Origin	Genotype	Pubescence color	Purpose
Harosoy- <i>T</i>	L66-707	Harosoy (6) x Clark	<i>TT TdTd</i>	Tawny	Genetic analysis and cDNA cloning
Clark- <i>td</i>	L66-260	Clark (6) x PI 91160 collected from China	<i>TT tdt</i>	Light tawny	Genetic analysis and cDNA cloning
Grant	–	Lincoln x Seneca	<i>TT tdt</i>	Light tawny	Genotyping
Korean	–	Collected from North Korea	<i>TT tdt</i>	Light tawny	Genotyping
Cloud	–	Selected from PI 16790 collected from China	<i>TT tdt</i>	Near-gray	cDNA cloning
Kingwa	–	Selected from Peking	<i>TT tdt</i>	Near-gray	Genotyping
Sooty	–	Selected from Cloud	<i>TT tdt</i>	Near-gray	Genotyping
Seneca	–	Selected from FC 03654 collected from China	<i>tt tdt</i>	Gray	Genotyping
PI 157421	–	Collected from South Korea	<i>TT tdt</i>	Near-gray	Next-generation sequencer (NGS) data analysis
PI 84631	–	Collected from South Korea	<i>TT tdt</i>	Near-gray	NGS data analysis
PI 549046	–	Collected from China	<i>TT tdt</i>	Near-gray	NGS data analysis
Bay	–	Selected from York x R62-550 (Essex x <i>G. soja</i>)	<i>tt TdT</i>	Gray	NGS data analysis
Williams 82	–	Williams (7) x Kingwa	<i>TT TdT</i>	Tawny	cDNA and DNA cloning

TABLE 2 | PCR primers used in this study.

Purpose	Target gene	Forward primer (5'-3')	Reverse primer (5'-3')
Simple sequence repeat (SSR) analysis (Gm03:45226162-45226406)	–	GGAGGATTAATTCGCTCCA	TTCGGTGAGGTCTCTAAAAGAT
cDNA cloning	Glyma.03G258700	TAAGGTGCTCGATCTGCTT	AATATGTTACAGTGCTAGCA
DNA cloning	Glyma.03G258700	TAAGGTGCTCGATCTGCTT ACTGAGATGGATGAACATC AATCAGGGTTAAGGCTTTAT	ATGGGTATTCCAGTAGTTCT TGCTAGCACGTGAACATATT AATATGTTACAGTGCTAGCA
Derived cleaved amplified polymorphic sequences (dCAPS) analysis for single-nucleotide polymorphism (SNP)	2nd exon of Glyma.03G258700	AAAGGGTCAGAGAAGACCTTA	TCCAATATTGTCACTTACCGC
dCAPS analysis for indel	Glyma.03G258700	AGACAATGGTGGGATTGGGA	AATATGTTACAGTGCTAGCA
Real-time PCR	Glyma.03G258700	GGAAAAAGTTGCCGACTGAG	TGGATCCGTTCCCTTGATTTTC
	<i>GmFNSII-1</i>	AGGTCCAGTTAACTCGGAACCAT	GCTTCACACACTAATTGCGTGT
	<i>GmFNSII-2</i>	AGGTCCAGTTAACTCGGAACCAT	GCTTCACACACTAATCGTGT
	<i>actin 1</i>	GTCCTTTCAGGAGGTACAACC	CCACATCTGCTGGAAGGTGC

to construct a phylogenetic tree of MYB genes related to flavonoid biosynthesis (**Additional File 1: Table S1**) using the neighbor-joining method with MEGA version 10.0.5 (<http://www.megasoftware.net/>) (Kumar et al., 2018). Bootstrap test of 1,000 replications was performed. Nucleotide sequences of the putative promoter region (up to 1,500 bp upstream from the start codon) of two *FNSII* genes, *GmFNSII-1* (Glyma.12G067000) and *GmFNSII-2* (Glyma.12G067100) of Williams 82 were downloaded from Phytozome. *Cis*-acting regulatory elements of these genes were investigated using New PLACE, a database of plant *cis*-acting regulatory DNA elements (<https://sogo.dna.affrc.go.jp/cgi-bin/sogo.cgi?lang=en&pj=640&action=page&page=newplace>), using default settings (Higo et al., 1999).

Derived Cleaved Amplified Polymorphic Sequences Analysis

Two pairs of PCR primers containing mismatched nucleotides were designed to detect a single-nucleotide polymorphism (SNP) existing in Clark-*td* and one base deletion in Cloud. The SNP of Clark-*td* was expected to result in the presence of the restriction site of *Bsr*BI in the amplicons. The deletion in Cloud was expected to result in the presence of the restriction site of *Eco*RV in the PCR products. The total volume of 25 μ l PCR mixture contained 10 pmol of each primer, 1 \times PrimeSTAR Max Premix, and 50 ng genomic DNA. The PCR program was 35 cycles of 10 s denaturation at 98°C, 5 s annealing at 56°C, and 10 s extension at 72°C. The PCR products were digested with *Bsr*BI or *Eco*RV, and the digests were separated on 8% nondenaturing polyacrylamide gels. After electrophoresis, the gels were stained with ethidium bromide and the DNA fragments were visualized under UV light.

Quantitative Real-Time PCR

Tissue samples were collected in three replicates. Pubescence was sampled with razor blades from pods and stems of Harosoy-*T* and Clark-*td* at the R3 stage (Fehr et al., 1971). Petals were collected on the day of opening from Harosoy-*T*. Leaves, stems, roots, root nodules, and immature seeds were sampled from Harosoy-*T* at the R3 stage. Total RNA was extracted from 50 mg

of pubescence, and 100 mg of the other tissue samples. cDNA was synthesized by reverse transcription polymerase chain reaction of 1 μ g of total RNA using the PrimeScript RT reagent Kit with gDNA Eraser (Perfect Real Time) (Takara) and an oligo (dT) primer following the manufacturer's instructions. Six concentration gradients (0, 1/2, 1/4, 1/8, 1/16, 1/32) were used to set up a standard curve. The testing samples of cDNA were diluted five times. Three biological replications and three technical repetitions were tested for quantification. The real-time PCR mixture contained 1 \times TB Green *Premix Ex Taq* (Tli RNaseH Plus) (Takara), 10 pmol of each primer, 1 \times ROX reference dye, 50 ng of cDNA, and water to a final volume of 20 μ l. The PCR was performed using the StepOnePlus Real-Time PCR instrument (Thermo Fisher Scientific). The initial 30 s denaturation at 95°C was followed by 40 cycles of 5 s denaturation at 95°C and 30 s annealing at 60°C. The expression level of the target gene was normalized using soybean *actin 1* gene (GenBank accession number: J01298). The PCR primers are listed in **Table 2**.

