

## A single-base deletion in soybean flavonol synthase gene is associated with magenta flower color

Ryoji Takahashi · Stephen M. Githiri · Kouta Hatayama · Emilyn G. Dubouzet · Norimoto Shimada · Toshio Aoki · Shin-ichi Ayabe · Tsukasa Iwashina · Kyoko Toda · Hisakazu Matsumura

Received: 26 June 2006 / Accepted: 12 August 2006 / Published online: 28 September 2006  
© Springer Science+Business Media B.V. 2006

**Abstract** The *Wm* locus of soybean [*Glycine max* (L.) Merr.] controls flower color. Dominant *Wm* and recessive *wm* allele of the locus produce purple and magenta flower, respectively. A putative full-length cDNA of flavonol synthase (FLS), *gmfls1* was isolated by 5' RACE and end-to-end PCR from a cultivar Harosoy with purple flower (*WmWm*). Sequence analysis revealed that *gmfls1* consisted of 1,208 nucleotides encoding 334 amino acids. It had 59–72% homology with FLS proteins of other plant species. Conserved

dioxygenase domains A and B were found in the deduced polypeptide. Sequence comparison between Harosoy and Harosoy-*wm* (magenta flower mutant of Harosoy; *wmwm*) revealed that they differed by a single G deletion in the coding region of Harosoy-*wm*. The deletion changed the subsequent reading frame resulting in a truncated polypeptide consisting of 37 amino acids that lacked the dioxygenase domains A and B. Extracts of *E. coli* cells expressing *gmfls1* of Harosoy catalyzed the formation of quercetin from dihydroquercetin, whereas cell extracts expressing *gmfls1* of Harosoy-*wm* had no FLS activity. Genomic Southern analysis suggested the existence of three to four copies of the FLS gene in the soybean genome. CAPS analysis was performed to detect the single-base deletion. Harosoy and Clark (*WmWm*) exhibited longer fragments, while Harosoy-*wm* had shorter fragments due to the single-base deletion. The CAPS marker cosegregated with genotypes at *Wm* locus in a F<sub>2</sub> population segregating for the locus. Linkage mapping using SSR markers revealed that the *Wm* and *gmfls1* were mapped at similar position in the molecular linkage group F. The above results strongly suggest that *gmfls1* represents the *Wm* gene and that the single-base deletion may be responsible for magenta flower color.

R. Takahashi (✉) · S. M. Githiri · K. Hatayama · E. G. Dubouzet · K. Toda · H. Matsumura  
National Institute of Crop Science and University of Tsukuba, Tsukuba, Ibaraki 305-8518, Japan  
e-mail: masako@affrc.go.jp

**Present Address:**  
S. M. Githiri  
Department of Crop Science, University of Nairobi,  
P.O. Box 30197, Nairobi, Kenya

**Present Address:**  
K. Hatayama  
National Institute of Advanced Industrial Science and Technology, Biomass Technology Research Center, Kure, Hiroshima 737-0197, Japan

**Present Address:**  
E. G. Dubouzet  
National Institute of Advanced Industrial Science and Technology,  
Tsukuba, Ibaraki 305-8666, Japan

N. Shimada · T. Aoki · S. Ayabe  
Nihon University, Fujisawa, Kanagawa 252-8510, Japan

T. Iwashina  
Tsukuba Botanical Garden, National Science Museum,  
Tsukuba, Ibaraki 305-0005, Japan

**Keywords** Deletion · FLS · Flavonol synthase · Flower color · Soybean · *Wm* locus

**DDBJ accession number** AB246668

### Abbreviations

FLS Flavonol synthase  
F3'H Flavonoid 3'-hydroxylase  
F3'5'H Flavonoid 3'5'-hydroxylase

## Introduction

Flavonoids are important secondary metabolites in the plant kingdom. Individual plant species synthesize a variety of flavonoid compounds that function in providing flower pigmentation to attract pollinators, in defending plants against pathogens, in acting as signal molecules in plant–microbe interactions, and in protecting plants from UV radiation (Shirley 1996, 2001; Dixon and Steele 1999).

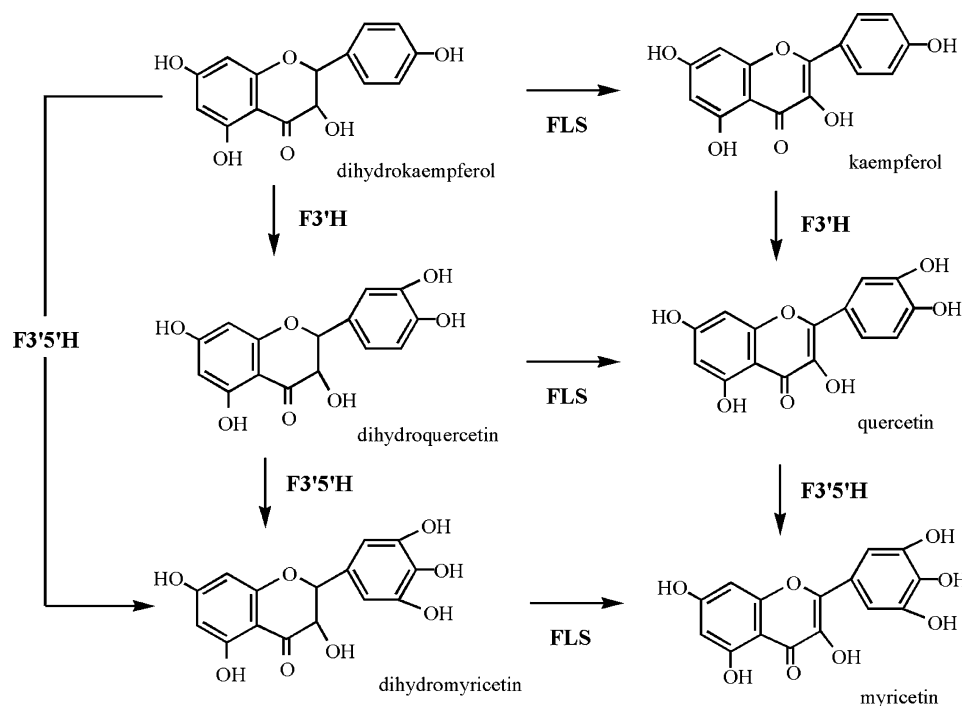
The biosynthetic pathway of flavonoids is well established and many of the structural and some of the regulatory genes have been cloned in several model plants including maize, petunia and snapdragon (Holton and Cornish 1995). Anthocyanins and flavonols are both derived from dihydroflavonols (dihydrokaempferol, dihydroquercetin, and dihydromyricetin). The B-ring of dihydrokaempferol can be hydroxylated at the 3' position by flavonoid 3'-hydroxylase (F3'H), or at the 3' and 5' positions by flavonoid 3',5'-hydroxylase (F3'5'H) to produce dihydroquercetin and dihydromyricetin, respectively. Flavonols are formed from dihydroflavonols, as illustrated in Fig. 1, by the introduction of a double bond between C-2 and C-3 through the action of flavonol synthase (FLS) (Forkmann 1991).

