

Loss-of-function mutation of soybean R2R3 MYB transcription factor reduces flavone content and dilutes tawny pubescence color

Fan Yan (✉ fenfeiyongyuan@163.com)

Jilin University

Stephen M. Githiri

Jomo Kenyatta University of Agriculture and Technology

Yu Sang

Jilin University

Qingyu Wang

Jilin University

Ryoji Takahashi

National Institute of Crop Science

Research article

Keywords: Soybean, MYB transcription factor, Flavone synthase II, Pubescence color, Td gene

Posted Date: July 3rd, 2019

DOI: <https://doi.org/10.21203/rs.2.10857/v1>

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

[Read Full License](#)

Abstract

Background 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 catalyzed by 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. Results 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 TdT*) 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 binds to the promoter of *FNS II* genes and upregulates their expression, resulting in increased flavone content and deeper pubescence color. In contrast, mutated Glyma.03G258700 protein fails to upregulate the expression of *FNS II* genes, resulting in decreased flavone content and dilute pubescence color. Conclusions This study revealed that soybean Glyma.03G258700 encoding the R2R3 MYB transcription factor corresponds to the *Td* gene. The wild type of MYB protein binds to the promoter of *FNS II* genes and upregulates their expression, resulting in higher flavone content and deeper pubescence color. Loss-of-function mutation of the gene fails to promote expression of *FNS II* genes, resulting in lower flavone content and dilute pubescence color.

Background

Pubescence color of soybean (*Glycine max* (L.) Merr.) is controlled by two genes (*T* and *Td*) [1]. 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 [1-3]. The *T* gene encodes a flavonoid 3-hydroxylase and hydroxylates flavonoids at the 3-position of the B-ring to generate dihydroxylated flavonoids [4, 5]. 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 [5].

Bernard [6] 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 TdT* or *tt tdt*) [6]. In contrast to the *T* gene, *Td* gene affects only pubescence color [1, 2].

Existence of different flavonoids is responsible for variation of tissue coloration in soybean. Iwashina et al. [7] 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, 3 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 [7]. 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 [7]. Biological roles of surface flavonoids remain to be investigated [8]. Most pigments in pubescence could not be extracted from pubescence, suggesting that flavone derivatives are highly polymerized [7]. In soybean, flavones predominantly exist in pubescence and are quite rare in other tissues [7].

Plants have evolved two independent enzyme systems to catalyze flavone synthesis using the same substrates, flavone synthase (FNS) I and II [9]. Both enzymes never occur side by side in the same organism. FNS I, 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 genes, *FNSII-1* (Glyma.12G067000) and *FNSII-2* (Glyma.12G067100) with 93% amino acid identity existing with an approximately 8 kb distance in chromosome 12 [10, 11].

Genome-wide association studies (GWAS) 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 [12]. 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 [13]. 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.

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 [6] so F₃ progeny tests were performed. A total of 98 F₂ plants were randomly selected, and their F₃ seeds were planted in the field. 98

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 [6].

Mapping of pubescence color gene

Bulked segregant analysis suggested that the *Td* gene was located in chromosome 3. Seven simple sequence repeat (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 Satt202, 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 (Fig. 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 (Fig. 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 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 (Fig. 2). Genome sequence alignment of the plant introductions and cultivar Bay indicated that no next generation sequencer (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: Fig. S1). We designed a forward primer in the upstream region to amplify the entire coding region of Glyma.03G258700 (Table 2 and Additional File 1: Fig. S1).

Gene cloning

RT-PCR using primers for Glyma.03G258700 generated fragments of approximately 1 kb length in Harosoy-*T*, Clark-*td*, and Cloud. The coding region of Harosoy-*T* was 840 bp long and encoded 279 amino acids (Fig. 3A). BLAST analysis confirmed that Glyma.03G258700 belongs to an MYB transcription factor of the R2R3 type (Fig. 3A). A single base was substituted from G to A at the nucleotide position 633 in Clark-*td* compared with Harosoy-*T*. The single nucleotide polymorphism (SNP) generated a premature stop codon (TGA) resulting in a short polypeptide consisting of 210 amino acids (Fig. 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 (Fig. 3B). Sequencing of the 2nd 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 [26] (Fig. 4). Particularly, the upper region, including R2R3 MYB-binding domains as well as the tail region, were quite similar. Plant MYB proteins are classified into 28 subclasses (S1 to S28), and many of them are involved in flavonoid biosynthesis [27]. 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 [28-31]. Based on phylogenetic analysis, Glyma.03G258700 belongs to S5 together with *DkMYB4*, whereas *R* and *W2* genes belong to S27 and S6, respectively (Fig. 5 and Additional File 2: Table S1).

