Molecular Genetic Studies on Soybean Mutants for Anthocyanin Biosynthesis

A Dissertation Submitted to

the Graduate School of Life and Environmental Sciences,

the University of Tsukuba

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in Agricultural Sciences

(Doctoral Program in Advanced Agriculture Technology and Sciences)

Fan YAN

Molecular Genetic Studies on Soybean Mutants for Anthocyanin Biosynthesis

May 2015

Fan YAN

Table of Contents

CHAPTER 1 Introduction

Soybean (*Glycine max* (L.) Merr.) is one of the earliest commercially grown plant that occupies the global market providing a remarkable amount of protein and oil. Flavonoids comprise of a large family of secondary metabolites including chalcones, flavanones, dihydroflavonols, flavonols, anthocyanins and proanthocyanidins in soybean (Haslam, 1998). The most important physiological effects of flavonoids are protecting plants from UV damage, against pathogen, attracting of pollinators and seed disperser (Winkel-Shirley, 2001). The consumption of soy-based food has increased worldwide these years, because it is considered to be an abundant source of isoflavonoid (Winkel-Shirley, 2001; Phommalth *et al*., 2008). Isoflavonoids are one of the specific form of flavonoids found in soybean and is considered to be beneficial to prevent human cancer (Dellapenna, 1999).

Flavonoids also determined pigmentation of flowers, seeds and pubescence in soybean. Maize, petunia, *Antirrhinum* and Arabidopsis as the model plants since 1900s, have attained remarkable results on flavonoid synthesis system affecting pigmentation of flower or seed coat (Holton and Cornish, 1995; Mol *et al*., 1998; Lepiniec *et al*., 2006). Although soybean mutants with flower color and seed coat color variation have been found, the genetic and molecular basis were poorly understood because contribution to yield and stress tolerance in unknown.

Generally, soybean cultivars have purple or white flowers. However, some other types were found having purple-blue (*w2*), magenta (*wm*), pink (*wp*), pale (*w1*

or *w4*) and dilute (*w1* or *w4*) flower colors (Takahashi *et al*., 2007; Takahashi *et al*., 2008; Takahashi *et al*., 2010; Xu *et al*., 2010) (Fig. 1.1). Previous studies revealed that *W4* gene was mutated in flower color variants T322, T321 and T369 (Hartwig and Hinson, 1962; Xu *et al*., 2010). A *Glycine soja* accession B09121 with light purple flower was controlled by a new allele of *W1* locus (Takahashi *et al*., 2010). The flower color variation was associated with different expressing level or mutation of anthocyanin genes in biosynthetic pathway. Schematic diagram of the anthocyanin and flavonol biosynthetic pathways is presented in Fig. 1.2.

Five genes have been identified contributing to seed coat color (*I, R, T, O* and *W1*) (Palmer *et al*., 2004). The *R* gene was assigned to classical linkage group II (molecular linkage group K, chromosome 9) (Nagai and Saito, 1923; Weiss, 1970). The *R* locus consists of multiple alleles of *R*, *r* and *r-m* (black stripes on brown seed) (Bernard and Singh, 1969). The *r-m* allele produces seeds that are highly variable not only among seeds in individual plants but also across generations. So, a transposable element may exist in *r-m* allele.

Transposable elements (TEs) were first discovered in maize (McClintock, 1948). Later, TEs have been found to existe in all organisms. TEs are DNA fragments that can change their positions in genome. They are useful tools for gene cloning because insertion of a transposon cause disruption of gene function, often producing variegation phenotypes when inserted in genes responsible for pigmentation (Toda *et al.*, 2002; Zabala and Vodkin, 2005; Zabala and Vodkin, 2008; Xu *et al.*, 2010; Takahashi *et al.*, 2012). It was a powerful method to use known transposable elements for cloning genes and investigating its function.

Moreover, transposon tagging is a valuable method to clone numerous genes, produce null alleles and identify the functional roles of genes.

Based on the beneficial effects of flavonoids, there is an increasing interest on breeding or cultivating soybean with high levels of flavonoids (Mol *et al*., 1998; Dixon and Steele, 1999; Forkmann and Martens, 2001). Our study was conducted to investigate the genetic, molecular and chemical basis of anthocyanin mutants to obtain information contributing to breeding science in soybean.

First, we examined genetic and molecular mechanism of flower color variants in *G. max* (Clark-*w4*, T321, T369, 222-A-3, E30-D-1 and E023-H-12) and in *G. soja* (kw4, PI 424008A and PI 424008C). Second, we identified a possible causal genetic factor responsible for seed coat variegation. Finally, we analyzed the components of anthocyanins, favonol glycosides, dihydroflavonols and isoflavones in the sprout of mutants and near-isogenic lines for pigmentation expecting to find some unique components.

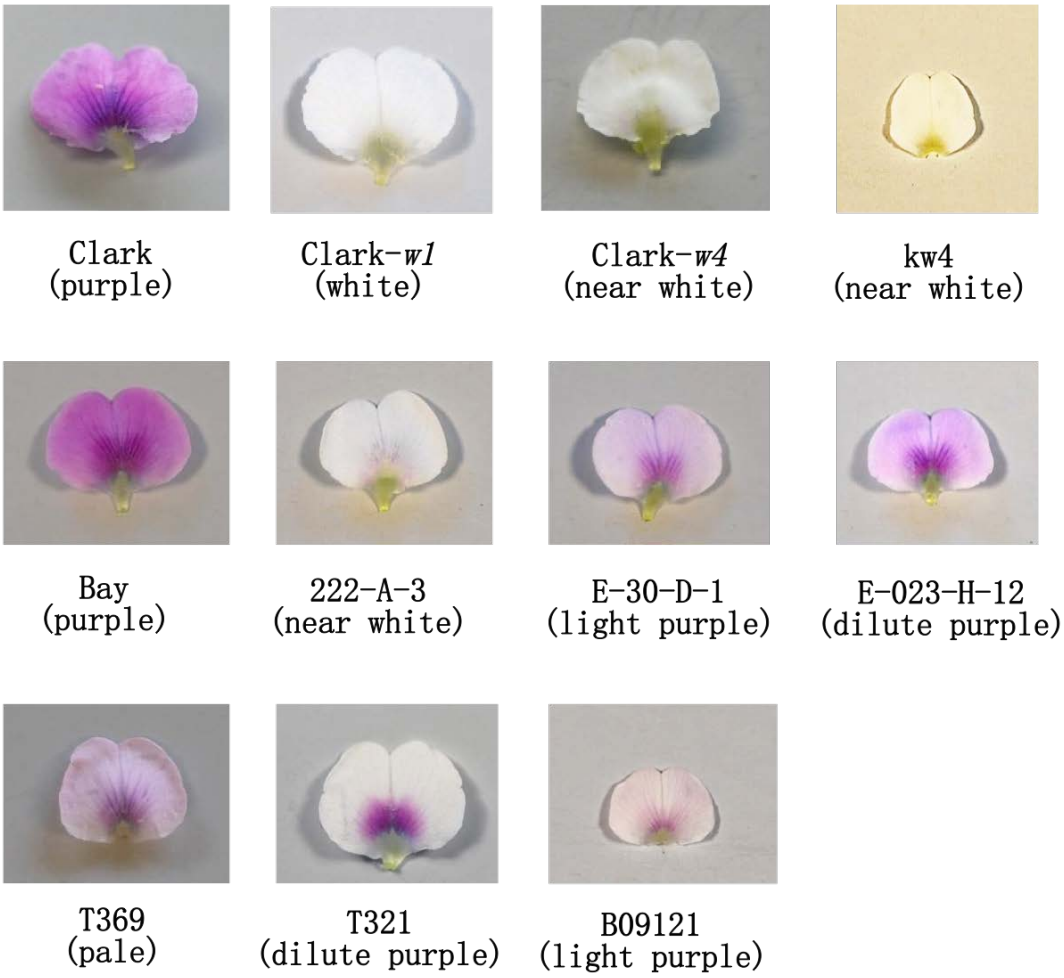

Figure 1.1

Banner petals of flower color variants of soybean and *Glycine soja*

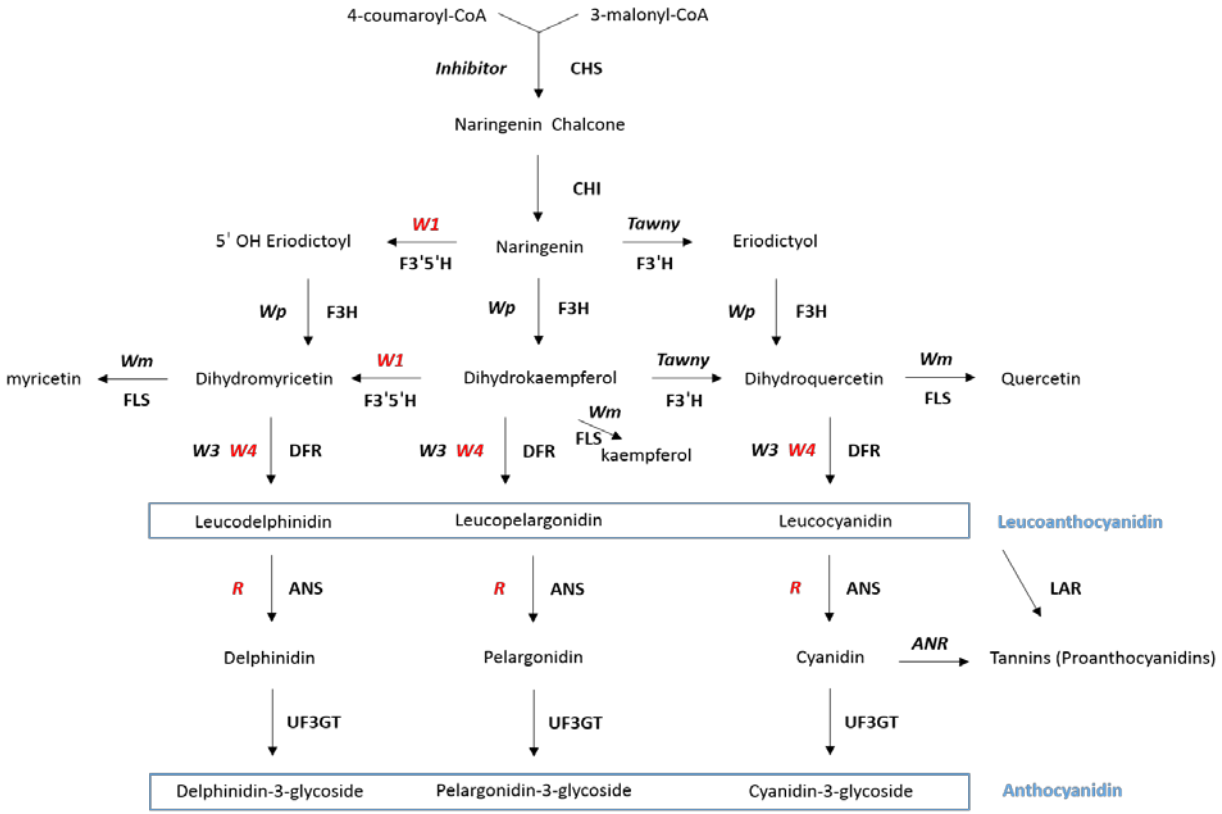

Figure 1.2

Schematic diagram of the anthocyanin and flavonol biosynthetic pathways

Enzyme names are abbreviated as follow: chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), dihydroflavonol 4-reductase (DFR), flavonoid 3′-hydroxylase (F3′H), flavonoid 3′5′-hydroxylase (F3′5′H), flavonol synthase (FLS), anthocyanidin synthase (ANS), anthocyanidin reductase (ANR), leucoanthocyanidin reductase (LAR), UDP-flavonoid 3-*O*-glucosyltransferase (UF3GT). Soybean pigmentation genes were in red font, which we investigated.