Five genes (*T*, *W1*, *I*, *R* and *O*) control the color of the seed coat, five genes (*W1*, *W3*, *W4*, *Wm* and *Wp*) control flower color, and two genes (*T* and *Td*) control the color of the pubescence in soybean (reviewed by

Palmer et al. 2004). RFLP marker of dihydroflavonol reductase co-segregated with flower color variants in an F<sub>2</sub> population segregating for *W3* locus (Fasoula et al. 1995). Genotypes at *I* locus were associated with chalcone synthase gene duplications (Todd and Vodkin 1996). Accordingly, *W3* and *I* are presumed to encode dihydroflavonol reductase and chalcone synthase, respectively. Chromatographic experiments indicated that alleles at *T* and *W1* loci are associated with 3'- and 3',5'-B-ring hydroxylation, respectively. The results suggested that *T* and *W1* encode F3'H and F3'5'H, respectively (Buttery and Buzzell 1973; Buzzell and Buttery 1982). Toda et al. (2002) cloned and characterized the F3'H cDNA from a pair of near-isogenic lines for the *T* locus, To7B (*TT*, tawny pubescence) and To7G (*tt*, gray pubescence). Sequence analysis revealed that they differed by a single-base deletion of C in the coding region of To7G. The deletion generated a truncated polypeptide lacking the GGEEK consensus sequence of F3'H gene and the heme-binding domain resulting in non-functional protein. Further, cDNA microarray suggested that *Wp* corresponds to flavanone 3-hydroxylase gene (Zabala and Vodkin 2005).

Magenta flower mutant of Harosoy (Harosoy-*wm*, T235) was found at Urbana, IL in 1957 (Fig. 2, Buzzell and Buttery 1977). They found that a single-gene controlled the phenotype and designated the gene as *Wm*. They found a linkage between *W1* and *Wm* with a

**Fig. 1** Schematic diagram of flavonol biosynthesis from dihydroflavonols. Enzymes involved in each step are indicated as follows:  
F3'H flavonoid 3'-hydroxylase,  
F3'5'H flavonoid 3',5'-hydroxylase and FLS flavonol synthase



**Fig. 2** Flowers of Harosoy (purple, genotype *WmWm*) and Harosoy-*wm* (magenta, genotype *wmwm*)



recombination value of  $2.2 \pm 0.7\%$ . The two loci were assigned to the middle region of molecular linkage group F (Cregan et al. 1999). Buzzell and Buttery (1977) further found that the recessive allele *wm* is associated with low levels of flavonol glycosides in the leaves and flowers. It suggests that *Wm* is responsible for the production of flavonol and may encode either FLS or a regulatory gene controlling expression of FLS gene. This study was conducted to isolate and characterize the FLS gene of soybean, and to investigate the relationship between FLS gene and the *Wm* locus.

## Materials and methods

### Plant materials

Purple flower cultivars, Clark and Harosoy (*WmWm*), and Harosoy-*wm* (*wmwm*) were used. Seeds of Harosoy-*wm* were provided by the USDA Soybean Germplasm Collections. Plants were grown in fields supplemented with a synthetic fertilizer (3.0 N–10.0P<sub>2</sub>O<sub>5</sub>–10.0 K<sub>2</sub>O kg/10a) at National Institute of Crop Science, Tsukuba, Japan. For genetic analysis, Harosoy-*wm* was crossed with Clark in 2004. Hybridity of F<sub>1</sub> plants was ascertained by purple flower color. One hundred and fourteen plants of the F<sub>2</sub> population were grown in fields in 2005.

### RNA extraction

Total RNA was extracted from frozen trifoliolate leaves (5 g) sampled at R3 stage (Fehr et al. 1971) and from frozen banner flower petals (500 mg) of Clark, Harosoy and Harosoy-*wm* using the TRIZOL Reagent (Invitrogen) according to the manufacturer's instructions.

### cDNA cloning

Three reverse primers, gmfls-GSP1 (5'-CACTCCTATTAGGTGCTTGCA-3'), gmfls-GSP2 (5'-AGGCCACACAATGTGGAACAA-3') and gmfls-GSP3 (5'-ATGATCCACCCACCCTTCTT-3') were constructed for 5' RACE. 2.5 µg of the total RNA from flower petals of Harosoy was reverse transcribed using gmfls-GSP1 as a primer. After purification and terminal transferase reaction, the resulting TdT product was used as a template in first-round PCR in the presence of the Abridged Anchor Primer (5' RACE System ver. 2.0, Gibco-BRL) and gmfls-GSP2. A 2 min denaturation at 94°C was followed by 35 cycles of 30 s denaturation at 94°C, 30 s annealing at 55°C, and 1 min extension at 72°C. A final 5 min extension at 72°C completed the program. The PCR was performed in an Applied Biosystems 9700 thermal cycler. For the second round PCR, 5 µl of the diluted first-round PCR product was reamplified with AUAP and gmfls-GSP3. The PCR product with molecular size of about 500 bp was cloned into pCR 2.1 vector (Invitrogen) and sequenced. A putative full-length clone of about 1,200 bp was obtained from Harosoy and Harosoy-*wm* by PCR using specific primers at the extreme 5' end (5'-GAAATCAACACACAAACACAACAAC-3') and 3' end (5'-CTACTTCCCACAAGTCACACTTTC-3'), cloned in pCR 2.1, and sequenced.

### DNA sequencing

Nucleotide sequences of both strands were determined with BigDye terminator cycle method using an ABI3100 Genetic Analyzer (Applied Biosystems). Nucleotide and amino acid sequences were analyzed with the BLAST program (Altschul et al. 1997).