Derived cleaved amplified polymorphic sequences (dCAPS) analysis

The scheme of dCAPS analysis for SNP and indel are shown in Fig. 6A and 6B, respectively. Results of dCAPS analyses are shown in Fig. 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 *BsrBI* generated a fragment of approximately 185 bp in Korean, Seneca, Kingwa, Grant, Sooty, and Clark-*td*, whereas PCR 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 with PI 157421, whereas Cloud has the same deletion with PI 549046.

Gene expression

The transcript level of Glyma.03G258700 in pubescence was extremely high (151.6 times of the immature seed), whereas that in the other tissues were comparable with immature seed (leaf: 0.57 times; stem: 0.46 times; root: 2.26 times; root nodule: 5.02 times; flower: 1.38 times) in Harosoy-*T* (Fig. 7A). In pubescence, the transcript level of Glyma.03G258700 was approximately 12% in Clark-*td* compared with that in Harosoy-*T* (Fig. 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 (Fig. 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 (Fig. 8).

Discussion

Pubescence color of soybean is controlled by two genes *T* and *Td* [1, 6]. The *T* gene encodes a flavonoid 3-O-hydroxylase and hydroxylates the 3-O-position of the B-ring to generate dihydroxylated flavonoids [4, 5]. The dominant *T* allele generates luteolin derivatives and tawny color in pubescence, whereas the recessive *t* allele produces apigenin derivatives and gray color [7]. 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 [7]. 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 at the end of chromosome 3, which is consistent with the results of GWAS analysis [12, 20]. Flavones are catalyzed by two functional copies of *FNS II* genes, *FNSII-1* and *FNSII-2*, in chromosome 12 [10, 11]. 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 amplified the entire cDNA of Harosoy-*T*, Clark-*td*, and Cloud. The 58-nucleotide sequence of these cDNAs was completely different from the reference sequence of Williams 82 (Additional File 1: Fig. S1). Genome sequence of Williams 82 needs to be confirmed at this region.

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 [6]. R2R3 MYB transcription factors have a conserved DNA binding domain at the N terminus and a more variable C-terminal transcriptional activation domain [32]. These results suggest that Glyma.03G258700 corresponds to the *Td* gene and that premature stop codons dysfunction the transactivating ability of the MYB protein.

There are three types of mutations in the gene (type 1: Grant, Korean, Kingwa, Sooty, Seneca Clark-*td*, and PI 157421; type 2: Cloud and PI 549046; type 3: PI 84631). These types were not associated with pubescence color (light tawny or near-gray), 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, a mRNA surveillance mechanism that eliminates aberrant mRNAs [33]. Similar observations were made with pubescence color gene *T* (flavonoid 3'-hydroxylase) and flower color gene *Wm* (flavonol synthase) in soybean [5, 34].

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, proanthocyanins, or flavonols (Additional File 2: Table S1) [27]. The *Td* gene specifically controls color of pubescence, but not changes in the color of flower, seed, or hypocotyl [2, 6]. The allele at the *Td* locus did not affect the amounts of anthocyanins, flavonol glycosides, or dihydroflavonol in flower petals [35]. In addition, the present study revealed that Glyma.03G258700 expresses predominantly in pubescence where it almost exclusively deposits flavones. These results suggest a hypothesis that Glyma.03G258700 has evolved to upregulate biosynthesis of flavones. Further investigations including expression assays of other flavonoid biosynthesis genes may be necessary to ascertain the hypothesis.

Glyma.03G258700 protein may bind to the promoter of *FNS II* genes and upregulate their transcription. Essentially, 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 proved to bind to the MYBCORE element by electrophoretic mobility shift assays [26]. These results strongly suggest that the Glyma.03G258700 protein may also bind to a similar element in the promoter of *FNSII* genes.

In summary, the wild type Glyma.03G258700 protein may bind to the promoter of *FNS II* genes and upregulates 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 [7]. Mutated MYB protein may have failed to promote catalysis of flavanones to flavones. The remaining flavanones may have been catalyzed by isoflavone

synthase and generated isoflavonoids. Electrophoretic mobility shift assays may be necessary to ascertain the interaction between the Glyma.03G258700 protein and the cis-acting regulatory elements existing 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 elements involved in pubescence-specific gene expression and develop a genetic system to overproduce specific proteins in the pubescence of soybean.