RESULTS

Inheritance of Pubescence Color

From a total of 120 F₂ seeds planted, 104 plants grew normally. Some of the F₂ plants were difficult to classify into tawny or light tawny pubescence class as previously reported (Bernard, 1975) so F₃ progeny tests were performed. A total of 98 F₂ plants were randomly selected, and their F₃ seeds were planted in the field. Ninety-eight F₃ families were segregated into 24 families fixed for tawny pubescence, 50 segregating families, and 24 families fixed for light tawny pubescence. The segregation fitted to a single gene control model of 1:2:1 ($\chi^2 = 0.04$, $P = 0.98$), in accordance with the previous report (Bernard, 1975).

Mapping of Pubescence Color Gene

Bulked segregant analysis suggested that the *Td* gene was located in chromosome 3. Seven SSR markers in chromosome 3 had good separation and were used for linkage mapping. The *Td* gene was mapped toward the end of chromosome 3. To fill the gap

between the *Td* gene and Satt022, a new SSR marker Gm03:45226162-45226406 was designed (**Table 2**). The primers were constructed flanking nine repeats of the AT motif. The *Td* gene was mapped 1.1 cM away from the marker toward the chromosome end, generating a linkage group spanning 86.1 cM (**Figure 1**).

Candidate Gene Identification

There are 71 genes (Glyma.03G257900 to Glyma.03G254900) from Gm03:45226162-45226406 to the chromosome end in Williams 82. Based on genome sequence comparison with Williams 82 (*Glycine max* Wm82.a2.v1), plant introductions with near-gray pubescence (PI 157421, PI 84631, and PI 549046) had mutations in the coding region of Glyma.03G258700, which is annotated as a MYB transcription factor (**Figure 2**). In PI 157421, base substitution from TGG to TGA generated a stop codon in the coding region. In PI 549046 and PI 84631, a single-base deletion occurred at different positions 3 bp apart from each other in the second exon, resulting in premature translation termination. These results suggest that Glyma.03G258700 might correspond to the *Td* gene.

However, according to the reference genome of Williams 82, cDNA of Glyma.03G258700 is truncated and lacks a start codon (**Figure 2**). Genome sequence alignment of the plant introductions and cultivar Bay indicated that no NGS reads were allocated to a 58-nucleotide region (Gm03:45302852 to Gm03:45302909) corresponding to the middle of the first exon in all genotypes (**Additional File 1: Figure S1**). We designed a forward primer in the upstream region to amplify the entire coding region of Glyma.03G258700 (**Table 2, Additional File 1: Figure S1, and Additional File 3: Figure S2**).

Gene Cloning

RT-PCR using primers for Glyma.03G258700 generated fragments of approximately 1 kb length in Harosoy-*T*, Clark-*td*, Cloud, and Williams 82. The coding region of Harosoy-*T* and Williams 82 was 840 bp long and encoded 279 amino acids (**Figure 3A**). BLAST analysis confirmed that Glyma.03G258700 belongs to a MYB transcription factor of the R2R3 type (**Figure 3A**). A single base was substituted from G to A at the nucleotide position 633 in Clark-*td* compared with Harosoy-*T*. The SNP generated a premature stop codon (TGA) resulting in a short polypeptide consisting of 210 amino acids (**Figure 3B**). Cloud had one base deletion at the nucleotide position 677. This deletion changed a subsequent reading frame and generated a truncated polypeptide consisting of 232 amino acids (**Figure 3B**). Sequencing of the second exon revealed that Korean, Seneca, Kingwa, Grant, and Sooty had the same SNP as Clark-*td*.

Glyma.03G258700 of Harosoy-*T* had an amino acid identity of 58% with a persimmon MYB gene, *DkMYB4*, which activates proanthocyanidin biosynthesis (Akagi et al., 2009) (**Figure 4**). Particularly, the N-terminal region, including R2R3 MYB-binding domains as well as the C-terminal region, were quite similar. Amino acid identity of the MYB-binding domain was 89%. Plant MYB proteins are classified into 28 subclasses (S1 to S28), and many of them are involved in flavonoid biosynthesis

chromosome 3

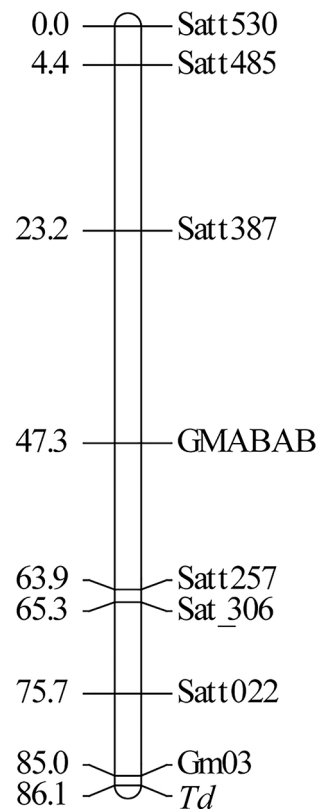


FIGURE 1 | Linkage mapping of soybean pubescence color gene *Td* using an F_2 population derived from a cross of near-isogenic lines, Clark-*td* x Harosoy-*T*. Distances of markers (cM) from the top of linkage group are shown on the left. Gm03 is an abbreviation of a newly developed simple sequence repeat (SSR) marker, Gm03:45226162-45226406.