### Expression in *E. coli* cells and enzyme assay

The entire coding region of *gmfls1* was cloned from Harosoy and Harosoy-*wm* by RT-PCR using the above 5' primer and a reverse primer (GTGCCAGAA TTCTCATTGAGGGATCTTATTG) that included an additional *EcoRI* site. The amplified bands were digested at the intrinsic *NdeI* and the introduced *EcoRI* sites, and were inserted into a pET21a vector (Novagen). *E. coli* RosettaTM2 (DE3) was transformed with the vectors harboring *gmfls1* or an empty vector. Expression of FLS proteins was induced by adding 1 mM isopropyl-thio- $\beta$ -D-galactoside to 50 ml of overnight culture for 3 h at 28°C. Cells were harvested by centrifugation, washed with 0.1 M potassium-phosphate (pH 7.0) containing 10% sucrose and 1.4 mM 2-mercaptoethanol and collected by centrifugation.

The cell pellets were suspended in 10 ml of 0.1 M potassium phosphate (pH 7.0) containing 10% sucrose and 1.4 mM 2-mercaptoethanol, disrupted by sonication and then centrifuged. The supernatant was used as a crude enzyme extract for the following FLS assay. A 500  $\mu$ l reaction mixture, consisting 0.1 M potassium-phosphate (pH 7.0), 10% sucrose, 1.4 mM 2-mercaptoethanol, 250  $\mu$ M 2-oxoglutarate, 5 mM ascorbic acid, 50  $\mu$ M ferrous sulfate, 10  $\mu$ g ( $\pm$ )-taxifolin (dihydroquercetin; Sigma) and 50  $\mu$ l of crude enzyme extract, was incubated for 1 h at 28°C. The reaction was terminated by extraction with 500  $\mu$ l of ethyl acetate, and the extract was analyzed by HPLC-photodiode array spectrophotometry. HPLC was performed using a TSK-GEL ODS 80TM column (4.6  $\times$  150 mm; TOSOH). Mobile phase solution A was MeOH and solution B consisted of distilled water containing 0.1% phosphoric acid. A gradient program was used for separation at the flow rate of 1 ml/min at 40°C. The gradient program was as follows: 0–15 min, 40–70% (v/v) A; 15–16 min, 70–100% (v/v) A; 16–17 min, 100% (v/v) A; 17–18 min, 100–40% (v/v) A; 18–24 min, 40% (v/v) A. The chromatograms were recorded at the maximum absorbance in the range of 230–400 nm. Quercetin was identified based on the retention time and its UV-Vis spectrum was compared with the standard samples.

### DNA blot analysis

Genomic DNA of Clark, Harosoy and Harosoy-*wm* was isolated from trifoliolate leaves by the CTAB method (Murray and Thompson 1980), and digested with three restriction enzymes, *EcoRI*, *HindIII* and *BglII*. The digested DNA of 5  $\mu$ g per lane was separated by electrophoresis, and blotted on to nylon

membranes (Hybond-N+, Amersham Pharmacia Biotech). Southern hybridization was performed with *gmfls1* probes labeled by the ECL direct labeling system (Amersham Pharmacia Biotech), and hybridization signals were detected by chemiluminescence on X-ray films.

### CAPS analysis

A pair of PCR primers (5'-GAAATCAACACAC ACAAACACAACAAC, 5'-CAAAGAACATTTT-CCCAACACTTTGCA) was designed to amplify a band including the single-base deletion of G found in Harosoy-*wm*. The deletion was expected to generate a *BglII* site (AGATCT) in the amplified band of Harosoy-*wm*. The PCR mixture contained 30 ng of genomic DNA, 5 pmol of each primer, 10 pmol of nucleotides, and 1 unit of ExTaq in 1 $\times$  ExTaq Buffer supplied by the manufacturer (TAKARA BIO) in a total volume of 50  $\mu$ l. After an initial 30 s denaturation at 94°C, there were 30 cycles of 30 s denaturation at 94°C, 1 min annealing at 57°C, and 1 min extension at 72°C. A final 7 min extension at 72°C completed the program. The amplified products were digested with *BglII* and the digests were separated on 8% polyacrylamide gel in 1 $\times$  TBE buffer (90 mM Tris-borate, 2 mM EDTA, pH 8.0). After electrophoresis, the gel was stained with ethidium bromide and the DNA fragments were visualized under UV light.

### SSR analysis and linkage mapping

The PCR mixture for SSR analysis contained 1  $\mu$ l of the extracted DNA that was diluted by ten times in TE-buffer, 5 pmol of primers, 10 pmol of nucleotides, and 1 unit of ExTaq in 1 $\times$  ExTaq Buffer in a total volume of 20  $\mu$ l. A 12 min denaturation at 94°C was followed by 40 cycles of 30 s denaturation at 94°C, 30 s annealing at 50°C, and 30 s extension at 68°C. A final 5 min extension at 68°C completed the program. The PCR products were separated on 8% acrylamide gels in 1 $\times$  TBE buffer, and the fragments were visualized by ethidium bromide staining. A linkage map was constructed using MAPMAKER/ EXP. ver. 3.0 (Lander et al. 1987).

### Semi-quantitative RT-PCR analysis

Semi-quantitative RT-PCR was conducted by reverse-transcribing 5  $\mu$ g of total RNA using the Superscript III First-Strand Synthesis System (Invitrogen) and oligo d(T) primer according to the manufacturer's instruction. To test the transcription level of *gmfls1*, PCR reactions were carried out for 15 or 22 cycles of

94°C 1 min/57°C 1 min/72°C 30 s in a volume of 20 µl, using 1 µl of cDNA. The primers were 5'-ATGGAG GTGCTAAGGGTGCAA-3' (forward) and 5'-AT GATCCACCCACCCTTTCTT-3' (reverse). *SacI*, a soybean actin gene (Shah et al. 1983), was used as a control. PCR reactions were carried out for 30 cycles of 94°C 1 min/57°C 1 min/72°C 30 s in a volume of 20 µl, using 1 µl of cDNA. The primers were 5'-CTGGGG-ATGGTGTGTCAGCCACAC-3' (forward) and 5'-CACCGAACTTTCTCTCGGAAGGTG-3' (reverse). PCR products were loaded on a 2% agarose gel, stained by ethidium bromide and visualized under UV light.