Conclusion

This study revealed that soybean *Td* gene encodes a unique R2R3 MYB transcription factor, which controls flavone content and pubescence color. The wild type of MYB protein binds to the promoter of *FNSII* genes and upregulates their expression, resulting in higher flavone content and deeper pubescence color. Loss-of-function mutation of the MYB gene fails to promote expression of *FNSII* genes, resulting in lower flavone content and dilute pubescence color.

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 TdT*) in 2004. Flowers of Clark-*td* were emasculated one 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 [14]. 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 [6] 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. 30 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 [15]. 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 subjected to bulked segregant SSR analysis. Polymorphic SSR markers in Harosoy x Clark (Nebraska) population [16] were used for analyzing variation between the two bulked DNA samples. PCR conditions were similar to those in a previous report [17]. Variation between the two bulked samples were used to determine which chromosome harbored the pubescence color gene. SSR markers on this chromosome [18] were then tested for variation among the 94 F₂ plants. The linkage map was constructed using MAPMAKER/EXP version 3.0 [19] 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>).

NGS data analysis

Based on information provided by Zhou et al. [20], 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 NGS data of US cultivar Bay with gray pubescence were uploaded to the Integrative Genomics Viewer (<http://software.broadinstitute.org/software/igv/>) [21] 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*, and Cloud 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, <http://www.phytozome.net/soybean.php>). The total volume of

50 μ l PCR mixture contained 10 pmol of each primer, 1 \times PrimeSTAR Max Premix (Takara) and 200 ng cDNA. The PCR program was 35 cycles of 10 sec denaturation at 98 $^{\circ}$ C, 15 sec annealing at 56 $^{\circ}$ C and 1 min extension at 72 $^{\circ}$ 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*, and Clark-*td* by the CTAB method. The second exon was amplified using the PCR primers in Table 2. PCR mixture contained 10 pmol of each primer, 1 \times PrimeSTAR Max Premix, and 50 ng genomic DNA in a total volume of 50 μ l. The PCR program was 35 cycles of 10 sec denaturation at 98 $^{\circ}$ C, 5 sec annealing at 56 $^{\circ}$ C, and 10 sec extension at 72 $^{\circ}$ 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 [22]. Nucleotide and amino acid sequences were aligned using ClustalW (<http://clustalw.ddbj.nig.ac.jp/top-j.html>) with default settings. An additional data file shows the amino acid alignment used 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/>) [23]. Bootstrap test of 1000 replications was performed. Nucleotide sequences of the putative promoter region (up to 1500 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 [24].

dCAPS analysis

Two pairs of PCR primers containing mismatched nucleotides were designed to detect a 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 sec denaturation at 98 $^{\circ}$ C, 5 sec annealing at 56 $^{\circ}$ C, and 10 sec extension at 72 $^{\circ}$ 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 3 replicates. Pubescence was sampled with razor blades from pods and stems of Harosoy-*T* and Clark-*td* at the R3 stage [25]. Flower 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 µg 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. The real-time PCR mixture contained 1 × TB Green *Premix Ex Taq* (Tli RNaseH Plus) (Takara), 10 pmol of each primer, 1 × ROX reference dye, 50 ng of cDNA, and water to a final volume of 20 µl. The PCR was performed using the StepOnePlus Real-Time PCR instrument (Thermo Fisher Scientific). The initial 30 sec denaturation at 95 °C was followed by 40 cycles of 5 sec denaturation at 95 °C and 30 sec 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.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Availability of data and material

Sequence data from this article has been deposited to DDBJ, with the accession numbers LC485152, LC485153 and LC485154. The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

FY is supported by National Natural Science Foundation of China, Transgenic Breeding Program of China (2016ZX08004-003) and the Fundamental Research Funds for the Central Universities. SMG is funded by

the Invitational Fellowships for Research in Japan (The Japan Society for the Promotion of Science).

Authors' contributions

The authors declare that they have no competing interests.

Acknowledgements

We thank Dr. R.L. Nelson at USDA/ARS Univ. of Illinois for supplying the seeds of the NILs, the Genetic Type Collections and the cultivars. We are grateful to Dr. Qiu Lijuan at Institute of Crop Science, Beijing, China for help in NGS analysis of Bay.