CHAPTER 2 Genetic analysis of flower color variation

Background

Flower color of soybean (*Glycine max* (L.) Merr.) is primarily controlled by six genes (*W1*, *W2*, *W3*, *W4*, *Wm* and *Wp*) (Palmer *et al*., 2004; Takahashi *et al*., 2008). Dominant and recessive allele of the *W1* locus controls purple and white flower color, respectively (Palmer *et al*., 2004). Under *W1* genotype, soybean genotype with *W3W4* has dark purple, *W3w4* has dilute purple or purple throat, *w3W4* has purple, and *w3w4* has near white flowers (Hartwig and Hinson, 1962). Flower color of genotypes with allelic combination *W1w3w4* was indistinguishable from those with white flowers under many environments, suggesting that environments affect flower color under the allelic combination (Hartwig and Hinson, 1962). *W1* encodes flavonoid 3'5'-hydroxylase (F3'5'H), *W2* encodes MYB transcription factor, *W3* and *W4* encode dihydroflavonol 4-reductase (DFR), *Wm* encodes flavonol synthase (FLS) and *Wp* encodes flavanone 3-hydroxylase (F3H) (Fasoula *et al*., 1995; Zabala and Vodkin, 2005; Zabala and Vodkin, 2007; Takahashi *et al*., 2007; Xu *et al*., 2010; Takahashi *et al*., 2011; Takahashi *et al*., 2013). The roles of these genes in the biosynthesis of anthocyanin and flavonol pathway are presented in Fig. 1.2.

The components of flavonoids in flower petals of soybean were analyzed (Iwashina *et al*., 2007; Iwashina *et al*., 2008; Takahashi *et al*., 2010). The primary components of anthocyanin were malvidin 3,5-di-*O*-glucoside, petunidin 3,5-di-*O*-glucoside, delphinidin 3,5-di-*O*-glucoside and delphinidin 3-*O*-glucoside. In addition, eight flavonol glycosides, kaempferol 3-*O*-gentiobioside, kaempferol 3-*O*-rutinoside,

kaempferol 3-*O*-glucoside, kaempferol 3-*O*-glycoside, kaempferol 3-*O*-rhamnosyl-(1→2)-[glucosyl-(1→6)-galactoside], kaempferol 7-*O*-glucoside, kaempferol 7-*O*-diglucoside and quercetin 3-*O*-gentiobioside, and one dihydroflavonol, aromadendrin 3-*O*-glucoside were identified. No anthocyanins were detected in Clark-*w1*, a near-isogenic line (NIL) of US cultivar Clark at the *W1* locus. Anthocyanins were not detected in Clark-*w4* in 2003 and 2004, but trace amounts were detected in 2007 (Iwashina *et al*., 2007; Iwashina *et al*., 2008), indicating slight responsiveness to environmental conditions in agreement with the previous report (Hartwig and Hinson, 1962). Similarly, white flower petals of PI 424008C contained no anthocyanins, whereas they contained similar amount of flavonols and dihydroflavonol with PI 424008A (Takahashi *et al*., 2010).

A mutable allele of the *W4* locus was discovered in a cross between two experimental lines with white and purple flowers, respectively (Palmer *et al*., 1989). The mutant line was designated as T322, and the mutable allele was designated as *w4-m*. Mutant lines T321 with *w4-dp* allele (dilute purple flower) and T369 with *w4-p* allele (pale flower) were isolated from descendants of T322 (Palmer and Groose, 1993; Xu and Palmer, 2005) (Fig. 1.1). A 20.5-kb transposable element (*Tgm9*) was isolated from the second intron of the *DFR2* gene (Xu and Palmer, 2005; Xu *et al*., 2010). In T321, *Tgm9* was excised and left 4 bp footprints. No footprint was left in T369 (Xu and Palmer, 2005). A 5' end fragment of *Tgm9* (944bp) was integrated at different upstream positions of the start codon in T321 and T369. Soybean has two other *DFR* genes, *DFR1* and *DFR3* (Yang *et al*., 2010). DNA marker analysis suggested that *W3* locus might correspond to the *DFR1* (Yang *et al*., 2010).

In soybean, about 33 % accessions have white flowers. In contrast, flower color of *G. soja* is almost exclusively purple. Allele of the *W1* locus is associated with hydroxylation of the 3'- and 5'-positions of the B-ring and is presumed to encode flavonoid 3'5'-hydroxylase (F3'5'H) (Buzzell *et al*., 1987). White flower color is associated with 65 bp insertion of tandem repeats in the coding region of the *F3'5'H* gene (Zabala and Vodkin, 2007). There are two types of *F3'5'H* gene. *GmF3'5'H-a* and *GmF3'5'H-b* have three SNPs different from each other (position 1059, 1424 and 1509) (Guo and Qiu, 2013).

One white-flowered plant was found in 1998 among the progeny of a purple-flowered *G. soja* accession PI 424008A that was introduced from South Korea (Chen and Nelson, 2004). The white-flowered plant was designated as PI 424008C. According to the DNA marker analysis and complementation tests, Chen and Nelson (2004) presumed that PI 424008C was caused by a mutation at *W1* locus. However, molecular mechanism of the white-flowered plant PI424008C has not been identified.

In 2002, a *G. soja* accession with light purple flowers, B09121 was discovered in southern Japan (Takahashi *et al*., 2010) (Fig. 1.1). Genetic and molecular analysis suggested that light purple color was controlled by a new allele at *W1* locus (*w1-lp*). B09121 had a unique amino acid substitution at amino acid position 210 of *F3'5'H* gene. Takahashi *et al*. (2012) found another mutant line having variegated flowers (purple/white) among *G. soja* accessions. The genetic and molecular analysis suggested that the mutable allele (*w1-m*) was caused by insertion of an active transposable element *Tgs1* in the first exon of the *F3'5'H* gene. The

mutable line was designated as B00146-m. The progeny of B00146-m have white flowers and purple flowers, designated as B00146-w and B00146-r, respectively. *Tgs1* left behind 2 bp footprint in white flower plant B00146-w at the time of transposition (Takahashi *et al*., 2012). In summary, four alleles with different flower colors were identified responsible to *W1* gene (*W1*, *w1*, *w1-lp* and *w1-m*).

Clark-*w4* and Clark-*w1* are near-isogenic lines (NILs) of US cultivar Clark at the *W1* and *W4* locus (Bernard *et al*., 1991). Flower color variants 222-A-3, E30-D-1 and E023-H-12 were obtained from mutagenized populations of US cultivar Bay (Anai, 2012) (Fig. 1.1). *G. soja* kw4 was an accession with near white flowers introduced from South Korea (Fig. 1.1).

Dr. Iwashina at Tsukuba Botanical Garden analyzed the flavonoids in flower petals of T369, T321, E30-D-1, 222-A-3 E023-H-12 and kw4. Four anthocyanin components, A1: malvidin 3,5-di-*O*-glucoside, A2: petunidin 3,5-di-*O*-glucoside, A3: delphinidin 3,5-di-*O*-glucoside, A4: delphinidin 3-*O*-glucoside were detected in agreement with previous studies (Iwashina *et al*., 2007; Iwashina *et al*., 2008) (Table 2.1). Flowers of T369 contained 59.8% of total anthocyanins compared with Clark. Less anthocyanins were detected in E023-H-12 (58.4%), T321 (44.7%) and E30-D-1 (39.3%). Near white flowers of 222-A-3 had the lowest level of anthocyanins (15.6%). Near white flowers of kw4 had only trace amount of the two components, A1 and A2. All cultivars and lines except for 222-A-3 and kw4 had all four components with the amounts decreasing in the following order: A1 > A2 > A3 > A4.

All cultivars and lines had eight flavonol glycoside components, F1 (kaempferol 3-*O*-gentiobioside), F2 (kaempferol 3-*O*-rutinoside), F3 (kaempferol 3-

O-glucoside), F4 (kaempferol 3-*O*-glycoside), F5 (kaempferol 3-*O*-rhamnosyl-(1→2)-[glucosyl-(1→6)-galactoside]), F6 (quercetin 3-*O*-gentiobioside), F7 (kaempferol 7-*O*-glucoside), F8 (kaempferol 7-*O*-diglucoside) in accordance with previous studies (Iwashina *et al*., 2007; Iwashina *et al*., 2008; Takahashi *et al*., 2010) (Table 2.2). The total amounts of flavonol glycosides were not very different among cultivars and lines except for T369. F1 was most abundant and accounted for about 80% of flavonol glycosides in these cultivars and lines in accordance with previous studies (Iwashina *et al*., 2007; Iwashina *et al*., 2008). The amount of F2 was extremely low in kw4 and comprised only 0.1% of the total amount of flavonol glycosides. Flowers of T369 had substantially lower amount of flavonol glycosides (16.0% of Clark). Petals of E023-H-12 contained higher amounts of F1 to F5, F8 and less amounts of F6 and F7 (Table 2.2). In total, contents of flavonol glycosides deposited in banner petals of E023-H-12 was 34 % higher than Bay.

Only one kind of dihydroflavonol (D1, aromadendrin 3-*O*-glucoside) was detected in all cultivars and lines except for T369 (Table 2.3). The amount varied from 57.4% (E30-D-1) to 163.8% (kw4) compared with Clark. In contrast, flower petals of T369 contained only 11.4% of D1 compared with Clark, in addition to two unique peaks corresponding to dihydroflavonols, D2 and D3 (Fig. 2.1).

This study was conducted to investigate the genetic and molecular mechanism of flower color variants in *G. max* (Clark-*w4*, T321, T369, 222-A-3, E30-D-1 and E023-H-12) and in *G. soja* (kw4, PI424008A and PI 424008C).