Analysis of flavonoids in flower petals

Two hundred milligrams of banner petals were collected from field-grown Harosoy and Harosoy-*wm* in 2 ml of MeOH containing 0.1% HCl (for anthocyanin) or absolute MeOH (for flavonol) in three replications in 2003 and 2004. Crude extracts of flower petals were analyzed with Shimadzu HPLC systems using Shim-pack CLC-ODS [I.D. 6.0 mm × 150 mm (Shimadzu)] at a flow rate of 1.0 ml/min. Detection wavelength was 190–700 nm and eluents were MeCN/HOAc/H<sub>2</sub>O/

H<sub>3</sub>BO<sub>3</sub> (6:8:83:3) for anthocyanins and MeCN/H<sub>2</sub>O/H<sub>3</sub>BO<sub>3</sub> (22:78:0.2) for flavonols. Amount of flavonoids was evaluated by their peak areas in HPLC chromatogram.

Results

Isolation of a flavonol synthase cDNA clone from soybean

A soybean EST clone that is homologous to FLS gene (GenBank accession number: BE059354) was purchased from the American Type Culture Collection (ATCC), Manassas, VA. Sequence analysis revealed that the clone was a chimera of ubiquitin and a truncated FLS gene lacking the 5'-region. 5' RACE was performed to obtain the entire cDNA. A DNA fragment with approximate size of 500 bp was cloned and sequenced. A putative full-length cDNA of about 1.2 kb, *gmfls1* was obtained by PCR using primers specific to the extreme 5' and 3' ends. Sequence analysis revealed that it contained 1,208 nucleotides (Fig. 3). The polypeptide (GmFLS1) of 334



**Fig. 3 (A)** The cDNA and deduced amino acid sequence of *gmfls1* (DDBJ accession number: AB246668) of Harosoy with purple flower (*WmWm*). The entire sequence shown is 1,208 bases long and contains a putative open reading frame of 334 amino acids shown below the nucleotide sequence. The conserved dioxygenase domains *A* and *B* are underlined. A guanine residue at the nucleotide position 75 (shown in bold) was deleted in Harosoy-*wm* (*wmwm*). **(B)** Comparison of cDNA and

deduced amino acid sequences of *gmfls1* from Harosoy and Harosoy-*wm*. Deletion of a guanine residue (shown in bold) altered the subsequent reading frame, changed the following 22 amino acids, and prematurely terminated the transcription of Harosoy-*wm*. The conserved dioxygenase domains *A* and *B* are absent in Harosoy-*wm*. The deletion generated a *Bgl*II site in Harosoy-*wm* (underlined)

amino acids was deduced from the DNA sequence based on an open reading frame that stretched from the putative 5' end of the cDNA to nucleotide position 1,031. The apparent molecular mass and isoelectric point of GmFLS1 were 38,150 and 5.78, respectively. GmFLS1 had amino acid homology with the previously reported FLS from *Eustoma grandiflorum* (EgFLS, AF240764, 71.6%), *Malus domestica* (MdFLS, AY965343, 71.5%), *Rosa hybrid* (RhFLS, AB038247, 71.0%), *Petunia hybrida* (PhFLS, Z22543, 69.6%), *Solanum tuberosum* (StFLS, X92178, 67.2%), and *Arabidopsis thaliana* (AtFLS, U84259, 59.0%) (Fig. 4). Conserved dioxygenase domains A and B (Holton et al. 1993) were found in the deduced polypeptide.

A deletion mutation occurs in the mutant with a recessive *wm* allele

Sequence analysis of FLS gene from Harosoy and Harosoy-*wm* revealed that a single-base deletion of G occurs at the nucleotide position 75 in Harosoy-*wm* (Fig. 3). Amino acids were identical between

Harosoy and Harosoy-*wm* until amino acid number 15. However, the deletion changed the subsequent reading frame in Harosoy-*wm*, altered the following 22 amino acids, and the polypeptide was terminated thereafter. The truncated polypeptide consisted of only 37 amino acids and lacks the dioxygenase domains A and B.

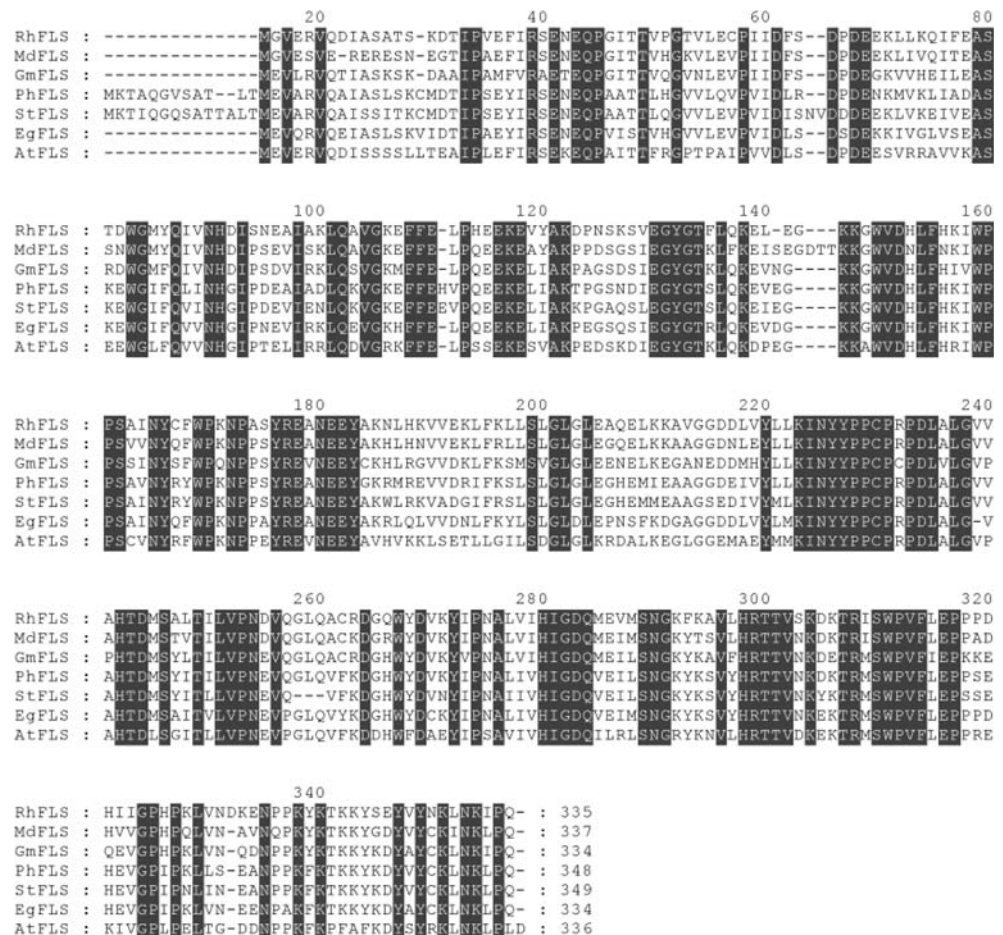
FLS activity

The HPLC chromatogram revealed that extracts of *E. coli* cells expressing Harosoy *gmfls1* catalyzed dihydroquercetin into quercetin, indicating that *gmfls1* encodes FLS (Fig. 5). In contrast, trace amounts of quercetin was detected in extracts of cells transformed with *gmfls1* of Harosoy-*wm* or empty vector, and in the substrate (dihydroquercetin), indicating that GmFLS1 of Harosoy-*wm* has no FLS activity.