References

1. Palmer R, Pfeiffer T, Buss G, Kilen T: Qualitative genetics. In: *Soybeans: Improvement, Production, and Uses*. Edited by Boerma H, Specht J, 3rd edition edn. Madison, WI: American Society of Agronomy; 2004.
2. Palmer RG, Payne RC: Genetic-Control of Hypocotyl Pigmentation among White-Flowered Soybeans Grown in Continuous Light. *Crop Sci* 1979, 19(1):124-126.
3. Murai Y, Yan F, Iwashina T, Takahashi R: Analysis of anthocyanin pigments in soybean hypocotyl. *Canadian Journal of Plant Science* 2016, 96(6):935-938.
4. Buttery BR, Buzzell RI: Varietal Differences in Leaf Flavonoids of Soybeans. *Crop Sci* 1973, 13(1):103-106.
5. Toda K, Yang DJ, Yamanaka N, Watanabe S, Harada K, Takahashi R: A single-base deletion in soybean flavonoid 3'-hydroxylase gene is associated with gray pubescence color. *Plant Molecular Biology* 2002, 50(2):187-196.
6. Bernard RL: The inheritance of near-gray pubescence color. *Soyb Genet Newsl* 1975, 2:31-35.
7. Iwashina T, Benitez ER, Takahashi R: Analysis of flavonoids in pubescence of soybean near-isogenic lines for pubescence color loci. *J Hered* 2006, 97(5):438-443.
8. Wollenweber E: Flavones and flavonols. In: *The flavonoids, advances in research since 1986*. Edited by Harborne JB. London, UK: Chapman & Hall; 1994: 259-335.
9. Martens S, Mithofer A: Flavones and flavone synthases (vol 66, pg 2399, 2005). *Phytochemistry* 2006, 67(5):521-521.

10. Fliegmann J, Furtwangler K, Malterer G, Cantarello C, Schuler G, Ebel J, Mithofer A: Flavone synthase II (CYP93B16) from soybean (*Glycine max* L.). *Phytochemistry* 2010, 71(5-6):508-514.
11. Jiang YN, Wang BA, Li H, Yao LM, Wu TL: Flavonoid Production Is Effectively Regulated by RNA1 Interference of Two Flavone Synthase Genes from *Glycine max*. *Journal of Plant Biology* 2010, 53(6):425-432.
12. Wen ZX, Boyse JF, Song QJ, Cregan PB, Wang DC: Genomic consequences of selection and genome-wide association mapping in soybean. *Bmc Genomics* 2015, 16.
13. Bandillo NB, Lorenz AJ, Graef GL, Arquin D, Hyten DL, Nelson RL, Specht JE: Genome-wide Association Mapping of Qualitatively Inherited Traits in a Germplasm Collection. *Plant Genome* 2017, 10(2).
14. Bernard RL, Nelson RL, Cremeens CR: USDA soybean genetic collections: isoline collection. *Soybean Genetics Newsletter* 1991, 18:27-57.
15. Murray MG, Thompson WF: Rapid Isolation of High Molecular-Weight Plant DNA. *Nucleic Acids Res* 1980, 8(19):4321-4325.
16. Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung L *et al*: An integrated genetic linkage map of the soybean genome. *Crop Sci* 1999, 39(5):1464-1490.
17. Rojas Rodas F, Rodriguez TO, Murai Y, Iwashina T, Sugawara S, Suzuki M, Nakabayashi R, Yonekura-Sakakibara K, Saito K, Kitajima J *et al*: Linkage mapping, molecular cloning and functional analysis of soybean gene Fg2 encoding flavonol 3-O-glucoside (1 → 6) rhamnosyltransferase. *Plant Mol Biol* 2014, 84(3):287-300.
18. Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB: A new integrated genetic linkage map of the soybean. *Theor Appl Genet* 2004, 109(1):122-128.
19. Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg LA: MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1987, 1(2):174-181.
20. Zhou ZK, Jiang Y, Wang Z, Gou ZH, Lyu J, Li WY, Yu YJ, Shu LP, Zhao YJ, Ma YM *et al*: Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology* 2015, 33(4):408-U125.
21. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: Integrative genomics viewer. *Nature Biotechnology* 2011, 29(1):24-26.

22. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997, 25(17):3389-3402.
23. Kumar S, Stecher G, Li M, Knyaz C, Tamura K: MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution* 2018, 35(6):1547-1549.
24. Higo K, Ugawa Y, Iwamoto M, Korenaga T: Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* 1999, 27(1):297-300.
25. Fehr WR, Caviness CE, Burmood DT, Penningt.Js: Stage of Development Descriptions for Soybeans, Glycine-Max (L) Merrill. *Crop Sci* 1971, 11(6):929-&.
26. Akagi T, Ikegami A, Tsujimoto T, Kobayashi S, Sato A, Kono A, Yonemori K: DkMyb4 Is a Myb Transcription Factor Involved in Proanthocyanidin Biosynthesis in Persimmon Fruit. *Plant Physiology* 2009, 151(4):2028-2045.
27. Liu JY, Osbourn A, Ma PD: MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. *Molecular Plant* 2015, 8(5):689-708.
28. Gillman JD, Tetlow A, Lee JD, Shannon JG, Bilyeu K: Loss-of-function mutations affecting a specific Glycine max R2R3 MYB transcription factor result in brown hilum and brown seed coats. *Bmc Plant Biology* 2011, 11.
29. Takahashi R, Benitez ER, Oyoo ME, Khan NA, Komatsu S: Nonsense Mutation of an MYB Transcription Factor Is Associated with Purple-Blue Flower Color in Soybean. *J Hered* 2011, 102(4):458-463.
30. Takahashi R, Yamagishi N, Yoshikawa N: A MYB Transcription Factor Controls Flower Color in Soybean. *J Hered* 2013, 104(1):149-153.
31. Yan F, Di SK, Takahashi R: CACTA-superfamily transposable element is inserted in MYB transcription factor gene of soybean line producing variegated seeds. *Genome* 2015, 58(8):365-374.
32. Pireyre M, Burow M: Regulation of MYB and bHLH Transcription Factors: A Glance at the Protein Level. *Molecular Plant* 2015, 8(3):378-388.
33. Chang YF, Imam JS, Wilkinson ME: The nonsense-mediated decay RNA surveillance pathway. *Annual Review of Biochemistry* 2007, 76:51-74.
34. Takahashi R, Githiri SM, Hatayama K, Dubouzet EG, Shimada N, Aoki T, Ayabe S, Iwashina T, Toda K, Matsumura H: A single-base deletion in soybean flavonol synthase gene is associated with magenta flower color. *Plant Molecular Biology* 2007, 63(1):125-135.

35. Iwashina T, Githiri SM, Benitez ER, Takemura T, Kitajima J, Takahashi R: Analysis of flavonoids in flower petals of soybean near-isogenic lines for flower and pubescence color genes. *J Hered* 2007, 98(3):250-257.

Supplementary Legends

Additional files Additional File 1: Figure S1 Comparison of nucleotide sequence around the 1st exon of soybean Glyma.03G258700 among reference genome of cultivar Williams 82 (Wm82.a2.v1), genome of three plant introductions and cultivar Bay, and cDNA of Harosoy-T. Genome alignment was created by the Integrative Genomics Viewer. Nucleotide position in chromosome 3 is indicated by arrows above panel. Genome sequence of Williams 82 is shown below panel and aligned with cDNA of Harosoy-T. Identical nucleotides are shown by asterisks. PCR primer for cDNA cloning is single-underlined. Start codon is double-underlined. Exon and 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.

Figures

chromosome 3

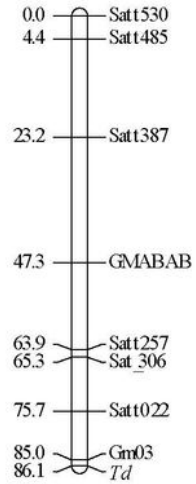


Figure 1

Linkage mapping of soybean pubescence color gene *Td* using an F₂ 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 SSR marker, Gm03:45226162-45226406.