(Liu et al., 2015). In soybean, two flavonoid-related genes, flower color gene *W2* (Glyma.14G154400) and seed coat color gene *R* (Glyma.09G235100), are known to encode R2R3 MYB transcription factors (Gillman et al., 2011; Takahashi et al., 2011; Takahashi et al., 2013; Yan et al., 2015). Based on phylogenetic analysis, Glyma.03G258700 belongs to S5 together with *DkMYB4*, whereas *R* and *W2* genes belong to S6 and S27, respectively (**Figure 5 and Additional File 2: Table S1**).

dCAPS Analysis

The scheme of derived cleaved amplified polymorphic sequences (dCAPS) analysis for SNP and indel are shown in **Figures 6A, B**, respectively. Results of dCAPS analyses are shown in **Figure 6C**. PCR using dCAPS primers for SNP generated amplified products of approximately 210 bp in all genotypes (Korean, Cloud, Seneca, Kingwa, Grant, Sooty, Harosoy-*T*, and Clark-*td*). Digestion with *Bsr*BI generated a fragment of approximately 185 bp in Korean, Seneca, Kingwa, Grant, Sooty, and Clark-*td*, whereas PCR

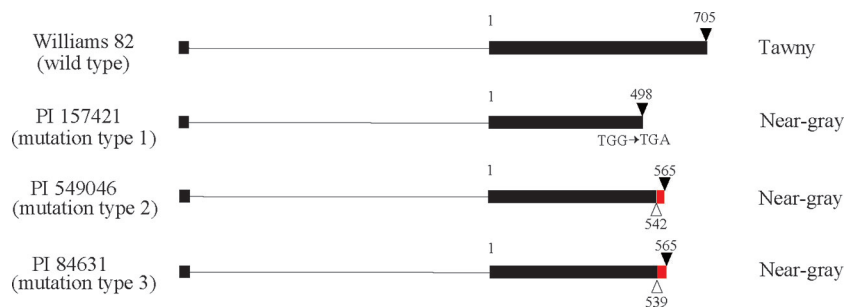


FIGURE 2 | Gene structure of Glyma.03G258700 in Williams 82 with tawny pubescence and three plant introductions with near-gray pubescence. Exons and introns are shown by boxes and lines, respectively. White and black triangles indicate deletions and stop codons, respectively. Red box represents a region where frameshift mutation occurred due to deletion. Numbers indicate nucleotide position starting from second exon.

A

Harosoy- <i>T</i>	MGRAPCCSKVGLHRGPWTPREDALLTKYIQAHGKGQWRS L P K T A G	45
	<u>LLRCGKSCRLRWMNYLRPDIK RGNITPEEDDLIIRMHSL L GNRWS</u>	90
	<u>LIAGRLPGRTDNEIKNYWNTHLSK KLRNQGTDPNTHDKLTEAPEK</u>	135
	KKGRKKNKQKSTENNKGSEKTLVYLPKPIRVKALSSCIPRTDSTL	180
	TLDSNSATTASTSQEKVQSPEAETKEVNMVWGVSDNIGDNGGIGI	225
	↓ FFGEDHDLVNSASYMECYSDDHDTLEKLYEYLLQLLNVEEKPDEL	270
	DSFAQSLLV*	279

B

Harosoy- <i>T</i>	ACAAAAGAGGTGAACATGGTTTGGGGGGTAAGTGACAATATT	651
	T K E V N M V W G V S D N G	
Clark- <i>td</i>	ACAAAAGAGGTGAACATGGTTTGA GGGGTAAGTGACAATATT	651
	T K E V N M V *	
Harosoy- <i>T</i>	AATGGTGGGATTGGAATATTCTTTGGTGAAGACCATGACCTA	699
	N G G I G I F F G E D H D L	
Cloud	AATGGTGGGATTGGAATAT-CTTTGGTGAAGACCATGACCTAG	699
	N G G I G I S L V K T M T *	

FIGURE 3 | Nucleotide and amino acid polymorphisms of the soybean gene Glyma.03G258700 in Harosoy-*T*, Clark-*td*, and Cloud. Asterisk represents stop codon. **(A)** Amino acid sequence of Harosoy-*T*. The position of single-nucleotide polymorphism (SNP) and indel found in Clark-*td* and Cloud are indicated by red and black arrows, respectively. MYB DNA-binding domains are underlined. **(B)** Alignment of partial cDNA and amino acid sequences. The SNP in Clark-*td* and an indel in Cloud are shown in red font. The altered amino acids caused by frameshift mutation in Cloud are shown in bold font.

products of Cloud and Harosoy-*T* were undigested. PCR using dCAPS primers for indel generated fragments of approximately 270 bp in all genotypes. Digestion with *EcoRV* generated a fragment of approximately 250 bp in Cloud, whereas PCR products of the other genotypes were unaffected. Thus, Korean, Seneca, Kingwa, Grant, Sooty, and Clark-*td* have the same SNP

as PI 157421, whereas Cloud has the same deletion as PI 549046 (**Additional File 3: Figure S2**).

Gene Expression

The transcript level of Glyma.03G258700 in pubescence was extremely high (151.6 times of the immature seed), whereas that