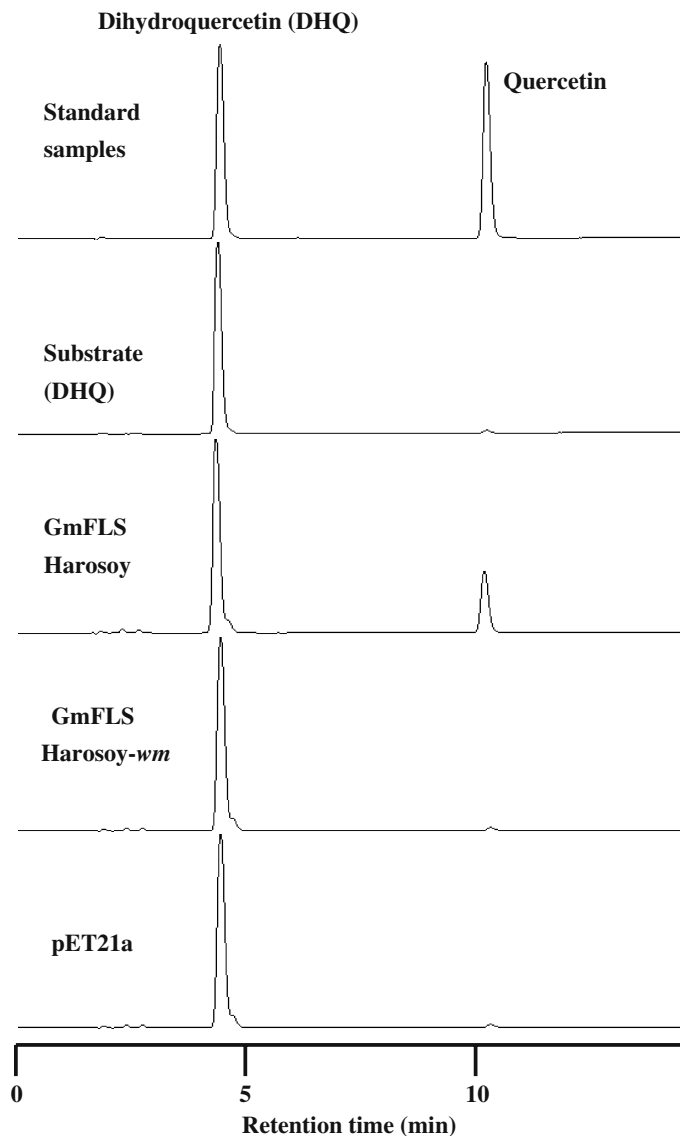
DNA blot analysis

Genomic DNA blots of Clark, Harosoy and Harosoy-*wm* were analyzed using the entire region of *gmfls1* as a

**Fig. 4** Amino acid sequence alignment for FLS cDNAs from *Rosa hybrid* (RhFLS, AB038247), *Malus domestica* (MdFLS, AY965343), *Glycine max* (GmFLS, AB246668), *Petunia hybrida* (PhFLS, Z22543), *Solanum tuberosum* (StFLS, X92178), *Eustoma grandiflorum* (EgFLS, AF240764), and *Arabidopsis thaliana* (AtFLS, U84259). Amino acids common to all the FLS genes are in reverse type



**Fig. 5** HPLC chromatogram of the FLS enzyme assay of *E. coli* cell extracts expressing FLS cDNAs (*gmfls1*) using dihydroquercetin as a substrate. FLS cDNA of Harosoy or Harosoy-*wm* was cloned into pET21a and transformed. *E. coli* cell extracts transformed with empty vector were also subjected to FLS assay as a negative control. The cell extracts expressing *gmfls1* of Harosoy catalyzed the formation of quercetin from dihydroquercetin. In contrast, trace amount of quercetin was detected in extracts of cells transformed with *gmfls1* of Harosoy-*wm* or empty vector, and in the substrate



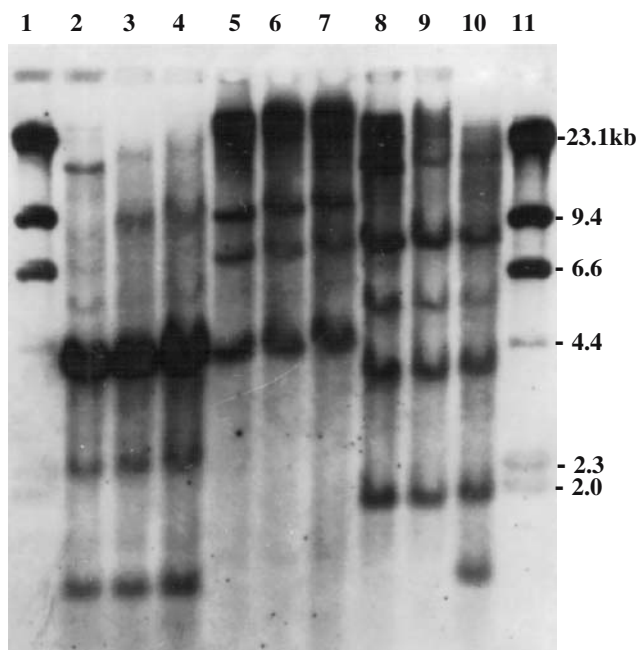
probe to estimate the copy number. Figure 6 shows the hybridization patterns after digestion with *EcoRI*, *HindIII*, and *BglII*. Three to four bands with various hybridizing intensities were observed. Polymorphism was observed between Harosoy and Harosoy-*wm* after *BglII* digestion probably due to the single-base deletion of G in Harosoy-*wm*.

#### CAPS analysis

CAPS analysis was conducted to detect the single-base deletion of G found in Harosoy-*wm* (Fig. 7). The deletion introduced *BglII* site in Harosoy-*wm* resulting in the two fragments (234 and 71 bp) after *BglII* digestion. In contrast, the amplified band of Harosoy was not digested (306 bp) because of the lack of the restriction site (Fig. 3).