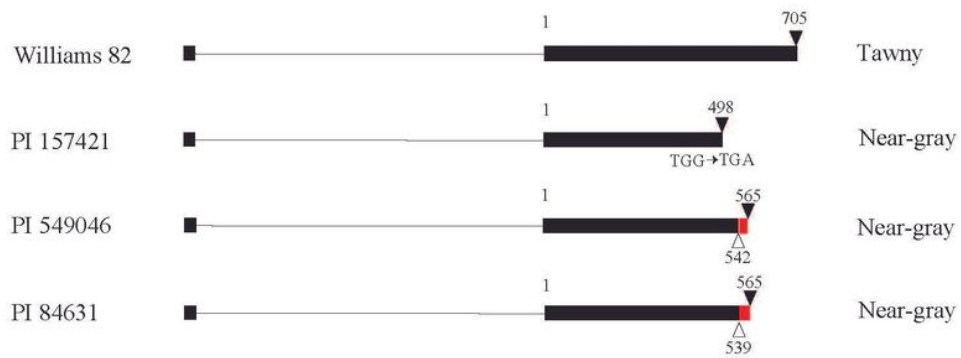


Figure 2

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-T	<u>MGRAPCCSKVGLHRGPWTPREDALLTKYIQAHGKGQWRSLEPKTAG</u>	45
	<u>LLRCGKSCRLRWMNYLRPDIKRGNITPEEDDLIIRMHSLGNRWS</u>	90
	<u>LIAGRLPGRTDNEIKNYWNTHLSKCLRNRQGTDPNTHDKLTEAPEK</u>	135
	KKGRKKNKQKSTENKNGSEKTLVYLPKPIRVKALSSCIPRTDSTL	180
	TLDSNSATTASTSQEKVQSPEAETKEVNMVWGVSDNIGDNGGIGI	225
	↓ FFGEDHDLVNSASYMECYSDDHD TLEKLYEEYLQLLNVEEK PDEL	270
	DSFAQSLLV*	279

B

Harosoy-T	ACAAAAGAGGTGAACATGGTTTGGGGGTAAGTGACAATGGT	663
	T K E V N M V W G V S D N G	
Clark-td	ACAAAAGAGGTGAACATGGTTTGGGGGTAAGTGACAATGGT	663
	T K E V N M V *	
Harosoy-T	AATGGTGGGATTGGAATATCTTTGGTGAAGACCATGACCTAG	700
	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. (A) Amino acid sequence of Harosoy-T. The position of SNP and indel found in Clark-td and Cloud are indicated by red and black arrows, respectively. R2R3 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

Glyma.03G258700	MGRAPCCSKVGLHRGPWTPRE <u>FA</u> LLTKYIQAHG QW RS LP KT AG LL RC GK SC RL RM NY	60
<i>DkMYB4</i>	MGRAPCCSKVGLHRGPWTK <u>ED</u> GL LL TKYIQVHG Q W RS LP KK AG LL RC GK SC RL RM NY	60
Glyma.03G258700	LRPDI KRG NI T P E D D L I R M S L L G N R W S L I A G R L P G R T D N E I K N Y W N T H L S K L R N Q G	120
<i>DkMYB4</i>	LRPDI KRG NI T P E D D L I R M S L L G N R W S L I A G R L P G R T D N E I K N Y W N T H L S K L R S Q G	120
Glyma.03G258700	TDP N T H K L T E A P E --- K R K G R K K N K O K S T E N N K G S E R T L V L P K P T R M K A L S S C I P R T	176
<i>DkMYB4</i>	TDP N T H K K L S D S H V Q E P K K S S N K K K K S K S N L D H T E K L V N E K P E R L K S L A S F S F S R	180
Glyma.03G258700	D S T L T L D N S A T T A S T S Q E K V Q S P E A E T K E V N M V W G V S N I G N G G I G L F F G E L - H D L V N	235
<i>DkMYB4</i>	D S S S F D W I T T A T A T A P S ----- G S E N H E G E R G M L G N G S N G H E V G F I G E D G S H Y H	232
Glyma.03G258700	S A S Y M C Y S D H D T L E K L Y E E Y L Q L L V E E K P T - E L D S F A Q S L L M	279
<i>DkMYB4</i>	H D H M L D S D L C Q S L E K L Y E E Y L Q L L T E E D D G L D S F A E S L L L	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. R2R3 MYB DNA-binding domains are underlined.