Glyma.03G258700	<u>MGRAPCCSKVGLHRGPWT</u> <u>PR</u> <u>EDALLTKYIQAHGK</u> <u>GW</u> <u>RS</u> <u>LP</u> <u>KTAGLLRCGK</u> <u>S</u> <u>C</u> <u>S</u> <u>R</u> <u>L</u> <u>R</u> <u>W</u> <u>M</u> <u>N</u> <u>Y</u>	60
<i>DkMYB4</i>	<u>MGRAPCCSKVGLHRGPWT</u> <u>CR</u> <u>EDG</u> <u>LLTKYIQVHG</u> <u>CS</u> <u>WR</u> <u>SL</u> <u>PK</u> <u>AGLLRCGK</u> <u>S</u> <u>C</u> <u>S</u> <u>R</u> <u>L</u> <u>R</u> <u>W</u> <u>M</u> <u>N</u> <u>Y</u>	60
Glyma.03G258700	<u>LRPDIK</u> <u>RGNITP</u> <u>ED</u> <u>DL</u> <u>IR</u> <u>MH</u> <u>SL</u> <u>LG</u> <u>NR</u> <u>W</u> <u>SL</u> <u>I</u> <u>AG</u> <u>RL</u> <u>P</u> <u>GR</u> <u>T</u> <u>D</u> <u>NE</u> <u>I</u> <u>K</u> <u>NY</u> <u>W</u> <u>N</u> <u>T</u> <u>H</u> <u>L</u> <u>S</u> <u>K</u> <u>L</u> <u>R</u> <u>N</u> <u>O</u> <u>G</u>	120
<i>DkMYB4</i>	<u>LRPDIK</u> <u>RGNITP</u> <u>ED</u> <u>DL</u> <u>IR</u> <u>MH</u> <u>AL</u> <u>LG</u> <u>NR</u> <u>W</u> <u>SL</u> <u>I</u> <u>AG</u> <u>RL</u> <u>P</u> <u>GR</u> <u>T</u> <u>D</u> <u>NE</u> <u>I</u> <u>K</u> <u>NY</u> <u>W</u> <u>N</u> <u>T</u> <u>H</u> <u>L</u> <u>S</u> <u>K</u> <u>L</u> <u>R</u> <u>S</u> <u>O</u> <u>G</u>	120
Glyma.03G258700	<u>TDPNTH</u> <u>DKL</u> <u>TEA</u> <u>PE</u> <u>---</u> <u>K</u> <u>KGR</u> <u>K</u> <u>K</u> <u>N</u> <u>K</u> <u>O</u> <u>K</u> <u>S</u> <u>T</u> <u>E</u> <u>N</u> <u>N</u> <u>K</u> <u>G</u> <u>S</u> <u>E</u> <u>K</u> <u>T</u> <u>L</u> <u>V</u> <u>Y</u> <u>L</u> <u>P</u> <u>K</u> <u>P</u> <u>R</u> <u>V</u> <u>K</u> <u>A</u> <u>L</u> <u>S</u> <u>S</u> <u>C</u> <u>I</u> <u>P</u> <u>R</u> <u>T</u>	176
<i>DkMYB4</i>	<u>TDPNTH</u> <u>KKL</u> <u>SDS</u> <u>HV</u> <u>Q</u> <u>E</u> <u>P</u> <u>K</u> <u>K</u> <u>S</u> <u>S</u> <u>N</u> <u>K</u> <u>K</u> <u>O</u> <u>K</u> <u>N</u> <u>K</u> <u>S</u> <u>N</u> <u>L</u> <u>D</u> <u>H</u> <u>E</u> <u>R</u> <u>L</u> <u>K</u> <u>V</u> <u>H</u> <u>N</u> <u>P</u> <u>K</u> <u>P</u> <u>R</u> <u>K</u> <u>S</u> <u>L</u> <u>A</u> <u>S</u> <u>F</u> <u>S</u> <u>F</u> <u>S</u> <u>R</u>	180
Glyma.03G258700	<u>DS</u> <u>L</u> <u>T</u> <u>L</u> <u>D</u> <u>N</u> <u>S</u> <u>A</u> <u>T</u> <u>A</u> <u>T</u> <u>S</u> <u>T</u> <u>S</u> <u>Q</u> <u>E</u> <u>K</u> <u>V</u> <u>Q</u> <u>S</u> <u>P</u> <u>E</u> <u>A</u> <u>E</u> <u>T</u> <u>E</u> <u>V</u> <u>N</u> <u>M</u> <u>V</u> <u>G</u> <u>V</u> <u>S</u> <u>N</u> <u>I</u> <u>G</u> <u>N</u> <u>G</u> <u>G</u> <u>I</u> <u>G</u> <u>I</u> <u>F</u> <u>F</u> <u>G</u> <u>E</u> <u>D</u> <u>-</u> <u>H</u> <u>D</u> <u>L</u> <u>V</u> <u>N</u>	235
<i>DkMYB4</i>	<u>DS</u> <u>S</u> <u>S</u> <u>F</u> <u>D</u> <u>W</u> <u>I</u> <u>T</u> <u>T</u> <u>A</u> <u>T</u> <u>A</u> <u>M</u> <u>P</u> <u>S</u> <u>-----</u> <u>G</u> <u>S</u> <u>N</u> <u>H</u> <u>E</u> <u>G</u> <u>E</u> <u>R</u> <u>G</u> <u>M</u> <u>L</u> <u>G</u> <u>N</u> <u>G</u> <u>S</u> <u>V</u> <u>G</u> <u>H</u> <u>E</u> <u>V</u> <u>G</u> <u>E</u> <u>F</u> <u>G</u> <u>E</u> <u>D</u> <u>G</u> <u>S</u> <u>D</u> <u>H</u> <u>Y</u> <u>H</u>	232
Glyma.03G258700	<u>S</u> <u>A</u> <u>S</u> <u>Y</u> <u>M</u> <u>E</u> <u>C</u> <u>Y</u> <u>S</u> <u>D</u> <u>H</u> <u>D</u> <u>L</u> <u>E</u> <u>K</u> <u>L</u> <u>Y</u> <u>E</u> <u>E</u> <u>Y</u> <u>L</u> <u>Q</u> <u>L</u> <u>L</u> <u>N</u> <u>V</u> <u>E</u> <u>K</u> <u>P</u> <u>-</u> <u>E</u> <u>L</u> <u>D</u> <u>S</u> <u>F</u> <u>A</u> <u>O</u> <u>S</u> <u>L</u> <u>L</u> <u>V</u>	279
<i>DkMYB4</i>	<u>H</u> <u>D</u> <u>H</u> <u>M</u> <u>E</u> <u>D</u> <u>S</u> <u>D</u> <u>L</u> <u>E</u> <u>C</u> <u>O</u> <u>S</u> <u>L</u> <u>E</u> <u>K</u> <u>L</u> <u>Y</u> <u>E</u> <u>E</u> <u>Y</u> <u>L</u> <u>Q</u> <u>L</u> <u>L</u> <u>A</u> <u>T</u> <u>E</u> <u>D</u> <u>D</u> <u>G</u> <u>O</u> <u>L</u> <u>D</u> <u>S</u> <u>F</u> <u>A</u> <u>S</u> <u>L</u> <u>L</u> <u>I</u>	277

FIGURE 4 | Amino acid alignment of soybean MYB gene, Glyma.03G258700 with persimmon *MYB4* gene, *DkMYB4*. Dashes represent gaps introduced to improve the alignment. Identical and similar amino acids are shown in white font highlighted in black and gray, respectively. MYB DNA-binding domains are underlined.

in the other tissues was comparable with immature seed (leaf: 0.57 time; stem: 0.46 time; root: 2.26 times; root nodule: 5.02 times; flower: 1.38 times) in Harosoy-*T* (Figure 7A). In pubescence, the transcript level of Glyma.03G258700 was approximately 12% in Clark-*td* compared with that in Harosoy-*T* (Figure 7B). Similarly, the transcript levels of two flavone synthase genes, *GmFNSII-1* and *GmFNSII-2*, in Clark-*td* were approximately 23% and 10%, respectively, of that in Harosoy-*T*.

Promoter of FNSII Genes

The nucleotide sequence of the promoter region had generally a low identity (18%) among the two flavone synthase II genes, *GmFNSII-1* and *GmFNSII-2*, in Williams 82. However, nucleotides around the end of the region were relatively similar (Figure 8). There are six kinds of *cis*-acting regulatory elements for binding of MYB transcription factors (MYBCORE, MYB2CONSENSUSAT, MYBCOREATCYCB1, MYBPZM, MYB1AT, and MYBPLANT) slightly upstream of the coding region, majority of which are shared by the two genes (Figure 8).