#### Co-segregation of deletion mutation with flower color

Flower color of the  $F_2$  population segregated into 90 purple and 24 magenta flower plants. It fitted to a 3:1 ratio ( $\chi^2 = 0.95$ ,  $0.3 < P < 0.5$ ) similar to the report of Buzzell and Buttery (1977). To test whether the deletion mutation co-segregated with flower color, genomic DNA from randomly selected 89  $F_2$  plants was subjected to CAPS analysis. Magenta flower plants had the 234 and 71 bp bands, whereas purple flower plants had fragment of either only the 306 bp band or all of the three bands (Fig. 8 and Table 1). The former and latter were presumed to be homozygotes (*WmWm*) and heterozygotes (*Wmwm*) at the *Wm* locus, respectively.



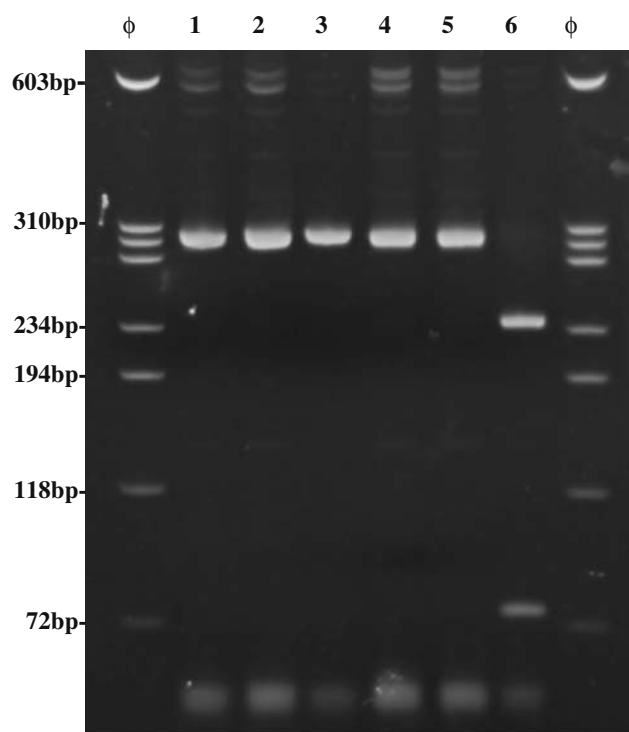
**Fig. 6** Restriction fragment length polymorphisms among Clark, Harosoy and Harosoy-*wm* using *gmfls1* cDNA as probe. Genomic DNA was digested with restriction enzymes, *EcoRI*, *HindIII* and *BglII*, separated by electrophoresis of a 0.8% agarose gel, transferred to nylon membrane, and hybridized with *gmfls1*. Lanes 1 and 11 molecular marker  $\lambda$ /*HindIII*, 2 Clark (*EcoRI* digest), 3 Harosoy (*EcoRI* digest), 4 Harosoy-*wm* (*EcoRI* digest), 5 Clark (*HindIII* digest), 6 Harosoy (*HindIII* digest), 7 Harosoy-*wm* (*HindIII* digest), 8 Clark (*BglII* digest), 9 Harosoy (*BglII* digest), and 10 Harosoy-*wm* (*BglII* digest). The migration of size markers is shown to the right of the blots

### Linkage mapping

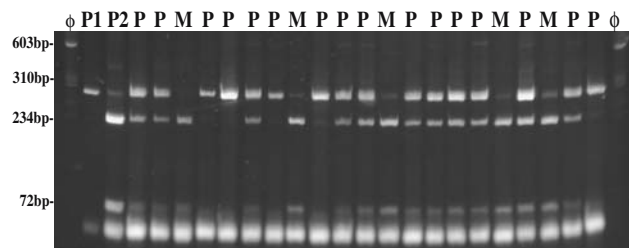
For linkage mapping, SSR markers that are presumed to locate around the *Wm* locus in molecular linkage group F, Satt193, Satt252, Satt193, Satt423, Satt425 and Sat\_039 were used for analysis. Among the SSR markers, Satt269, Satt252, Satt193 and Satt425 exhibited a good separation in the  $F_2$  population. Genotypes of the SSR markers and the CAPS marker were used for linkage analysis. Linkage mapping revealed that the *Wm* locus and *gmfls1* were mapped at similar position in the molecular linkage group F (between Satt252 and Satt425) in accordance with the previous report (Fig. 9) (Cregan et al. 1999).

### Expression of FLS gene

Because RNA blot analysis with labeled *gmfls1* could not detect any signal, semi-quantitative RT-PCR analysis was performed. Figure 10 illustrates the results of RT-PCR analysis using total RNA from trifoliolate leaves (R3 stage) and flower petals of Harosoy and



**Fig. 7** Results of CAPS analysis of GmFLS in Clark, Harosoy and Harosoy-*wm*. PCR products were digested with *BglII* and separated on a 8% polyacrylamide gel.  $\phi$  molecular marker  $\phi$ x174/*HaeIII*, Lanes 1 Clark (undigested), 2 Harosoy (undigested), 3 Harosoy-*wm* (undigested), 4 Clark (digested), 5 Harosoy (digested), and 6 Harosoy-*wm* (digested). The single-base deletion of G in Harosoy-*wm* introduced *BglII* site resulting in the two fragments (234 and 71 bp) after *BglII* digestion. In contrast, amplified bands of Clark or Harosoy were not digested (306 bp) because of the lack of the restriction site. The migration of size markers is shown to the left of the gel



**Fig. 8** Results of CAPS analysis of a  $F_2$  population derived from a cross between Harosoy-*wm* and Clark. PCR products were digested with *BglII* and separated on a 8% polyacrylamide gel.  $\phi$  molecular marker  $\phi$ x174/*HaeIII*, P1 Clark, P2 Harosoy-*wm*, P  $F_2$  plants with purple flower and M  $F_2$  plants with magenta flower. The CAPS markers co-segregated with the genotypes at *Wm* locus;  $F_2$  plants with magenta flower had 234 and 71 bp bands, whereas  $F_2$  plants with purple flower had fragment of either only 306 bp band or all of the three bands. The former and latter were presumed to be homozygotes (*WmWm*) and heterozygotes (*Wmwm*) at the *Wm* locus, respectively. The migration of size markers is shown to the left of the gel