Table 2.1 Anthocyanin content [mean ± SD (x 10^3)] according to HPLC analysis of flower petals from soybean and *Glycine soja* (Dr. Iwashina at Tsukuba Botanical Garden)

Line name	A1[a]	A2	A3	A4	Total
Clark	933±40	538±21	399±13	255±20	2,125±77
Bay	1,745±326	610±99	323±62	267±14	2,945±331
222-A-3	142±28	189±30	0±0	0±0	331±16
E30-D-1	345±4	241±20	151±3	98±65	835±65
kw4	t[b]	t	0±0	0±0	-
T321	369±107	255±22	178±36	148±4	950±155
T369	513±32	348±17	232±44	178±8	1,271±89
E023-H-12	578±230	290±60	188±17	185±42	1,240±334

[a]A1: malvidin 3,5-di-*O*-glucoside, A2: petunidin 3,5-di-*O*-glucoside, A3: delphinidin 3,5-di-*O*-glucoside, A4: delphinidin 3-*O*-glucoside, [b]Trace amount

Table 2.2 Flavonol glycoside content [mean ± SD (x 10^3)] according to HPLC analysis of flower petals from soybean and *Glycine soja* (Dr. Iwashina at Tsukuba Botanical Garden)

Line name	F1[a]	F2	F3	F4	F5	F6	F7	F8	Total
Clark	9,432±103	772±34	177±6	441±26	353±38	138±10	13±0	128±13	11,454±177
Bay	8,508±278	788±52	162±5	246±10	429±26	131±8	53±0	179±15	10,496±385
222-A-3	7,836±426	698±39	134±9	275±33	459±42	124±2	323±55	365±15	10,214±477
E30-D-1	8,001±491	732±37	168±78	335±24	465±20	117±32	274±43	318±56	10,409±630
kw4	10,947±386	16±2	432±5	695±26	802±50	154±7	354±52	417±38	13,816±533
T321	9,417±476	805±50	159±17	371±20	523±16	100±2	174±10	287±12	11,838±598
T369	703±9	214±26	102±2	135±1	243±5	151±2	130±6	158±16	1,837±55
E023-H-12	11,484±286	1,083±12	193±14	387±23	656±80	79±7	14±1	213±23	14,108±400

[a]F1 (kaempferol 3-*O*-gentiobioside), F2 (kaempferol 3-*O*-rutinoside), F3 (kaempferol 3-*O*-glucoside), F4 (kaempferol 3-*O*-glycoside), F5 (kaempferol 3-*O*-rhamnosyl-(1→2)-[glucosyl-(1→6)-galactoside]), F6 (quercetin 3-*O*-gentiobioside), F7 (kaempferol 7-*O*-glucoside), F8 (kaempferol 7-*O*-diglucoside)

Table 2.3 Dihydroflavonol content [mean ± SD (x 10^3)] according to HPLC analysis of flower petals from soybean and *Glycine soja* (Dr. Iwashina at Tsukuba Botanical Garden)

Line name	D1[a]	D2[b]	D3[b]	Total
Clark	843±53	0±0	0±0	843±53
Bay	758±83	0±0	0±0	758±83
222-A-3	593±40	0±0	0±0	593±40
E30-D-1	484±24	0±0	0±0	484±24
kw4	1,381±58	0±0	0±0	1,381±58
T321	646±22	0±0	0±0	646±22
T369	96±8	153±39	54±10	303±54
E023-H-12	773±35	0±0	0±0	773±35

[a]D1: aromadendrin 3-*O*-glucoside

[b]D2, D3: unidentified dihydroflavonols

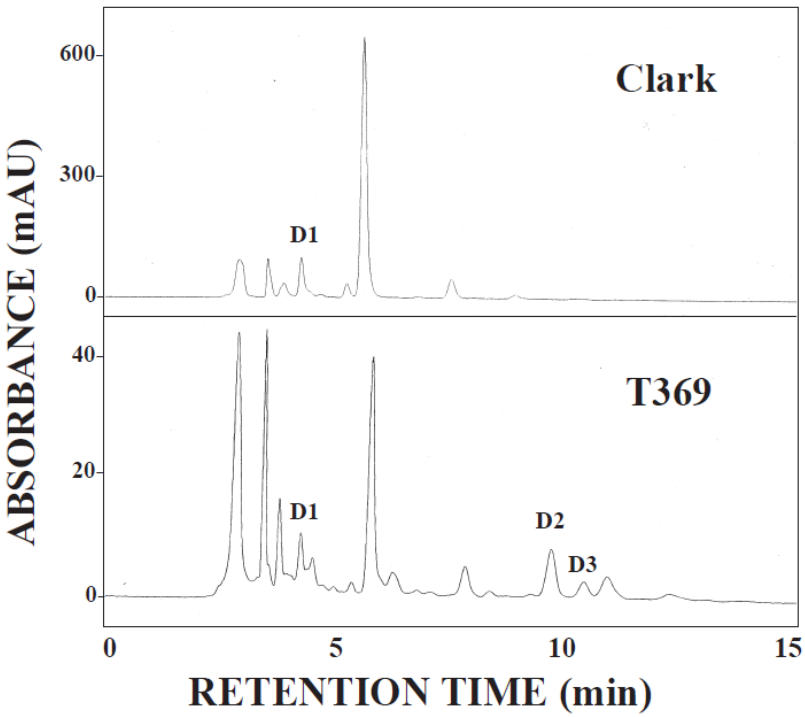

Figure 2.1

HPLC chromatogram of dihydroflavonols extracted from flower petals of a soybean cultivar Clark and T369 (Dr. Iwashina at Tsukuba Botanical Garden)

A total of 200 mg of banner petals was extracted with 2 ml of MeOH. Eluents: MeCN/H_2O/H_3BO_3 (22:78:0.2). Flow-rate: 1.0 ml/min. Injection: 10 μl. Detection: 290 nm. D1, aromadendrin 3-*O*-glucoside; D2 and D3, unidentified dihydroflavonols.

2.1 Allelic variation of soybean flower color gene *W4* encoding dihydroflavonol 4-reductase 2

2.1.1 Materials and methods

Plant materials

The plant materials used in this study are presented in Fig. 1.1 and listed in Table 2.4. Cultivar Clark has purple flower and tawny pubescence (*T*). Bay has purple flower and gray pubescence (*t*). A NIL of a US cultivar Clark with *w4* allele, Clark-*w4* (L68-1774) has near white flower and tawny pubescence. A NIL of a Canadian cultivar Harosoy with *w4* allele, Harosoy-*w4* (L72-1138) has near white flower and gray pubescence. Seeds of NILs, T321 and T369 were provided by the USDA Soybean Germplasm Collection. The NILs were developed by backcrossing the near white flower trait six times from the cultivar Laredo into Clark or Harosoy (Bernard *et al*., 1991). Seeds of 222-A-3 and E30-D-1 were obtained from Dr. Anai at Saga University. Flower color variants 222-A-3 and E30-D-1 were found from mutagenized populations of US cultivar Bay. 222-A-3 with near white flowers was from an X-ray treated population, whereas E30-D-1 with light purple and was developed from an EMS-treated population (Anai, 2012). Seeds of kw4 were obtained from Dr. Xu (JIRCAS, Japan). kw4 was a *G. soja* accession with near white flowers, among accessions introduced from South Korea (personal communication, 2007). It is unknown if the accession has near white flowers in the natural habitat.

Genetic analysis

222-A-3 and E30-D-1 were crossed with Clark-*w4* and E30-D-1 was also

crossed with Clark. Flowers of 222-A-3 and E30-D-1 were emasculated one day before opening and fertilized with pollen from Clark or Clark-*w4*. Harosoy-*w4* was crossed with kw4. Hybridity of the F_1 plants was ascertained by tawny pubescence color.

A total of seven F_1 and 130 F_2 seeds derived from Harosoy-*w4* × kw4 were field-planted at the National Institute of Crop Science, Tsukuba, Japan (36°06'N, 140°05'E). Similar numbers of F_1 and F_2 seeds derived from two crosses (222-A-3 × Clark-*w4* and E30-D-1×Clark-*w4*) were planted. A bulk of 30 seeds each of fifty F_3 families derived from E30-D-1 × Clark-*w4* were planted. A total of six F_1 and 130 F_2 seeds derived from E30-D-1 × Clark were planted. N, P and K were applied at 3.0, 4.4 and 8.3 g m^{-2}, respectively. Plants were individually grown with spacing of 70 cm between rows and 10 cm between plants. Flower color was recorded in individual F_1, F_2 and F_3 plants.

Molecular cloning

Total RNA of seven cultivars was extracted from banner petals (200 mg) using the TRIZOL Reagent (Invitrogen) according to the manufacturer's protocol. Then, RT-PCR was conducted to obtain the cDNA by reverse transcription using the Superscript III First-Strand Synthesis System (Invitrogen) and an oligo(dT) primer.

Genomic DNA of Clark, kw4, Bay and E-30-D-1 was extracted from trifoliolate leaves by CTAB method (Murray and Thompson, 1980). Genome sequences containing the entire coding region (about 3.3 kb) and the 5' upstream region (about 1.2 kb) of Clark and kw4 were determined by cloning two fragments

overlapping each other. The 5' upstream region was also cloned from Bay and E30-D-1. PCR primers were listed in Table 2.5.

The PCR mixture contained 10 ng of genomic DNA, 10 pmol of each primer, 10 pmol of nucleotides and 1 unit of ExTaq in 1×ExTaq Buffer in a total volume of 50 µl. A 30 sec denaturation at 94°C was followed by 30 cycles of 30 sec denaturation at 94°C, 1 min annealing at 59°C and 1 min extension at 72°C. A final 7 min extension at 72°C completed the program. The PCR was performed in an Applied Biosystems 9700 thermal cycler. PCR products were separated on a 0.8% agarose gel. After electrophoresis, the gel was stained with ethidium bromide and the DNA fragments were visualized under UV light.

Ligation

PCR fragments with expected size were ligated with pCR® 2.1-vector (Invitrogen). By mixing the following components, the mixture was mixed and kept at 16°C overnight.

pCR®2.1-vector	1.0 µl
PCR product	2.0 µl
Ligation Mix	3.0 µl

Transformation

ECOS™ Competent *E.coli DH5α* were used for transformation. After transformation, 10 colonies were picked and cultured in LB medium containing ampicillin to select positive clones. Transformation method followed the protocol

provided by the manufacturer as below.

a. Thaw the competent cell on ice

b. Add ligation mixture into the cell immediately

c. Vortex for 1 sec immediately

d. Incubation on ice for 5 minutes

e. Incubation for 45 sec at 42°C

f. Vortex for 1sec immediately

g. Transfer the cell to the LB plate with antibiotic ampicillin

h. Incubation for 12-16 hours at 37°C

Sequencing and analysis

Four recombinant plasmids were selected for sequencing analysis. The reaction mixture contained 50 ng of plasmid, 4.0 µl terminator ready reaction mix, 3.2 pmol of primer, deionized water and adjusted to 10 µl. An initial denaturation at 96°C for 5 min was followed by 25 cycles of 96°C denaturing for 30 sec, 50°C annealing for 5 sec and 60°C extension for 4 min. Then, denatured at 95°C for 5 min. Samples were sequenced using the ABI 3100 Genetic Analyzer. Nucleotide sequences of both strands were determined. Primers are exhibited in Table 2.5. Nucleotide sequences and the postulated amino acid were analyzed with GENETYX ver. 8.1.2 (GENETYX). Sequences were aligned using ClustalW (http://clustalw.ddbj.nig.ac.jp/top-j.html) at default settings.