DISCUSSION

Pubescence color of soybean is controlled by two genes *T* and *Td* (Bernard, 1975; Palmer et al., 2004). The *T* gene encodes a flavonoid 3'-hydroxylase and hydroxylates the 3'-position of the B-ring to generate dihydroxylated flavonoids (Buttery and Buzzell, 1973; Toda et al., 2002). The dominant *T* allele generates luteolin derivatives and tawny color in pubescence, whereas the recessive *t* allele produces apigenin derivatives and gray color (Iwashina et al., 2006). Thus, *T* gene alters pubescence color by changing the structure of flavones. Under the dominant *T* allele, dominant *Td* allele generates higher amount of luteolin derivatives and tawny color in pubescence (Iwashina et al., 2006). In contrast, recessive *td* allele produces less amount of luteolin derivatives and light tawny (or near-gray) pubescence color. These results suggest that the *Td* gene may control amounts of flavones.

Genetic analysis and linkage mapping using the F₂ population and F₃ families derived from a cross between NILs with tawny

and light tawny pubescence suggested that a gene responsible for pubescence color was located around the end of chromosome 3, which is consistent with the results of GWAS analysis (Wen et al., 2015; Zhou et al., 2015). Flavones are dependent for their biosynthesis on two functional copies of *FNS II* genes, *FNSII-1* and *FNSII-2*, on chromosome 12 (Fliegmann et al., 2010; Jiang et al., 2010). Accordingly, the *FNS II* genes may not correspond to the *Td* gene. There remains a possibility that the *Td* gene encodes a transcription factor that controls the expression of *FNS II* genes. Genome sequence alignment of plant introductions revealed that near-gray pubescence is associated with premature stop codons in the coding region of Glyma.03G258700 encoding an R2R3 MYB transcription factor.

However, cDNA of Glyma.03G258700 is truncated and lacks a start codon in the reference genome of Williams 82. Genome sequence alignment indicated that no NGS reads were allocated to the 58-nucleotide region in the middle of the first exon. We presume that the reference sequence of Williams 82 has a mistake at this region. We designed a forward primer upstream of the region and successfully cloned the entire cDNA of Harosoy-*T*, Clark-*td*, Cloud, and Williams 82 as well as the entire genomic fragment of Williams 82. The 58-nucleotide sequence of these cDNAs including Williams 82 was completely different from the reference sequence of Williams 82 (Additional File 1: Figure S1 and Additional File 3: Figure S2). Reference sequence of Williams 82 needs to be corrected.

Glyma.03G258700 of Harosoy-*T* consisted of 840 nucleotides and encoded 279 amino acids. In contrast, cDNA of Clark-*td* had a nonsense mutation at the nucleotide position 633 and generated a truncated polypeptide comprising 210 amino acids. Further, Cloud had one base deletion at the nucleotide position 677, resulting in a truncated polypeptide comprising 232 amino acids. dCAPS analysis revealed that Grant, Korean, Kingwa, Sooty, and Seneca with *td* allele had the same SNP with Clark-*td*. Thus, dCAPS analysis confirmed that Seneca with gray pubescence has the *td* allele as expected by pedigree information (Bernard, 1975). R2R3 MYB transcription factors have a conserved DNA-binding domain at the N terminus and a more variable C-terminal transcriptional activation domain (Pireyre and Burow, 2015). These results suggest that