**Table 1** Association between flower color and band patterns in CAPS analysis in the F<sub>2</sub> population derived from a cross between Harosoy-*wm* with magenta flower and Clark with purple flower

Band pattern	Number of plants	
	Purple flower	Magenta flower
Only 306 bp band	22	0
234 and 71 bp bands	0	18
306, 234 and 71 bp bands	49	0

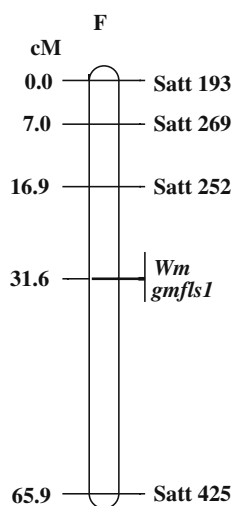
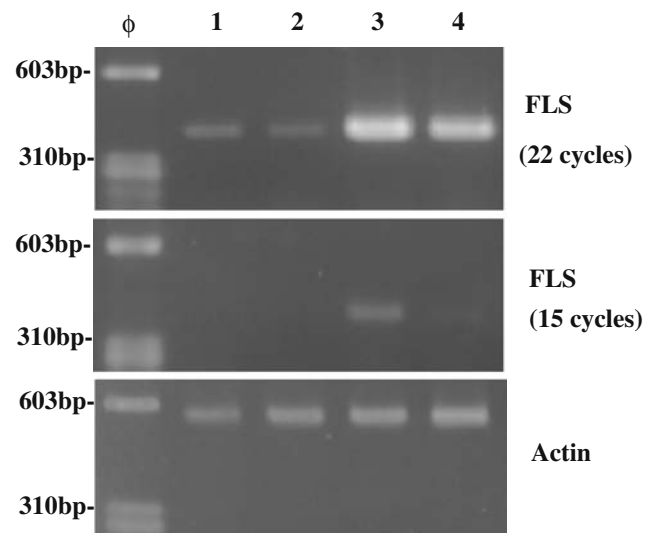
Harosoy-*wm*. Amount of RT-PCR products in flower petals were substantially higher than that in leaves. Further, RT-PCR products of Harosoy-*wm* were lower than Harosoy in both leaves and flower petals.

#### Amount of flavonoids in flower petals

Amount of anthocyanins in banner petals was similar between Harosoy and Harosoy-*wm* in 2003 and 2004 (Table 2). In contrast, amount of flavonols in Harosoy-*wm* was significantly lower (17% in 2003; 9% in 2004) compared to Harosoy.

## Discussion

Flavonol glycosides (kaempferol and quercetin glycosides) are the most abundant flavonoids in soybean leaves (Buttery and Buzzell 1973). We cloned a putative full-length cDNA of soybean FLS gene responsible

**Fig. 9** Linkage mapping of *gmfls1* and *Wm* locus using 89 plants of a F<sub>2</sub> population developed by a cross between Harosoy-*wm* and Clark. The name of linkage group is indicated at the top. Distances of markers from the top of the linkage group are shown on the left. The prefix Satt indicates SSR markers. *gmfls1* and *Wm* locus were mapped at similar position in the molecular linkage group F**Fig. 10** RT-PCR analysis of RNA isolated from leaves and flower petals of Harosoy and Harosoy-*wm*. Specific primers for *gmfls1* of Harosoy (*upper* and *middle* panels) and *actin* control (*lower* panel) generated fragments of 383 and 552 bp, respectively.  $\phi$  molecular marker  $\phi$ x174/*Hae*III, Lanes 1 Harosoy (leaf), 2 Harosoy-*wm* (leaf), 3 Harosoy (flower petal), and 4 Harosoy-*wm* (flower petal). PCR reaction was performed for 22 (*upper* panel) and 15 cycles (*middle* panel) to evaluate FLS gene expression and 30 cycles for actin control. The amount of RT-PCR products using the *gmfls1* primer in leaves was lower than in flower petals. Less PCR product was generated from Harosoy-*wm* than from Harosoy. The migration of size markers is shown to the left of the gel

for conversion from dihydroflavonols to flavonols, *gmfls1*. *gmfls1* is 1,208 bp long and it is presumed to encode a polypeptide, GmFLS1, with 334 amino acids. GmFLS1 has conserved dioxygenase domains A and B and exhibits a significant homology with previously reported FLS. The GmFLS1 of Harosoy had FLS activity as shown by bacterial heterologous expression assay. Genomic Southern analysis revealed that soybean contains three to four copies of FLS gene in the genome. To our knowledge, this is the first report on the isolation of full-length cDNA of FLS from leguminous plants.

The single-base deletion in Harosoy-*wm* probably shortened the polypeptide to only 37 amino acids. The truncated polypeptide lacks the conserved dioxygenase domains A and B. Bacterial expression assay

**Table 2** Contents of anthocyanins and flavonols represented by peak areas in HPLC chromatogram  $\pm$  SE ( $\times 10^3$ ) in flower petals of soybean in 2003 and 2004 at Tsukuba, Japan

Lines	Year	Anthocyanins	Flavonols
Harosoy	2003	5,017 $\pm$ 266	19,101 $\pm$ 1,068
	2004	3,585 $\pm$ 117	16,215 $\pm$ 846
Harosoy- <i>wm</i>	2003	4,378 $\pm$ 55	3,265 $\pm$ 213
	2004	3,738 $\pm$ 153	1,452 $\pm$ 48

confirmed that GmFLS1 of Harosoy-*wm* has no FLS activity.

The CAPS marker that was designed to detect the single-base deletion co-segregated with alleles at *Wm* locus in a F<sub>2</sub> population segregating for the locus. The CAPS marker was mapped at the expected position in the molecular linkage group F. These results strongly suggest that *Wm* encodes FLS and the single-base deletion produced a truncated polypeptide lacking FLS activity, resulting in the substantial reduction of flavonols in flower petals and leaves. Flavonoid analysis revealed that amount of flavonols in flower petals of Harosoy-*wm* was 9–17% compared to Harosoy. It suggests that FLS encoded by other FLS gene member(s) may also have FLS activity.

Semi-quantitative RT-PCR analysis revealed that abundance of transcript in Harosoy-*wm* was lower than Harosoy in both leaves and flower petals. Nonsense-mediated mRNA decay (NMD) is a mRNA surveillance widespread among eukaryotes that detects and eliminates mRNAs harboring premature translation termination codons, thereby preventing the synthesis of truncated and potentially harmful proteins (Baker and Parker 2004; Conti and Izaurralde 2005). NMD may be responsible for the decrease in transcripts in the mutant similar to the F3'H gene in soybean (Toda et al. 2002).