CAPS analysis

Genomic DNA of Clark-*w4*, E30-D-1 and 40 F_2 plants that were used for F_3 progeny tests were isolated from trifoliolate leaves by CTAB method. A pair of PCR primers was designed to detect a single-base substitution found in E30-D-1 (Table 2.5). The base substitution within the restriction site is expected to result in the presence/absence of the restriction site of *Bsr*GI in the amplified product. Annealing temperature was at 56°C. Then, PCR products were digested with *Bsr*GI, and the digests were separated on an 8 % non-denaturing polyacrylamide gel in 1 x TBE buffer (90 mM Tris-borate, 2 mM EDTA, pH 8.0). After electrophoresis, the gel was stained with EtBr and the DNA fragments were visualized under UV light.

Quantitative real-time PCR

For quantitative real-time PCR, total RNA (5 μg) was extracted from banner petal samples (three replications) and was reverse-transcribed. The quantitative real-time PCR system contained 2.0 μl 1x SYBR Premix Dimer Eraser, 2.0 μl 1 x ROX reference dye, 0.6 μl of each primer (10 μM), 0.4 μl of cDNA synthesis reaction mixture, adjust to a final volume of 20 μl. The initial 30 sec denaturation at 95°C, 40 cycles of 3 sec denaturation at 95°C, 30 sec annealing at 58°C and 30 sec extension at 72°C. Using the Step-One-Plus Real-Time PCR System (Applied Biosystems) for analysis. Primer sequences are exhibited in Table 2.5. The expression level of the soybean *actin* gene (GenBank accession number: J01298) (Shah *et al*., 1983) was used to normalize target gene expression.

Accession numbers

Sequence data was deposited in the DDBJ Data Libraries under accession numbers AB872212 (cDNA of Bay), AB872213 (cDNA of Clark-*w4*), AB872214 (cDNA of 222-A-3), AB872215 (cDNA of E30-D-1), AB872216 (genomic DNA of Clark) and AB872217 (genomic DNA of kw4).

2.1.2 Results

Genetic analysis

F_1 plants derived from a cross between Harosoy-*w4* and kw4 had near white flowers (Table 2.6). All of the 116 plants of the F_2 population had near white flowers, suggesting that flower color of kw4 was controlled by the *w4* allele. F_1 plants derived from a cross between 222-A-3 and Clark-*w4* had near white flowers. All of the 109 plants of the F2 population had near white flowers, suggesting that flower color of 222-A-3 was also controlled by the *w4* allele.

F_1 plants derived from a cross between E30-D-1 and Clark had purple flowers. A total of 112 plants of the F_2 population segregated into 84 plants with purple flowers and 28 plants with light purple flowers. The segregation fitted a 3:1 ratio ($\chi^2 = 0.00$, P = 1.00) suggesting that a single gene controls flower color and that the allele for purple flower was dominant to light purple flower. F_1 plants derived from a cross between E30-D-1 and Clark-*w4* had light purple flowers. A total of 111 plants of the F_2 population segregated into 82 plants with light purple flowers and 29 plants with near white flowers. The segregation fitted a 3:1 ratio ($\chi^2 = 0.08$, P =

0.78) suggesting that the *W4* locus controls the flower color and the allele for light purple flower was dominant to near white flower. Ten F_3 lines derived from F_2 plants with near white flowers had near white flowers. Forty families derived from F_2 plants with light purple flowers segregated into 16 families fixed with light purple flowers and 24 families segregating for flower colors (Table 2.7). The segregation fitted a 1:2 ratio (χ^2 = 0.80, P = 0.37) confirming that an allele at the *W4* locus controls flower color. The new allele was designated as *w4-lp*. The dominance relationship of the alleles is *W4* > *w4-lp* > *w4*.

Molecular cloning

DNA fragments of about 1.1 kb were amplified by RT-PCR in Clark, Bay, 222-A-3 and E30-D-1 (Fig. 2.2). Fragments of about 1.4 kb were amplified from Clark-*w4*. No amplification product was observed in kw4. The *DFR2* transcripts of Clark and Bay were 1065 bp long and they encoded 354 amino acids. Amino acids were identified except for two substitutions around the C-terminus at positions 338 (valine or glutamic acid) and 353 (arginine or glutamine). Comparison of nucleotide sequences between cDNA and genomic DNA of Clark revealed that the *DFR2* gene has six exons and five introns similar to a previous report (Xu *et al*., 2010).

The sequencing results showed that 222-A-3 had a base (T) deletion at nucleotide position 29. This deletion probably generated a truncated polypeptide consisting of only 24 amino acids (Fig. 2.3 A and B). The polypeptide lacked the NADPH binding domain (Lacombe *et al*., 1997). In E30-D-1, a single base was substituted from G to A at nucleotide position 116 compared with Bay (Fig. 2.3 C).

The base-substitution altered amino acid from arginine to histidine at position 39. The 5' upstream region of E30-D-1 was identical with that of Bay and Clark. In Clark-*w4*, cDNA had a 344-bp insertion compared with Clark and Bay. The insertion corresponded to the fourth intron with five nucleotide substitutions compared with Clark, suggesting that the fourth intron was retained in Clark-*w4*. In Clark-*w4*, a single-base G at the start of the fourth intron was changed to A compared with the genome sequence of Clark (Fig. 2.3 D). The base substitution may have abolished the 5' splice site (GT) resulting in the retention of the fourth intron. The retention caused a mutation from amino acid position 217 and premature translation termination at amino acid position 227 (Fig. 2.4).

In kw4, transcripts of the *DFR2* gene in the flower petals were not detected by RT-PCR. The genomic fragments containing the entire coding region were amplified by PCR. Six exons and five introns were assumed similar to Clark. The amino acid sequence was identical with that of Clark. A 367-bp fragment was deleted in the third intron of kw4 (Fig. 2.4). The 5' upstream region of kw4 had six single-base substitutions, three single base deletion, two two-base deletions and a three-base alteration including one base deletion (Fig. 2.5).

CAPS analysis

PCR with CAPS primers generated a band of 377 bp in Bay, Clark, Clark-*w4*, 222-A-3 and E30-D-1. Digestion with *Bsr*GI generated a shorter band of 194 bp in E30-D-1, whereas the band was undigested in the other materials (Fig. 2.6 upper panel). The result showed that plants fixed with light purple flowers had only a

shorter band, plants fixed with near white flowers had only a longer band and plants segregating for flower colors had both bands (Fig. 2.6 lower panel). Thus, the CAPS marker co-segregated with flower color.

Quantitative real-time PCR

Results of real-time PCR are presented in Fig. 2.7. Transcript level of T321 was low and 16.8 % of Bay. Transcript levels of 222-A-3, E30-D-1 and Clark-*w4* were much lower at 7.7, 3.8 and 3.7 %, respectively. Transcripts of the *DFR2* gene of kw4 were not detected by real-time PCR. In contrast to the above flower color variants, the transcript level of T369 was about 2.3 times of Bay.

2.1.3 Discussion

Previous studies revealed that the *W4* gene was mutated in flower color variants Clark-*w4*, T321 and T369 (Hartwig and Hinson, 1962; Xu *et al*., 2010). In our study, genetic analysis, molecular experiment and complementation tests revealed that the flower colors of 222-A-3, E30-D-1 and kw4 (*G. soja* accession) were also controlled by this gene. Amino acid polymorphism (i.e., truncation) or null expression of the *DFR2* gene was connected to flower color variants.

In 222-A-3, a single-base deletion caused a frame-shift mutation from amino acid position 11. This deletion was postulated to produce a truncated polypeptide of only 24 amino acids that lacked the NADPH binding domain. Thus, the *DFR2* transcript of 222-A-3 may be nonfunctional. In Clark-*w4*, the first nucleotide of the fourth intron was substituted from G to A. The base substitution may have abolished

the 5’ splice site (recognition site). The retention changed the subsequent amino acids. Translation was prematurely terminated of consisting of 227 amino acids. *DFR* genes have many amino acids conserved across plant species in the downstream of the mutation (Shimada *et al*., 2004). The results strongly suggest that the *DFR2* gene of Clark-*w4* may not functional. In kw4, the *DFR2* gene was not expressed in flower petals. Compared to Clark, a 367-bp fragment was deleted in the third intron of *DFR2* gene. However, we are not sure whether the deletion in intron may be responsible for null gene expression. So, we investigated the 5’ upstream region to check if any mutation occurred in the promoter region. The results showed that there were many nucleotide polymorphisms in the 5’ upstream region, including six single base substitutions, three single base deletions, two continuous base deletion and a three-base alteration including one deletion. The upstream promoter regions of structural anthocyanin biosynthesis genes contain cis regulatory elements that affect pigmentation patterns or intensity (Xu *et al*., 2010). The accumulation of a substantial number of mutations might have resulted into mull gene expression. Promoter assays may be necessary to determine which polymorphism is critical for gene expression. Features of DNA sequences in Clark-*w4*, 222-A-3 and kw4 strongly suggest that loss-of-function of *DFR2* gene lead to near white flowers.

The 5’ upstream regions of the *DFR2* gene in E30-D-1, Bay and Clark were identical. In the first exon, however, a single-base substitution altered an amino acid at position 39 from arginine to histidine. The position of the residue was slightly downstream of the NADPH binding region. No catalytic domain has been assigned to the region, but the arginine residue is conserved across eight plant species

(Shimada *et al*., 2004). Further, CAPS marker of E30-D-1 was co-segregated with flower color. These results suggest that the amino acid substitution might have affected transcript abundance and/or DFR function resulting in less anthocyanin contents and light concentration of flower color. Transgenic experiments may be necessary to ascertain the functional of this single nucleotide polymorphism.

Flavonol glycoside content in flower petals of T321 was similar compared with Clark. In contrast, flavonol glycoside content of T369 was only 16.0 % compared with Clark (Table 2.2). In addition, the *DFR2* gene was over-expressed in flower petals of T369 and it was barely expressed in T321. The reduction of flavonol glycosides in T369 can be explained by substrate competition between over-expressed DFR and flavonol synthase (Fig. 1.2). Flower petals of T369 contained substantially lower amounts of D1 but it had unique dihydroflavonol components, D2 and D3. D2 and D3 are absent in the soybean and *G. soja* accessions analyzed so far. Over-expression of *DFR2* gene may be responsible for the unique dihydroflavonol composition. Chemical structure of D2 and D3 should be determined and investigated.

DFR2 gene of soybean controls intensity and distribution of pigmentation in flower petals. Mutation of the gene results in unique flavonoid composition and a wide variety of flower color patterns, from light purple, dilute purple, pale to near white.