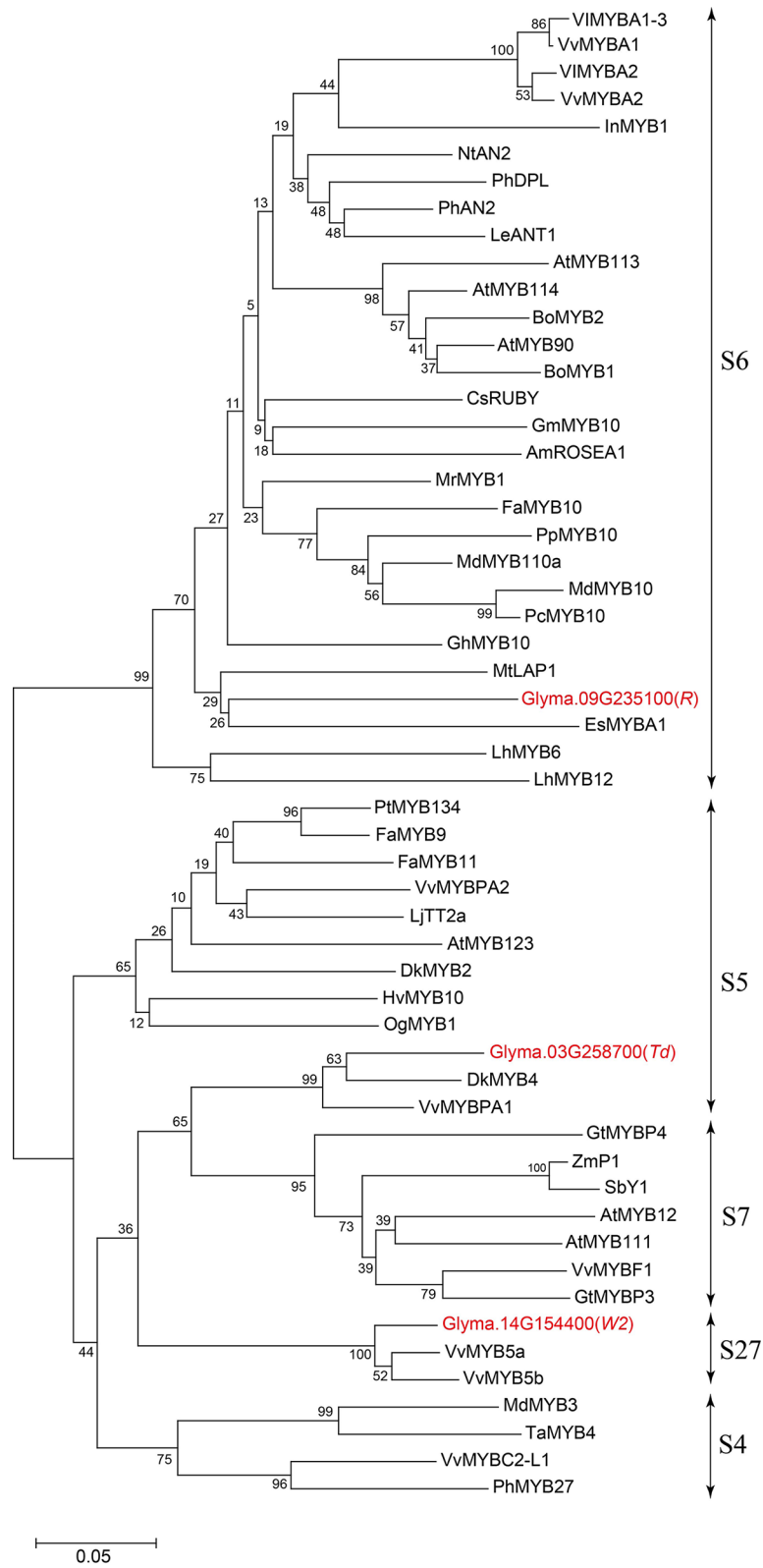


FIGURE 5 | Phylogenetic tree of plant R2R3 MYB transcription factors related to flavonoid biosynthesis. Soybean MYBs are shown in red font. Bar represents 0.05 amino acid substitution/site.

A**Partial cDNA sequence of Glyma.03G258700 gene**

PCR primers: 5' - AAAGGGTCAGAGAAGACCTTA GGGGTAAGTGACAATATTGGA - 5'
 Harosoy-*T*: 5' - AACAAAGGGTCAGAGAAGACCTTAGTTT----TGGGGGTAAGTGACAATATTGGAGA
 Clark-*td*: 5' - AACAAAGGGTCAGAGAAGACCTTAGTTT----TGAGGGTAAGTGACAATATTGGAGA

↓ PCR reaction

PCR product

Harosoy-*T*: 5' - AAAGGGTCAGAGAAGACCTTAGTTT----TGGGGTAAGTGACAATATTGGA
 Clark-*td*: 5' - AAAGGGTCAGAGAAGACCTTAGTTT----TGAGGTAAGTGACAATATTGGA
BsrBI site

B**Partial cDNA sequence of Glyma.03G258700 gene**

PCR primers: 5' - AGACAATGGTGGGATTGGGGA TGCTAGCACGTGAACATATT - 5'
 Harosoy-*T*: 5' - TGGAGACAATGGTGGGATTGGAATATTCTT----CCTGCTAGCACGTGAACATATTGTA
 Cloud: 5' - TGGAGACAATGGTGGGATTGGAATAT-CTT----CCTGCTAGCACGTGAACATATTGTA

↓ PCR reaction

PCR product

Harosoy-*T*: 5' - AGACAATGGTGGGATTGGGATATTCTT----CCTGCTAGCACGTGAACATATT
 Cloud: 5' - AGACAATGGTGGGATTGGGATAT-CTT----CCTGCTAGCACGTGAACATATT
EcoRV site

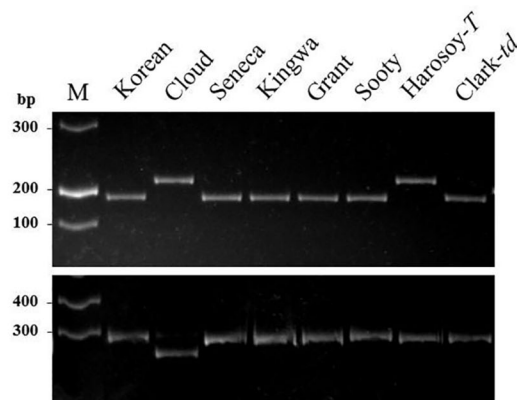
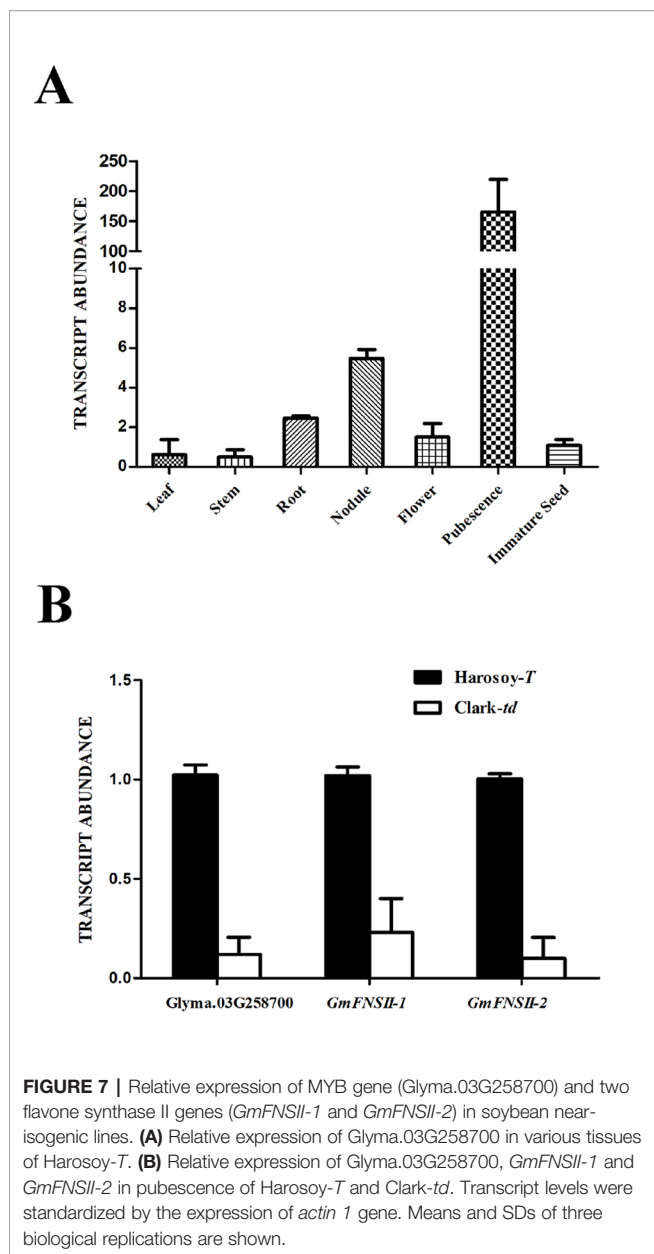
C

FIGURE 6 | Schematic presentation and results of derived cleaved amplified polymorphic sequences (dCAPS) analysis. **(A)** dCAPS analysis to detect an SNP. A mismatched nucleotide (C, in white font highlighted in black) was incorporated in the reverse primer to generate a *BsrBI* site in the PCR product of Clark-*td*. **(B)** dCAPS analysis to detect an indel. A mismatched nucleotide (G, in white font highlighted in black) was incorporated in the forward primer to generate an *EcoRV* site in the PCR product of Cloud. Annealing sites of PCR primers are single-underlined. Polymorphic nucleotides are shown in red font. **(C)** Results of dCAPS analysis in pubescence color variants. (Upper panel) Results of dCAPS analysis for SNP. PCR products were digested by *BsrBI*. (Lower panel) Results of dCAPS analysis for indel. PCR products were digested by *EcoRV*. M: DNA ladder molecular weight marker.