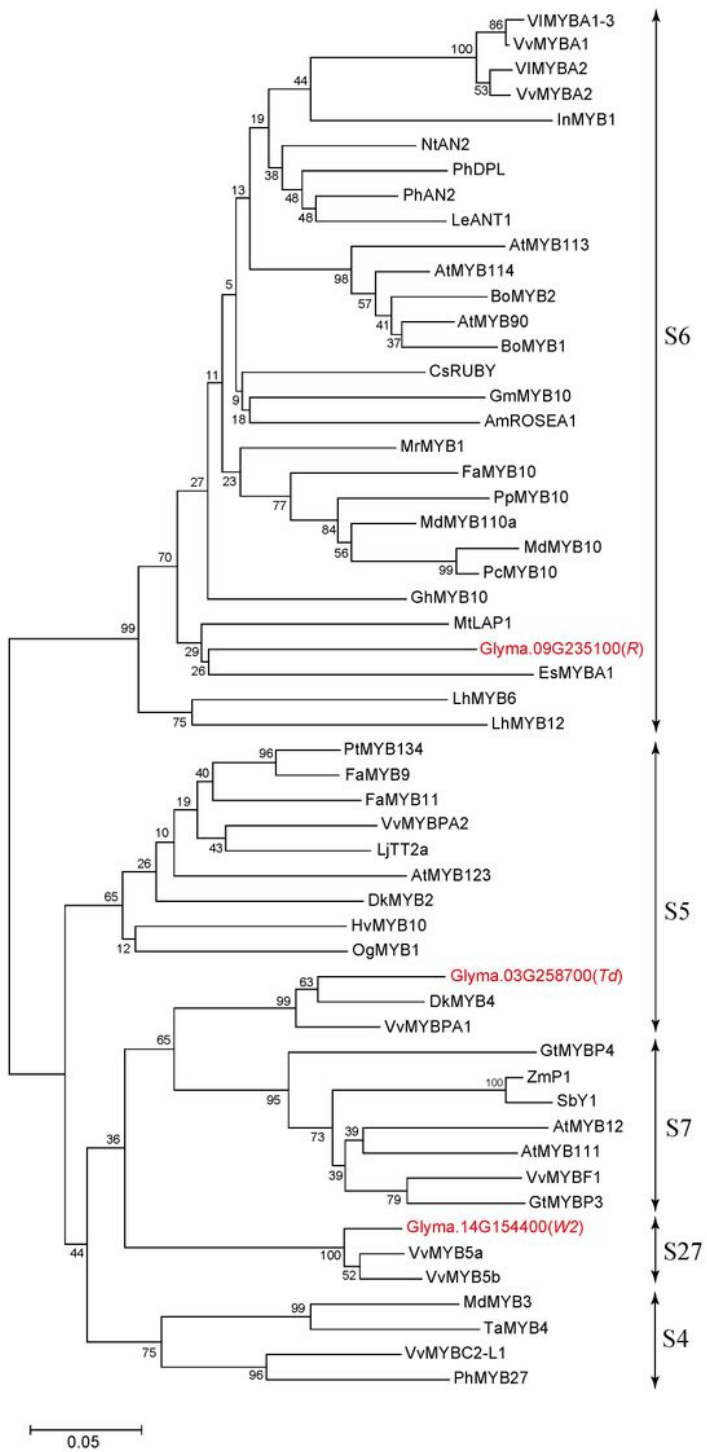


Figure 5

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 substitutions/site.

A

Partial cDNA sequence of Glyma.03G258700 gene

PCR primers: 5' - AAAGGGTCAGAGAAGACCTTA CGGGTAAGTGACAATATTGGA - 5'
Harosoy-T: 5' - AACAAAGGGTCAGAGAAGACCTTAGTTT---TGGGGGGTAAGTGACAATATTGGAGA
Clark-td: 5' - AACAAAGGGTCAGAGAAGACCTTAGTTT---TGAGCGGGTAAGTGACAATATTGGAGA

↓ PCR reaction

PCR product

Harosoy-T: 5' - AAAGGGTCAGAGAAGACCTTAGTTT---TGGCGGGTAAGTGACAATATTGGA
Clark-td: 5' - AAAGGGTCAGAGAAGACCTTAGTTT---TGAGCGGGTAAGTGACAATATTGGA
BsrBI site

B

Partial cDNA sequence of Glyma.03G258700 gene

PCR primers: 5' - AGACAATGGTGGGATTGGGA TGCTAGCACGTGAACATATT - 5'
Harosoy-T: 5' - TGGAGACAATGGTGGGATTGGAAATATTCTT---CCTGCTAGCACGTGAACATATTGTA
Cloud: 5' - TGGAGACAATGGTGGGATTGGAAATAT-CTT---CCTGCTAGCACGTGAACATATTGTA

↓ PCR reaction

PCR product

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

C

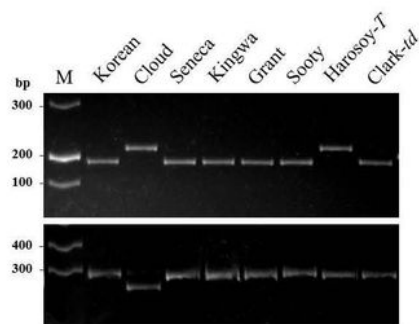
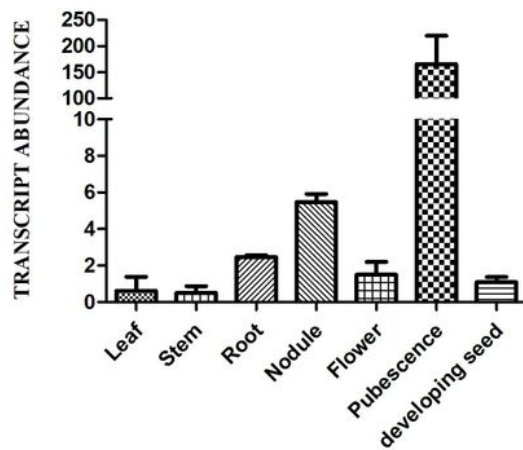