Table 2.4 Plant materials of soybean and *Glycine soja* used in this study

Line	Flower color	Genotype	Origin	Cross combination
Clark	purple	*W1W2w3W4WmWpT*	-	-
Bay	purple	*W1W2w3W4WmWpt*	-	-
L68-1774 (Clark-*w4*)	near white	*W1W2w3w4WmWpT*	L6[a](6) x (Laredo x Harosoy)	-
L72-1138 (Harosoy-*w4*)	near white	*W1W2w3w4WmWpt*	L2[b](6) x Laredo	-
222-A-3	near white	-	X-ray induced mutant of Bay	222-A-3 x Clark-*w4*
E30-D-1	light purple	-	EMS-induced mutant of Bay	E30-D-1 x Clark E30-D-1 x Clark-*w4*
kw4	near white	-	*G. soja* accession of South Korea	Harosoy-*w4* x kw4

[a] A *Phytophtora* and pustle-resistant Clark isoline with genes *Rps1* and *rxp*

[b] A *Phytophtora* and pustle-resistant Harosoy isoline with genes *Rps1* and *rxp*

Table 2.5 PCR primers used in this study

Purpose	Target	Forward primer (5'-3')	Reverse primer (5'-3')
cDNA cloning	*DFR2*	AACCAAAACAACGAGAGAGA	CTTATCCCTGATATGAAAGC
cDNA sequencing	*DFR2*	TGCTAGACATCATGAAAGCA CACTGCTCTTTCACTAATCA TACCCTGAGTATAATGTCCT	TGTGAACAGCATATGTACCT GATTAGTGAAAGAGCAGTGA TTCACGCATGCTTTCATGAT
cloning of genomic fragment	upstream fragment of *DFR2*	ACGGTTTCTTCCATTCCATT	ACTTGATTTCAGCCATGGTA
	downstream fragment of *DFR2*	GTTCATCAATGCACATAGAC	CTTATCCCTGATATGAAAGC
sequencing of genomic fragment	upstream fragment of *DFR2*	TACAAGTTGTCATCACGATC TTTGGTGTACACTCGTATGT ATGTAACATGATGGTTCGTG	GAAGCTTTGATGAAGCCATT CACAATTATATCATTGGGCA AACCACCATTGCTTAATACC
	downstream fragment of *DFR2*	CTTTTTCTCTGCAGGTTTCA AAGTACCATTCCAACATTAA TGTTGTGCTCTTTGGCATAT	TAGTGGATGAATATGATTCT GATAGATGACAGTTGTTGTC ACCCTGAGTATAATGTCCTT
CAPS analysis	*DFR2*	ACGGTTTCTTCCATTCCATT	CAAATGCTTCACCTTCTTCA
cloning of 5' upstream region	upstream fragment of *DFR2*	AGAGATATATAAGAAGTTAGGA	TATCACGAAATAGTTTTTGTAAT
	downstream fragment of *DFR2*	CCTTTACCATCTACAAGATAA	ATGATGTAATATTGGGAACCT
sequencing of 5' upstream region	*DFR2*	GAAAAGAGAAATAGGTATTATA	GTTTAACTAATCAAACTAAATT
real-time PCR	*DFR2*	CCAAGGACCCTGAGAATGAA	CAGAAGTCAACATCGCTCCA
	actin	GTCCTTTCAGGAGGTACAACC	CCACATCTGCTGGAAGGTGC

Table 2.6 Segregation of flower color in F_1 plants and F_2 populations

Generation	Number of plants				Expected ratio	χ^2 value	Probability (*P* value)
	Total	Purple	Light purple	Near white			
kw4	10	-	-	10	-	-	-
Harosoy-*w4* (H-*w4*)	10	-	-	10	-	-	-
H-*w4* x kw4 F_1	5	-	-	5	-	-	-
H-*w4* x kw4 F_2	116	-	-	116	-	-	-
222-A-3 (222)	10	-	-	10	-	-	-
Clark-*w4* (C-*w4*)	10	-	-	10	-	-	-
222 x C-*w4* F_1	5	-	-	5	-	-	-
222 x C-*w4* F_2	109	-	-	109	-	-	-
E30-D-1 (E)	10	-	10	-	-	-	-
Clark (C)	10	10	-	-	-	-	-
E x C F_1	4	4	-	-	-	-	-
E x C F_2	112	84	28	-	3:1	0.00	1.00
E x C-*w4* F_1	3	-	3	-	-	-	-
E x C-*w4* F_2	111	-	82	29	3:1	0.08	0.78

Table 2.7 Segregation of flower color of F_3 families derived from a cross between E30-D-1 and Clark-*w4*

Line	Number of families				Expected ratio	χ^2 value	Probability (*P* value)
	Total	Fixed for light purple	Segregating	Fixed for near white			
E30-D-1 x Clark-*w4* F_3 (light purple)[a]	40	16	24	-	1:2	0.80	0.37
E30-D-1 x Clark-*w4* F_3 (near white)[b]	10	-	-	10	-	-	-

[a]F_3 families derived from F_2 plants with light purple flowers

[b]F_3 families derived from F_2 plants with near white flowers

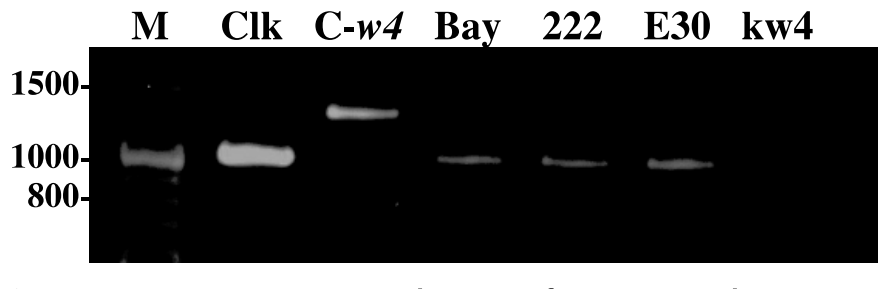

Figure 2.2

Agarose gel electrophoresis of RT-PCR products correspond to the entire coding region of *DFR2* gene in soybean and *Glycine soja*

λ, molecular marker λ/*Hind*III; C, Clark; C4; Clark-*w4*; 222, 222-A-3; E30, E30-D-1; K, kw4. The size of migration marker (bp) is shown on the left of the gel.

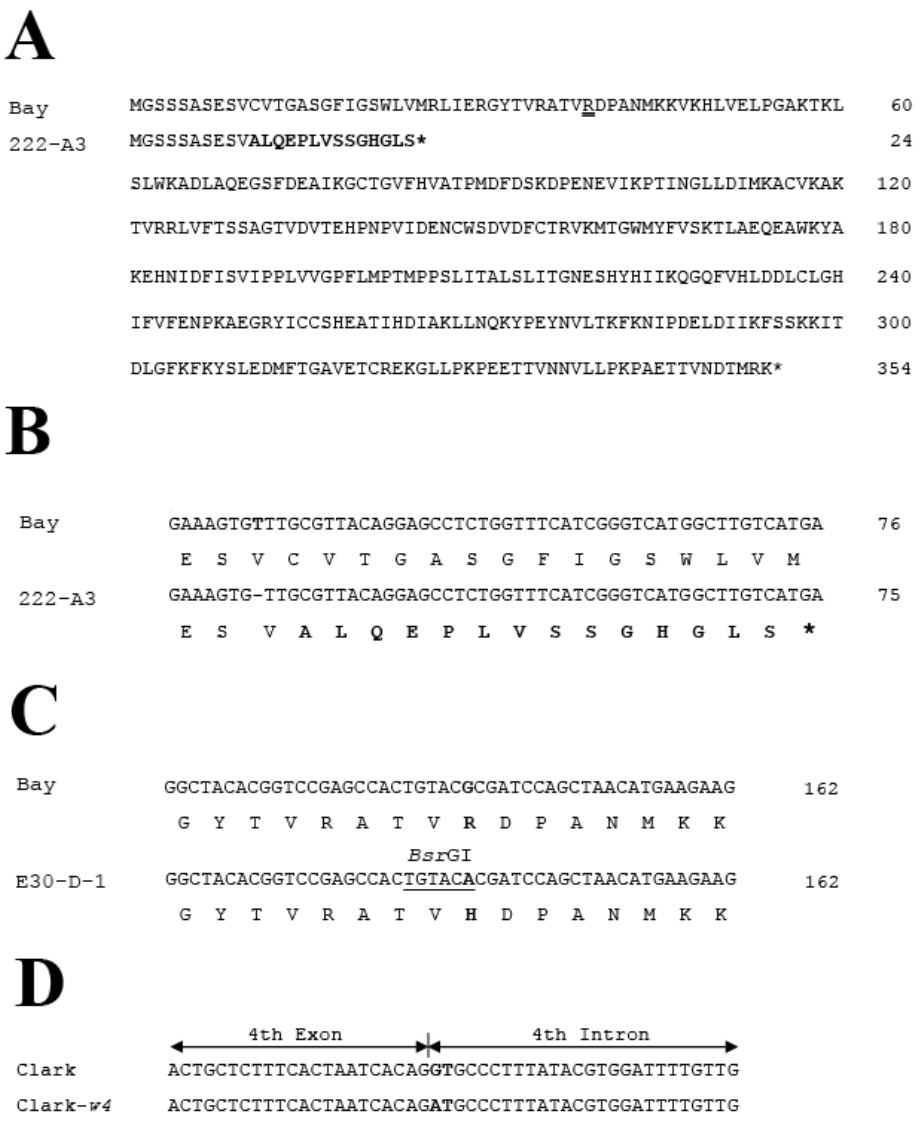

Figure 2.3

Nucleotide and amino acid polymorphisms of *DFR2* gene in flower color variants of soybean

(A) Amino acid sequence of *DFR2* gene from Bay. Amino acids polymorphic in 222-A-3 are shown in bold. The substitution of single amino acid is double-underlined in E-30-D-1. (B) Alignment of partial cDNA sequence and deduced amino acids from Bay and 222-A-3. Amino acids polymorphic in 222-A-3 are shown in bold. (C) Alignment of partial cDNA sequence and deduced amino acids from Bay and E30-D-1. Polymorphic nucleotides and amino acids are shown in bold. (D) Alignment of partial genome sequence around the end of the 4th exon of Clark and Clark-*w4*. Two nucleotides corresponding to the 5' splice site are shown in bold.