Glyma.03G258700 corresponds to the *Td* gene and that premature stop codons disable the transactivating ability of the MYB protein.

There are three types of mutations (type 1 to 3) in the gene (Figure 2 and Additional File 3: Figure S2). Under genetic background of dominant *TT* allele, cultivars of type 2 and 3 have near-gray pubescence whereas type 1 cultivars have either near-gray or light tawny pubescence. Thus, these types were not directly associated with pubescence color, suggesting that other genetic backgrounds may affect the lightness of pubescence color. The expression level of the gene in pubescence of Clark-*td* was very low compared with Harosoy-*T*, probably because of nonsense-mediated mRNA decay, an mRNA surveillance mechanism that eliminates aberrant mRNAs (Chang et al., 2007). Similar

observations were made with pubescence color gene *T* (flavonoid 3'-hydroxylase) and flower color gene *Wm* (flavonol synthase) in soybean (Toda et al., 2002; Takahashi et al., 2007).

A large number of R2R3 MYB transcription factors are involved in flavonoid biosynthesis. An additional data file shows that majority of the MYB proteins are involved in biosynthesis of anthocyanins, proanthocyanidins, or flavonols (Additional File 2: Table S1) (Liu et al., 2015). The *Td* gene specifically controls color of pubescence, but not changes the color of flower, seed, or hypocotyl (Bernard, 1975; Palmer and Payne, 1979). The allele at the *Td* locus did not affect the amounts of anthocyanins, flavonol glycosides, or dihydroflavonol in petals (Iwashina et al., 2007). In addition, the present study revealed that Glyma.03G258700 expresses predominantly in pubescence where flavones are almost exclusively deposited. These results suggest a hypothesis that Glyma.03G258700 has evolved to upregulate biosynthesis of flavones existing in pubescence. Further investigation including expression assays of other flavonoid biosynthesis genes may be necessary to ascertain this hypothesis.

Expression levels of Glyma.03G258700, *GmFNSII-1*, and *GmFNSII-2* in pubescence of Clark-*td* were extremely low compared with those in Harosoy-*T*. In addition, the promoter region of *FNSII-1* and *FNSII-2* shared *cis*-regulatory elements for MYB transcription factors, including the MYBCORE element. Persimmon DkMYB4 protein having a quite similar MYB-binding domain was proven to bind to the MYBCORE element by electrophoretic mobility shift assays (Akagi et al., 2009). These results strongly suggest that the Glyma.03G258700 protein may also bind to a similar element in the promoter of *FNSII* genes and upregulate their transcription.

In summary, the wild-type Glyma.03G258700 protein may bind to the promoter of *FNS II* genes and upregulate their transcription, resulting in high flavone amounts and tawny pubescence color. In contrast, premature stop codons in Glyma.03G258700 may cause a dysfunction in the transactivating ability of the MYB protein. The mutated proteins fail to promote the expression of *FNSII* genes, resulting in lower flavone amounts and dilute pubescence color. *FNS II* converts flavanones to flavones. Clark-*td* occasionally produces isoflavonoids in pubescence (Iwashina et al., 2006). Mutated MYB protein may have failed to promote catalysis of flavanones to flavones. The remaining flavanones may have been transformed by isoflavone synthase to generate isoflavonoids. Electrophoretic mobility shift assays may be necessary to ascertain the interaction between the Glyma.03G258700 protein and the *cis*-acting regulatory elements in the promoter of *FNSII* genes. Gene editing and transgenic experiments to prove our hypothesis that Glyma.03G258700 corresponds to the *Td* gene, upregulates flavone biosynthesis, and deepens pubescence color, remain as future work. Investigations of its promoter may help identify *cis*-acting regulatory

intron of *Harosoy-T* are drawn by black box harboring amino acids and black line, respectively. Gene orientation is indicated by an arrow underneath. Only amino acids in yellow font are indicated as 1st exon in Wm82.a2.v1.

ADDITIONAL FILE 2: TABLE S1 | List of R2R3 MYB transcription factors used in phylogenetic analysis.