Co-pigmentation of anthocyanins with tannins, flavones and flavonol glycosides can have major effects on flower color (Scott-Moncrieff 1936). When compared over a range of physiological pH values, co-pigmented anthocyanins are usually bluer than anthocyanins alone. The recessive *wm* allele substantially reduces flavonol glycoside content and it may inhibit co-pigmentation between anthocyanins and flavonol resulting in less bluer (magenta) flower color (Fig. 2) similar to the FLS mutant in petunia (Holton et al. 1993). Co-pigmentation between anthocyanins and flavonol glycosides may contribute to the purple flower color in soybeans having *Wm* allele.

Recessive *wm* allele reduces photosynthetic rate and seed yield, and causes earlier leaf senescence (Buzzell and Buttery 1977). Either low levels of flavonol glycosides or high levels of their precursors may be responsible for the deleterious effects. Flavonols are required for pollen tube growth, UV-protection, IAA metabolism or transport, and defense (Stafford 1990). In soybean, additional functions of flavonol glycosides have been reported. Four flavonol glycoside alleles, *Fg1* ( $\beta$ (1-6)-glucoside present), *Fg2* ( $\alpha$ (1-6)-rhamnoside present), *Fg3* ( $\beta$ (1-2)-glucoside present), and *Fg4* ( $\alpha$ (1-2)-rhamnoside present), have been identified in soybean (Buzzell and Buttery 1974). Plants with the *Fg1*

*Fg3* alleles have a lower rate of photosynthesis, lower leaf chlorophyll concentration, lower leaf weight, and lower seed yield (Buttery and Buzzell 1976). Further, *Fg1* and *Fg3* control waviness of leaf margins in soybean (Buzzell and Buttery 1998). However, in the presence of brown pubescence (*TT*), wavy leaf is not expressed (Rode and Bernard 1975). Further investigations are necessary to determine the mechanisms how flavonol glycosides in soybean affect a wide variety of characters such as photosynthesis, seed yield, leaf weight, leaf senescence, leaf chlorophyll concentration and leaf waviness.

**Acknowledgments** We thank Dr. R.L. Nelson at USDA/ARS University of Illinois for providing the seeds of T235. We are grateful to Dr. Joseph G. Dubouzet (Natl. Inst. Agrobiol. Sci.) for critical reading of the manuscript, and Dr. Eduardo R. Benitez (Nat 1. Inst. Crop. Sci.) for taking photographs.

## References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Baker KE, Parker R (2004) Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Curr Opin Cell Biol* 16:293–299
- Buttery BR, Buzzell RI (1973) Varietal differences in leaf flavonoids of soybeans. *Crop Sci* 13:103–106
- Buttery BR, Buzzell RI (1976) Flavonol glycoside genes and photosynthesis in soybean. *Crop Sci* 16:547–550
- Buzzell RI, Buttery BR (1974) Flavonol glycoside genes in soybeans. *Can J Genet Cytol* 16:897–899
- Buzzell RI, Buttery BR (1977) Inheritance and linkage of a magenta flower gene in soybeans. *Can J Genet Cytol* 19:749–751
- Buzzell RI, Buttery BR (1982) Genetics of black pigmentation of soybean seed coats/hila. *Soybean Genet Newsl* 9:26–29
- Buzzell RI, Buttery BR (1998) Genetics of leaf waviness in soybean. *Soybean Genet Newsl* 25:23–24
- Conti E, Izaurralde E (2005) Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr Opin Cell Biol* 17:316–325
- Cregan PB, Jarvic T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic map of the soybean genome. *Crop Sci* 39:1464–1490
- Dixon RA, Steele CL (1999) Flavonoids and isoflavonoids—a gold mine for metabolic engineering. *Trends Plant Sci* 4:394–400
- Fasoula DA, Stephens PA, Nickell CD, Vodkin LO (1995) Cosegregation of purple-throat flower color with dihydroflavonol reductase polymorphism in soybean. *Crop Sci* 35:1028–1031
- Fehr WR, Caviness CE, Burmood DT, Pennington JS (1971) Stage of development descriptions for soybeans, *GLYCINE MAX* (L.) Merrill. *Crop Sci* 11:929–931
- Forkmann G (1991) Flavonoids as flower pigments: the formation of the natural spectrum and its extension by genetic engineering. *Plant Breed* 106:1–26

- Holton TA, Brugliera F, Tanaka Y (1993) Cloning and expression of flavonol synthase from *Petunia hybrida*. *Plant J* 4:1003–1010
- Holton TA, Cornish EC (1995) Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* 7:1071–1083
- Lander ES, Green P, Abrahamson J, Barlow A, Day MJ, Lincoln SE, Newberg L (1987) Mapmaker: an interactive computer package for constructing primary genetic linkage map of experimental and natural populations. *Genomics* 1:174–181
- Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucl Acid Res* 8:4321–4325
- Palmer RG, Pfeiffer TW, Buss GR, Kilen TC (2004) Qualitative genetics. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production, and uses, 3rd edn. Agron Monogr 16. ASA, CSSA, and SSSA, Madison, WI, pp. 137–233
- Rode MW, Bernard RL (1975) Inheritance of wavy leaf. *Soybean Genet Newsl* 2:42–44
- Scott-Moncrieff R (1936) A biochemical survey of some Mendelian factors for flower color. *J Genet* 32:117–170
- Shah DM, Hightower RC, Meagher RB (1983) Genes encoding actin in higher plants: intron positions are highly conserved but the coding sequences are not. *J Mol Appl Genet* 2:111–126
- Shirley BW (1996) Flavonoid biosynthesis: new functions for an old pathway. *Trends Plant Sci* 1:377–382
- Shirley BW (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol* 126:485–493
- Stafford HA, (1990) Flavonoid metabolism. CRC Press, Boca Raton, FL, pp. 1–298
- Toda K, Yang D, Yamanaka N, Watanabe S, Harada K, Takahashi R (2002) A single-base deletion in soybean flavonoid 3'-hydroxylase gene is associated with gray pubescence color. *Plant Mol Biol* 50:187–196
- Todd JJ, Vodkin LO (1996) Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. *Plant Cell* 8:687–699
- Zabala G, Vodkin LO (2005) The *wp* mutation of *Glycine max* carries a gene-fragment-rich transposon of the CACTA superfamily. *Plant Cell* 17:2619–2632