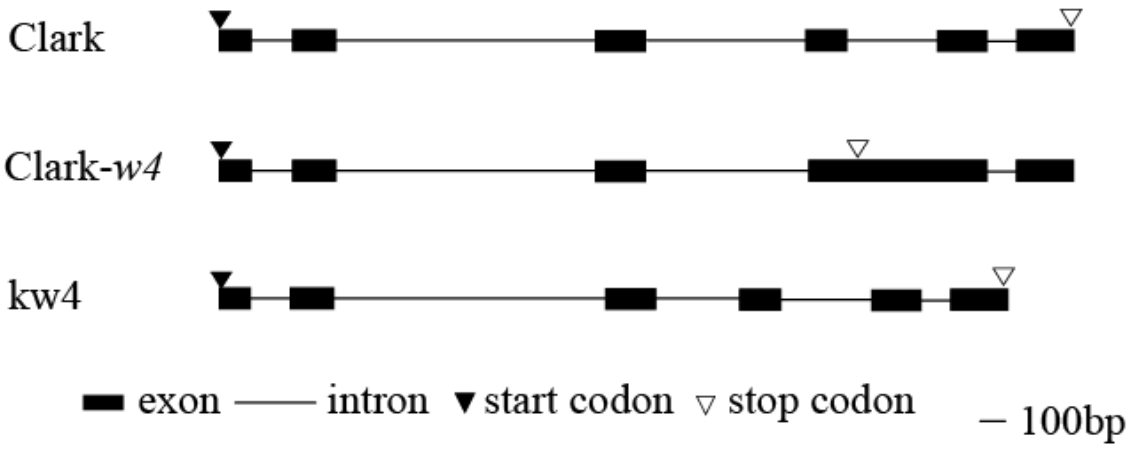

Figure 2.4

Intron/exon structure of *DFR2* gene from Clark (*G. max*), a Clark near-isogenic line with *w4* allele, Clark-*w4* and kw4, a *Glycine soja* accession

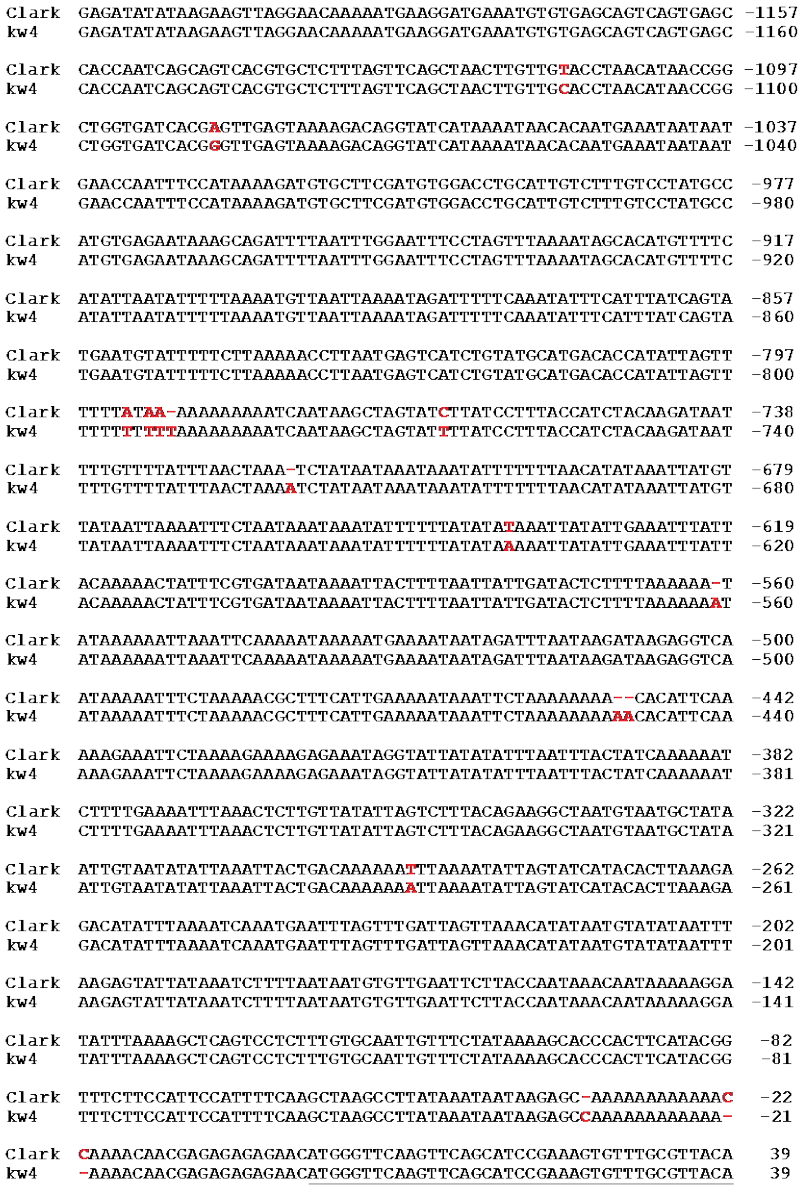

Figure 2.5

Alignment of the 5' upstream region of *DFR2* gene in soybean cultivar Clark and a *Glycine soja* accession kw4

Polymorphic nucleotides are shown in red font. Coding region of *DFR2* gene is underlined.

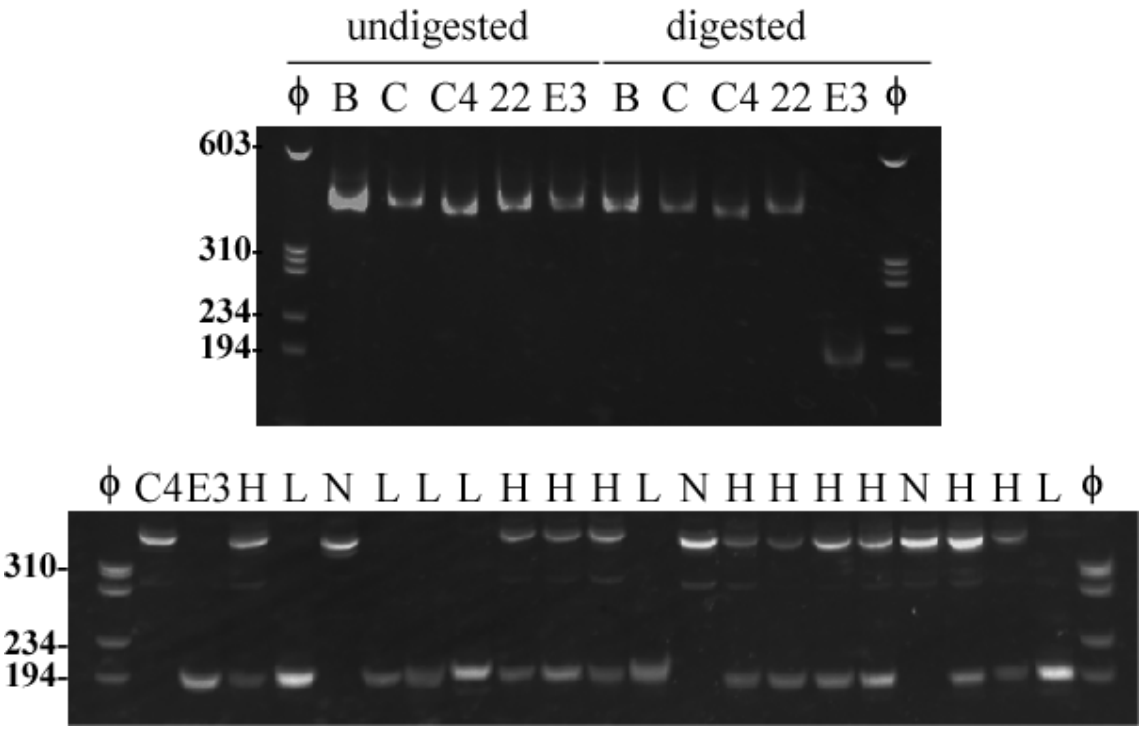

Figure 2.6

Upper panel-results of CAPS analysis for flower color variants

PCR products amplified with CAPS primers were digested by *Bsr*GI and the digests were separated on an 8 % polyacrylamide gel. ϕ: molecular marker ϕx174/*Hae*III; B: Bay; C: Clark; C4: Clark-*w4*; 22: 222-A3; E3: E30-D-1.

Lower panel-results of CAPS analysis in an F_2 population derived from a cross between E30-D-1 and Clark-*w4*

ϕ, ϕx174/*Hae*III; C4, Clark-*w4*; E3, E30-D-1; H, F_2 plants segregating for flower color; L, F_2 plants fixed for light purple flower; N, F_2 plants fixed for near white flower. The migration of size markers is shown to the left of the gel.

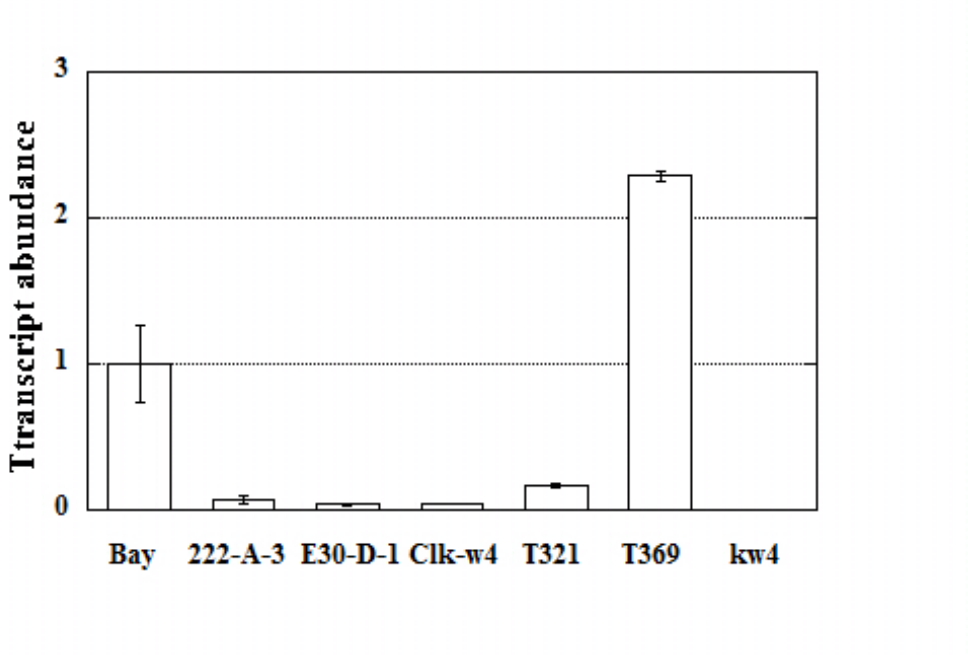

Figure 2.7

Expression of *DFR2* gene relative to the cultivar Bay of flower petals in soybean and *Glycine soja*

Transcript levels were standardized to the transcript level of actin. The means and SDs from three biological replications are exhibited.

2.2 Allelic variation of soybean flower color gene *W1* encoding flavonoid 3'5'-hydroxylase

2.2.1 Materials and methods

Plant materials

Flower color variant E023-H-12 with light purple flowers was isolated from EMS-treated population of Bay (Anai, 2012) (Fig. 1.1). A NIL of a cultivar Clark at *w1* locus, Clark-*w1* (L63-2373) has white flowers and tawny pubescence (*w1W2w3W4WmWpT*). PI 424008C is a *Glycine soja* line with white flowers that developed from the progeny of PI 424008A with purple flowers. Seeds of E023-H-12 were obtained from Dr. Anai at Saga University. Seeds of the NILs, PI 424008A and PI 424008C were provided by the USDA Soybean Germplasm Collection. The NIL was developed by backcrossing the white flower trait six times from T139 into Clark (Bernard *et al*., 1991).