REFERENCES

- Akagi, T., Ikegami, A., Tsujimoto, T., Kobayashi, S., Sato, A., et al. (2009). DkMyb4 is a Myb transcription factor involved in proanthocyanidin biosynthesis in persimmon fruit. *Plant Physiol.* 151, 2028–2045. doi: 10.1104/pp.109.146985
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Bandillo, N. B., Lorenz, A. J., Graef, G. L., Arquin, D., Hyten, D. L., et al. (2017). Genome-wide association mapping of qualitatively inherited traits in a germplasm collection. *Plant Genome* 10. doi: 10.3835/plantgenome2016.06.0054
- Bernard, R. L., Nelson, R. L., and Creameans, C. R. (1991). USDA soybean genetic collections: isoline collection. *Soyb. Genet. Newsl.* 18, 27–57.
- Bernard, R. L. (1975). The inheritance of near-gray pubescence color. *Soyb. Genet. Newsl.* 2, 31–35.
- Buttery, B. R., and Buzzell, R. I. (1973). Varietal differences in leaf flavonoids of soybeans. *Crop Sci.* 13, 103–106. doi: 10.2135/cropsci1973.0011183X001300010033x
- Chang, Y. F., Imam, J. S., and Wilkinson, M. E. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* 76, 51–74. doi: 10.1146/annurev.biochem.76.050106.093909
- Cregan, P. B., Jarvik, T., Bush, A. L., Shoemaker, R. C., Lark, K. G., et al. (1999). An integrated genetic linkage map of the soybean genome. *Crop Sci.* 39, 1464–1490. doi: 10.2135/cropsci1999.3951464x
- Fehr, W. R., Caviness, C. E., Burmood, D. T., and Pennington, J. S. (1971). Stage of development descriptions for soybeans, *glycine-max* (L) merrill. *Crop Sci.* 11, 929–931. doi: 10.2135/cropsci1971.0011183X001100060051x
- Fliegmann, J., Furtwangler, K., Malterer, G., Cantarello, C., Schuler, G., et al. (2010). Flavone synthase II (CYP93B16) from soybean (*Glycine max* L.). *Phytochemistry* 71, 508–514. doi: 10.1016/j.phytochem.2010.01.007
- Gillman, J. D., Tetlow, A., Lee, J. D., Shannon, J. G., and Bilyeu, K. (2011). Loss-of-function mutations affecting a specific *Glycine max* R2R3 MYB transcription factor result in brown hilum and brown seed coats. *BMC Plant Biol.* 11, 155. doi: 10.1186/1471-2229-11-155
- Higo, K., Ugawa, Y., Iwamoto, M., and Korenaga, T. (1999). Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* 27, 297–300. doi: 10.1093/nar/27.1.297
- Iwashina, T., Benitez, E. R., and Takahashi, R. (2006). Analysis of flavonoids in pubescence of soybean near-isogenic lines for pubescence color loci. *J. Hered.* 97, 438–443. doi: 10.1093/jhered/esl027
- Iwashina, T., Githiri, S. M., Benitez, E. R., Takemura, T., Kitajima, J., et al. (2007). Analysis of flavonoids in flower petals of soybean near-isogenic lines for flower and pubescence color genes. *J. Hered.* 98, 250–257. doi: 10.1093/jhered/esm012
- Jiang, Y. N., Wang, B. A., Li, H., Yao, L. M., and Wu, T. L. (2010). Flavonoid production is effectively regulated by RNAi interference of two flavone synthase genes from *Glycine max*. *J. Plant Biol.* 53, 425–432. doi: 10.1007/s12374-010-9132-9
- Kumar, S., Stecher, G., Li, M., Nknyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., et al. (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1, 174–181. doi: 10.1016/0888-7543(87)90010-3
- Liu, J. Y., Osbourn, A., and Ma, P. D. (2015). MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol. Plant* 8, 689–708. doi: 10.1016/j.molp.2015.03.012
- Martens, S., and Mithofer, A. (2006). Flavones and flavone synthases. *Phytochemistry* 67, 521–521. doi: 10.1016/j.phytochem.2006.01.004
- ADDITIONAL FILE 3: FIGURE S2** | Gene structure of Glyma.03G258700 deduced from sequencing or dCAPS analysis in pubescence color variants. Exons and introns are shown by boxes and lines, respectively. White and black triangles indicate start codons and stop codons, respectively. Red box represents a region where frameshift mutation occurred due to deletion. Numbers indicate nucleotide position starting from second exon.
- Murai, Y., Yan, F., Iwashina, T., and Takahashi, R. (2016). Analysis of anthocyanin pigments in soybean hypocotyl. *Can. J. Plant Sci.* 96, 935–938. doi: 10.1139/cjps-2016-0055
- Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4325. doi: 10.1093/nar/8.19.4321
- Palmer, R. G., and Payne, R. C. (1979). Genetic control of hypocotyl pigmentation among white-flowered soybeans grown in continuous light. *Crop Sci.* 19, 124–126. doi: 10.2135/cropsci1979.0011183X001900010033x
- Palmer, R. G., Pfeiffer, T. W., Buss, G. R., and Kilen, T. C. (2004). “Qualitative genetics,” in *Soybeans: Improvement, Production, and Uses*, 3rd edition. Eds. H. R. Boerma and J. E. Specht. (Madison, WI: American Society of Agronomy).
- Pireyre, M., and Burow, M. (2015). Regulation of MYB and bHLH transcription factors: a glance at the protein level. *Mol. Plant* 8, 378–388. doi: 10.1016/j.molp.2014.11.022
- Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi: 10.1038/nbt.1754
- Rojas Rodas, F., Rodriguez, T. O., Murai, Y., Iwashina, T., Sugawara, S., et al. (2014). Linkage mapping, molecular cloning and functional analysis of soybean gene *Fg2* encoding flavonol 3-O-glucoside (1→6) rhamnosyltransferase. *Plant Mol. Biol.* 84, 287–300. doi: 10.1007/s11103-013-0133-1
- Song, Q. J., Marek, L. F., Shoemaker, R. C., Lark, K. G., Concibido, V. C., et al. (2004). A new integrated genetic linkage map of the soybean. *Theor. Appl. Genet.* 109, 122–128. doi: 10.1007/s00122-004-1602-3
- Takahashi, R., Githiri, S. M., Hatayama, K., Dubouzet, E. G., Shimada, N., et al. (2007). A single-base deletion in soybean flavonol synthase gene is associated with magenta flower color. *Plant Mol. Biol.* 63, 125–135. doi: 10.1007/s11103-006-9077-z
- Takahashi, R., Benitez, E. R., Oyoo, M. E., Khan, N. A., and Komatsu, S. (2011). Nonsense mutation of an MYB transcription factor is associated with purple-blue flower color in soybean. *J. Hered.* 102, 458–463. doi: 10.1093/jhered/esr028
- Takahashi, R., Yamagishi, N., and Yoshikawa, N. (2013). A MYB transcription factor controls flower color in soybean. *J. Hered.* 104, 149–153. doi: 10.1093/jhered/ess081
- Toda, K., Yang, D. J., Yamanaka, N., Watanabe, S., Harada, K., et al. (2002). A single-base deletion in soybean flavonoid 3'-hydroxylase gene is associated with gray pubescence color. *Plant Mol. Biol.* 50, 187–196. doi: 10.1023/A:1016087221334
- Wen, Z. X., Boyse, J. F., Song, Q. J., Cregan, P. B., and Wang, D. C. (2015). Genomic consequences of selection and genome-wide association mapping in soybean. *BMC Genomics* 16, 671. doi: 10.1186/s12864-015-1872-y
- Wollenweber, E. (1994). “Flavones and flavonols,” in *The flavonoids, advances in research since 1986*. Ed. J. B. Harborne (London, UK: Chapman & Hall), 259–335.
- Yan, F., Di, S., and Takahashi, R. (2015). CACTA-superfamily transposable element is inserted in MYB transcription factor gene of soybean line producing variegated seeds. *Genome* 58, 365–374. doi: 10.1139/gen-2015-0054
- Zhou, Z. K., Jiang, Y., Wang, Z., Gou, Z. H., Lyu, J., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33, 408–414. doi: 10.1038/nbt.3096

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yan, Githiri, Liu, Sang, Wang and Takahashi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CCBY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.