Figure 6

Schematic presentation and results of dCAPS analysis. (A) dCAPS analysis to detect a 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 fond. (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

A



B

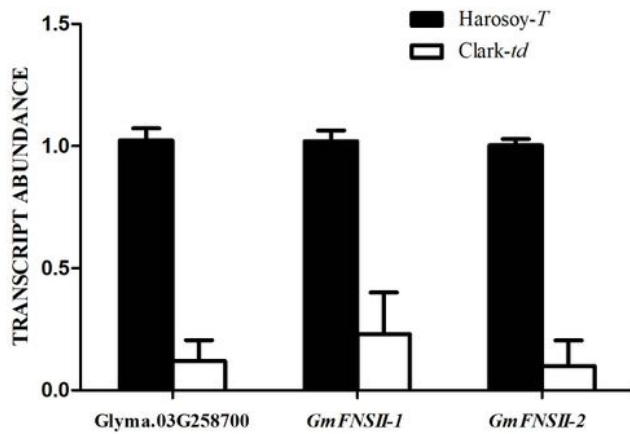


Figure 7

Relative expression of MYB gene (Glyma.03G258700) and two flavone synthase II genes (GmFNSII-1 and GmFNSII-2) in soybean near-isogenic lines. (A) Relative expression of Glyma.03G258700 in various tissues of Harosoy-T. (B) Relative expression of Glyma.03G258700, GmFNSII-1 and GmFNSII-2 in pubescence of Harosoy-T and Clark-td. Transcript levels were standardized by the expression of actin 1 gene. Means and SDs of three biological replications are shown.

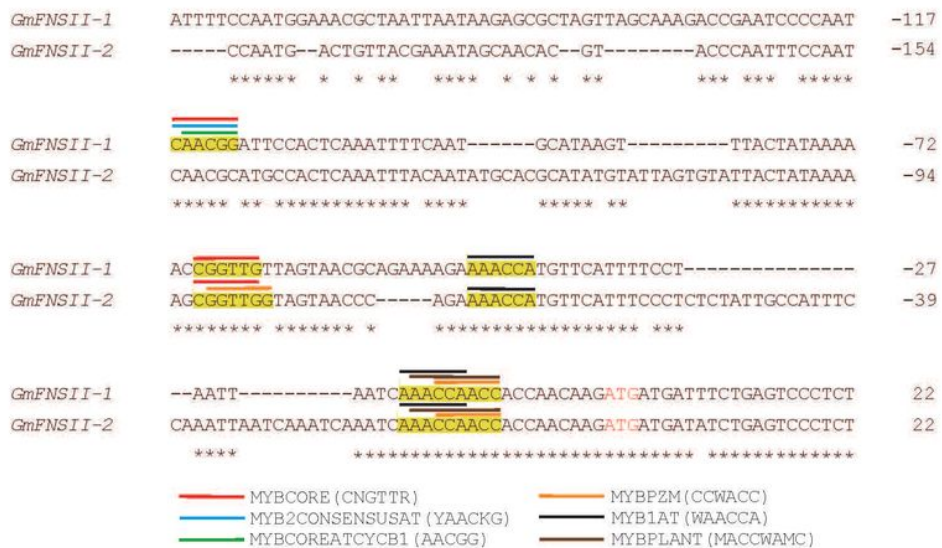


Figure 8

Nucleotide alignment of upstream region of two soybean flavone synthase II genes, GmFNSII-1 and GmFNSII-2, derived from reference genome of cultivar Williams 82. Dashes represent gaps introduced to improve the alignment. Identical nucleotides are indicated by asterisks. Start codon is shown by red font. Putative cis-acting regulatory element regions for MYB transcription factors are highlighted in yellow. Position of each element is shown by bars above nucleotide sequences.

Supplementary Files

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

- [supplement1.xlsx](#)
- [supplement2.docx](#)
- [supplement3.pdf](#)