Genetic analysis

Mutant E023-H-12 was crossed with Harosoy and Clark-*w1*. Flowers of E023-H-12 were emasculated one day before opening and fertilized with pollen from Harosoy or Clark-*w1*. Hybridity of the F_1 plants was ascertained either by flower color or pubescence color.

A total of seven F_1 and 100 F_2 seeds derived from E023-H-12 x Harosoy were field-planted at the National Institute of Crop Science. A total of seven F_1 and 120 F_2 seeds derived from E023-H-12 x Clark-*w1* were planted. F_3 progeny of E023-H-12 x Clark-*w1* crossing, thirty seeds each derived from forty F_2 plants having light

purple flowers and from four F_2 plants having white flowers were planted. In addition, fifteen F_3 seeds derived from a F_2 plant having purple flowers were planted. N, P and K were applied at 3.0, 4.4 and 8.3 g m-2, respectively. Plants were individually grown with spacing of 70 cm between rows and 10 cm between plants. Flower color was recorded in individual F_1, F_2 and F_3 plants.

Molecular cloning

Total RNA was extracted from banner petals (200 mg) of Bay, Clark-*w1*, E023-H-12, PI 424008A and PI 424008C using the TRIZOL Reagent (Invitrogen) according to the manufacturer's instructions. cDNA was synthesized by reverse transcription of 5 μg of total RNA using the Superscript III First-Strand Synthesis System (Invitrogen) and an oligo(dT) primer according to the manufacturer's instructions. The full-length cDNA was cloned by end-to-end PCR. Primers and PCR conditions followed those in a previous report (Takahashi*, et al.*, 2010). The PCR products were cloned into pCR 2.1 vector (Invitrogen) and sequenced.

DNA fragments including the entire coding region of *F3'5'H* gene were amplified by end-to-end PCR using genomic DNA of Clark-*w1*, E023-H-12 and F_3-9-2 (an F_3 plant fixed with purple flowers). Primers are shown in Table 2.8. The PCR mixture contained 10 ng of genomic DNA, 10 pmol of each primer, 10 pmol of nucleotides and 1 unit of ExTaq in 1×ExTaq Buffer in a total volume of 50 μl. A 30 sec denaturation at 94°C was followed by 30 cycles of 30 sec denaturation at 94°C, 1 min annealing at 60°C and 1 min extension at 72°C. A final 7 min extension at 72°C completed the program. The PCR products were cloned into the pCR 2.1 vector and

sequenced.

Sequencing analysis

Nucleotide sequences of both strands were determined with the BigDye terminator cycle method using an ABI3130 Genetic Analyzer (Applied Biosystems). Primers used for sequencing cDNA clones are those in a previous report (Takahashi *et al*., 2010). Primers used for sequencing the genomic clones are shown in Table 2.8. Sequences were aligned using ClustalW (http://clustalw.ddbj.nig.ac.jp/top-j.html) at default settings.

dCAPS and indel analysis

Genomic DNA of Clark-*w1*, E023-H-12, PI 424008A and PI 424008C was extracted from trifoliolate leaves by CTAB method (Murray and Thompson, 1980). For co-segregation test of the crossing from E023-H-12 and Clark-*w1*, total DNA was also extracted from 44 F_2 families and 15 F_3 families (Kamiya and Kiguchi, 2003). A bulk of five seeds each from F_2 or F_3 progeny were used. And the F_3 progeny was derived from a F_2 plant with purple flower. Primers for dCAPS analysis are designed (Table 2.8). A mismatched nucleotide (T) was incorporated in the end of forward primer to generate a *Bsr*GI site in the PCR product of E023-H-12 (Fig. 2.8). The SNP within the restriction site was expected to produce a polymorphism. 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 x ExTaq Buffer supplied by the manufacturer (Takara) in a total volume of 25 μl. The PCR thermal profile featured an initial 30 sec

denaturation at 94°C followed by 30 cycles of 30 sec denaturation at 94°C, 1 min annealing at 60°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 *Bsr*GI, and the digests were separated on an 8% nondenaturing polyacrylamide gel in 1 x TBE buffer (90 mM Tris-borate, 2mM EDTA, pH 8.0). After electrophoresis, the gel was stained with ethidium bromide and the DNA fragments were visualized under UV light.

PCR primers to recognize the 65 bp indel were designed (Table 2.8 and Fig. 2.9). The PCR profile and electrophoresis conditions were identical with the dCAPS analysis except that annealing temperature was performed at 62°C. PCR products were separated by polyacrylamide gel electrophoresis as described above.

Quantitative real-time PCR

For quantitative real-time PCR, total RNA (5 μg) from each of three replicate banner petal samples of Bay, E023-H-12, PI 424008A and PI 424008C was reverse-transcribed using the Superscript III First-Strand Synthesis System and an oligo(dT) primer. The expression level of the soybean *actin 1* gene (GenBank accession number: J01298) (Shah, Hightower and Meagher, 1983) was used to normalize target gene expression. Primers for *F3'5'H* gene are shown in Table 2.8. Primers for the *actin 1* gene and PCR conditions were similar to the last section.

Accession numbers

Sequence data was deposited in the DDBJ Data Libraries under accession numbers LC010617 (Bay), LC010619 (cDNA of Clark-w1), LC010618 (cDNA of

E023-H-12), LC010620 (cDNA of PI 424008A) and LC010621 (cDNA of PI 424008C).

2.2.2 Results

Genetic analysis

F_1 plants derived from a cross between E023-H-12 and Harosoy had purple flowers (Table 2.9). F_2 plants segregated into 68 plants with purple flowers and 25 plants with light purple flowers. The segregation ratio fitted 3 purple to 1 light purple. The results suggested that flower color of E023-H-12 was controlled by a single gene and that purple flower was dominant to light purple flower.

F_1 plants derived from a cross of E023-H-12 and Clark-*w1* had light purple flowers. A total of 112 F_2 plants segregated into 88 plants with light purple flowers, 23 plants with white flowers and a single plant with purple flowers.

Ignoring the single purple-flowered plant, the segregation ratio fitted 3 dilute to 1 white. The results suggested that the *W1* locus was responsible for light purple flower color and that light purple flower was dominant to white flower. Four F_3 families developed from F_2 white-flowered plants were fixed with white flowers (Table 2.10). Forty F_3 families derived from F_2 light purple-flowered plants segregated into 12 families with light purple flowers and 28 families segregating for light purple and white flowers. The segregation ratio was 1 line fixing with light purple flower and 2 lines segregating for flower colors. So, the segregation results indicated that a single gene controlled flower color. In addition, 15 F_3 plants generated from the single purple-flowered F_2 plant segregated into 11 plants with purple flowers and 4 plants with light purple flowers.

The new allele responsible for light purple flower at *W1* locus was

designated as *w1-lp2*. Dominance relationship was *W1* > *w1-lp2* > *w1*.

Molecular cloning

Fragments of about 1.7 kb were amplified by reverse transcription PCR from flower petals of Bay, Clark-*w1*, E023-H-12, PI 424008A and PI 424008C. Amplified fragments of Bay, E023-H-12 and PI 424008A consisted of 1,657 nucleotides. The coding region was 1,530 bp encoding 509 amino acids (Fig. 2.10). Nucleotide sequence of Bay had three SNPs in the coding region compared with Clark at nucleotide positions 1,059, 1,424 and 1,509. The second SNP was non-synonymous (valine in Bay and glutamic acid in Clark at amino acid position 475) whereas the other SNPs were synonymous. The results indicated that Bay and Clark had the gene structure of *GmF3'5'H-a* and *GmF3'5'H-b*, respectively (Guo and Qiu, 2013). Nucleotide sequence of PI 424008A had three SNPs compared with Bay; substitution in the coding region (nucleotide position 1,374 in the indel region) and two SNPs in the 3' untranslated region. Deduced amino acids of PI 424008A were identical with Bay. The nucleotide sequence of E023-H-12 was identical with Bay except for a SNP at nucleotide position 410 (Fig. 2.9). The SNP generated amino acid polymorphism (histidine in Bay and leucine in E023-H-12) at amino acid position 137 (Fig. 2.10).

The nucleotide sequence of the genomic fragment including the entire coding region was compared between Clark-*w1* and E023-H-12 (Fig. 2.11). There were two SNPs, a 181 bp indel and of a 3 bp indel in the first intron (Fig. 2.11). Eight SNPs and one 2 bp indel existed in the second intron. The nucleotide sequence of F_3-9-2 (an F_3 plant fixed with purple flowers derived from a cross between E023-H-12

and Clark-*w1*) was identical with Clark-*w1* from the first exon to the second intron. In contrast, the nucleotide sequence of the third exon was identical with E023-H-12 (Fig. 2.11).

PCR fragments of Clark-*w1* and PI 424008C consisted of 1,710 nucleotides. Their nucleotide sequence was identical with a white-flowered cultivar Williams 82. There was a 65 bp insertion of tandem repeats in the coding region.

dCAPS and indel analysis

For detecting the SNP in E023-H-12, dCAPS primers were designed. Amplified products were about 120 bp in Clark, Clark-*w1*, Bay and E023-H-12 as expected (Fig. 2.12A). After digestion with *Bsr*GI, the fragment of E023-H-12 was unaffected, whereas the other materials generated a shorted band of about 100 bp.

Indel analysis, Clark, Bay, E023-H-12 and PI 424008A generated a band of about 300 bp, but a longer band of about 350 bp in Clark-*w1* and PI 424008C. The result confirmed that there is an indel of 65 bp tandem repeats in the *F3′5′H* gene of Clark-*w1* and PI 424008C.

In Fig. 2.12B, dCAPS analysis of an F_2 population derived from E023-H-12 and Clark-*w1* revealed that plants fixed with light purple flowers had only longer band similar to E023-H-12, plants fixed with white flowers had only shorter band similar to Clark-*w1* and plants segregating for flower colors had both bands. In contrast, indel analysis revealed that plants fixed with light purple flowers had only shorter band, plants fixed with white flowers had only longer band and plants segregating for flower colors had both bands. Thus, the dCAPS and indel markers co-

segregated with flower color.

Among the 15 F_3 plants derived from a purple-flowered F_2 plant, 11 plants had purple flowers and four plants had light purple flowers. dCAPS analysis indicated that light purple-flowered plants had only longer band and purple-flowered plants had only shorter band or both bands (Fig. 2.12C upper panel). For indel analysis, 15 plants generated a band of about 300 bp in all plants that are about 50 bp shorter than that of Clark-*w1* (Fig. 2.12C lower panel).

Gene expression

The expression level of the *F3'5'H* gene in the flower petals of E023-H-12 was similar to Bay (Fig. 2.13). In contrast, the gene expression level was higher in the flower petals of PI 424008A (1.5 times of Bay) and PI 424008C (2.3 times of Bay).

2.2.3 Discussion

Flower color variant E023-H-12 was isolated from EMS-treated population of cultivar Bay. Genetic analysis showed that a new allele of the *W1* locus was responsible for dilute flower color. The new allele was designated as *w1-lp2*. Dominance relationship of the locus was *W1* > *w1-lp2* > *w1*.

Anthocyanin composition in light purple flower petals associated with the *w1-lp2* allele was qualitatively identical with that in purple flowers, but the total amount was decreased by 42 %. The reduction may be responsible for the light purple color. This is in contrast to the *w1-lp* allele which was associated with the production of unique anthocyanin components (Takahashi*, et al.*, 2010). Flavonol

glycosides F1 to F5 of light purple petals were increased, resulting in a 34 % increase of the total flavonol glycoside amount. This is in contrast to plants with the *w1* allele whose flavonol glycoside content were similar to those of plants with the *W1* allele (Iwashina*, et al.*, 2007).

Nucleotide sequence and amino acids of the *F3′5′H* gene was identified with Bay except one base substitution in E023-H-12. The substitution changed one amino acid at position 137. It was histidine in Bay and leucine in E023-H-12. The amino acid polymorphism was apart from the putative substrate recognition sites (SRS) or functional domains for hydroxylation activity. However, the polymorphism of B09121 was adjacent to the SRS2 (Gotoh, 1992; Seitz *et al*., 2007; Takahashi *et al*., 2010; Falginella *et al*., 2010) (Fig. 2.10). The dCAPS and indel analysis suggested that the SNP was co-segregated with flower colors. So, the amino acid substitution may be responsible for light purple flower colors in E023-H-12. The expression level of the *F3′5′H* gene between Bay and E023-H-12 had no significant difference. Therefore, the expression level may not responsible for light purple flower color. Transgenic experiment may be necessary to identify the functional importance of the amino acid substitution for flower color.

A single plant having purple flowers was found in the F_2 population generated from a cross between E023-H-12 and Clark-*w1*. The plant produced F_3 plants with purple and light purple flowers at a 3:1 ratio, suggesting that the F_2 plant had the heterozygous allele *W1w1-lp2*. The dCAPS markers of the F_3 plants co-segregated with flower color; plants with light purple flowers only had the longer fragment whereas plants with purple flowers only had the shorter fragment or both

fragments.

Intragenic recombination (Stadler, 1973) in the *F3'5'H* gene of the purple-flowered F_2 plant can explain these results. The nucleotide sequence of the genomic fragment of an F_3 plant fixed with purple flowers was identical with that of Clark-*w1* from the first exon to the second intron (Fig. 2.11). However, the nucleotide sequence of the third exon was identical with E023-H-12. Intragenic recombination may have occurred between the last SNP in the second intron (nucleotide position 3022) and the 65 bp indel (from nucleotide position 4211) (in the gray region in Fig. 2.11). The chimeric nature of the gene strongly suggests that outcrossing may not be responsible.

Purple-flowered plants were also generated in the F_2 and F_3 populations derived from a cross between plants with light purple flowers (*w1-lp* allele) and plants with white flowers (*w1* allele) (Takahashi*, et al.*, 2010). The frequent occurrence of intragenic recombination in populations derived from a cross with plants having *w1* allele suggests that existence of tandem repeats in the 65-bp insertion is possibly responsible for the high frequency of intragenic recombination.

The nucleotide sequence of the *F3'5'H* gene in PI 424008A had three unique SNPs that have not been observed in other purple-flowered germplasm analyzed so far. On the other hand, the nucleotide sequence of PI 424008C was completely identical with Clark-*w1*, Williams 82 and other white-flowered soybean cultivars. Thus, the *F3'5'H* gene of PI 424008C had the 65 bp insertion of tandem repeats and it did not have the SNPs unique to PI 424008A. These results suggest that the white flower of PI 424008C may have arisen from outcrossing of PI 424008A. Mutations of

W1 gene give rise to unique flavonoid compositions and display a wide variety of flower color patterns in soybean, from purple, light purple, purple/white variegation to white.

Table 2.8 PCR primers used in this study

Purpose	Forward primer (5'-3')	Reverse primer (5'-3')
Cloning of genomic clone	AACTAGCAAATTAATTAGCTT TTAGACATGGTAATGGCTCA TTTGCTGCATTTGGAATTGG	TTTATCACTCATGTAGGTCC GGACCACTCTATTATACTTG CAACCCAAACATTACTTAT
Sequencing of genomic clone	CTCTCTCACTAAGTGCTAAA TATGTCGTCATCTGTCATTG TCTTCCCTCTATTTGTACAT TTTGCTGCATTTGGAATTGG TGATGATGTAAATGTTAGTGC	TATTAGCTTCGGTTTGGTTC TGCATATAGAAAATAAGATGG ATGGTTCTAACGAAGTTTTAT TTTATCACTCATGTAGGTCC ATGTTCACATTCAGCCTAGT
dCAPS analysis[1]	GCTAAGAAAACTAAGTAACTTGT	GCCTCATCCCTCTTGTTACA
Indel analysis	TGGTGCTGGGAGGAGGATTT	TTTATGTAGCCACAGCCACA
Real-time PCR	GGACTTGCAAGGCATAGAGC	CCCCATCGGAGTTCTCACTA

[1]A mismatched nucleotide (T, double-underlined) was incorporated in the forward primer to generate a *Bsr*GI site.

Table 2.9 Segregation of flower color in F_1 plants, F_2 and F_3 populations derived from crosses between E023-H-12 and Harosoy or line Clark-*w1*

Generation	Number of plants				Expected ratio	χ^2 value	Probability (*P* value)
	Total	Purple	Light purple	White			
E023-H-12 (E)	10	-	10	-	-	-	-
Harosoy (H)	10	10	-	-	-	-	-
E x H F_1	5	5	-	-	-	-	-
E x H F_2	93	68	25	-	3:1	0.18	0.68
E023-H-12 (E)	10	-	10	-	-	-	-
Clark-*w1* (C-*w1*)	10	-	-	10	-	-	-
E x C-*w1* F_1	5	-	5	-	-	-	-
E x C-*w1* F_2	112	1	88	23	3:1[a]	1.08	0.30
E x C-*w1* F_3 (purple)[b]	15	11	4	-	3:1	0.22	0.88

[a]χ^2 value was estimated excluding a plant with purple flowers.

[b]Segregation of F_3 family derived from a F_2 plant having purple flowers.

Table 2.10 Segregation of flower color in F_3 families derived from a cross between E023-H-12 and Clark-*w1*

Line	Number of families				Expected ratio	χ^2 value	Probability (*P* value)
	Total	Fixed for light purple	Segregating	Fixed for white			
E023-H-12 x Clark-*w1* F_3 (light purple)[a]	40	12	28	-	1:2	0.20	0.65
E023-H-12 x Clark-*w1* F_3 (white)[b]	4	-	-	4	-	-	-

[a]F_3 families derived from F_2 plants with dilute flowers.

[b]F_3 families derived from F_2 plants with white flowers.

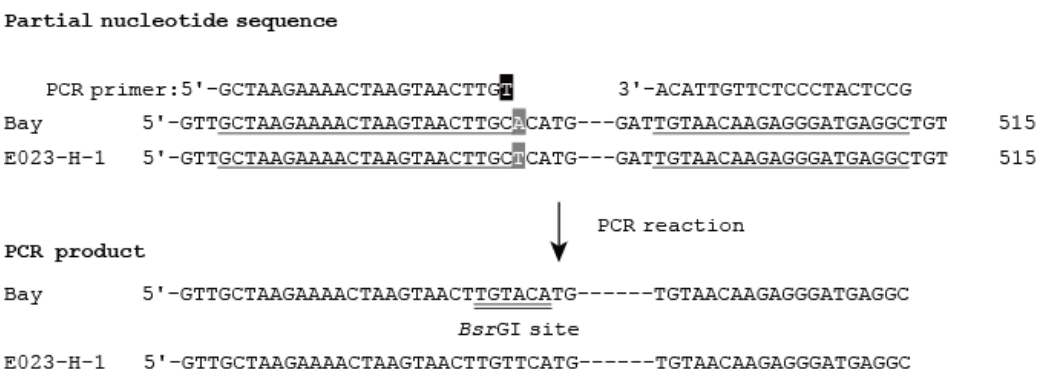

Figure 2.8

Schematic presentation of dCAPS analysis to detect a SNP between Bay and E023-H-12 (highlighted in gray)

A mismatched nucleotide (T) highlighted in black was incorporated in the forward primer to generate a *Bsr*GI site (double underlined) in the PCR product of Bay. Annealing sites of PCR primers are underlined.

```
Bay          GAAGTTGCTAAGAAAACTAAGTAACTTGCACATGCTTGGAGGAAAGGCACTTGATGATTGGGCCCAAATTCGAGAT   456
E023-H-12    GAAGTTGCTAAGAAAACTAAGTAACTTGCTCATGCTTGGAGGAAAGGCACTTGATGATTGGGCCCAAATTCGAGAT   456
Clark-w1     GAAGTTGCTAAGAAAACTAAGTAACTTGCACATGCTTGGAGGAAAGGCACTTGATGATTGGGCCCAAATTCGAGAT   456

Bay          GAAGAGATGGGGCACATGCTTGGTGCAATGTACGATTGTAACAAGAGGGATGAGGCTGTGGTGGTGGCGGAGATGT   532
E023-H-12    GAAGAGATGGGGCACATGCTTGGTGCAATGTACGATTGTAACAAGAGGGATGAGGCTGTGGTGGTGGCGGAGATGT   532
Clark-w1     GAAGAGATGGGGCACATGCTTGGTGCAATGTACGATTGTAACAAGAGGGATGAGGCTGTGGTGGTGGCGGAGATGT   532

Bay          TGACATATTCAATGGCCAACATGATTGGCCAAGTTATATTGAGTCGTCGAGTGTTTGAGACAAAGGGTTCGGAGTC   608
E023-H-12    TGACATATTCAATGGCCAACATGATTGGCCAAGTTATATTGAGTCGTCGAGTGTTTGAGACAAAGGGTTCGGAGTC   608
Clark-w1     TGACATATTCAATGGCCAACATGATTGGCCAAGTTATATTGAGTCGTCGAGTGTTTGAGACAAAGGGTTCGGAGTC   608

Bay          TAACGAGTTCAAGGACATGGTGGTTGAGCTCATGACCGTTGCTGGTTACTTCAACATTGGTGACTTCATACCCTTT   684
E023-H-12    TAACGAGTTCAAGGACATGGTGGTTGAGCTCATGACCGTTGCTGGTTACTTCAACATTGGTGACTTCATACCCTTT   684
Clark-w1     TAACGAGTTCAAGGACATGGTGGTTGAGCTCATGACCGTTGCTGGTTACTTCAACATTGGTGACTTCATACCCTTT   684

Bay          TTGGCCAAGTTGGACTTGCAAGGCATAGAGCGTGGCATGAAGAAGTTGCACAAGAAGTTTGATGCGTTGTTAACGA   760
E023-H-12    TTGGCCAAGTTGGACTTGCAAGGCATAGAGCGTGGCATGAAGAAGTTGCACAAGAAGTTTGATGCGTTGTTAACGA   760
Clark-w1     TTGGCCAAGTTGGACTTGCAAGGCATAGAGCGTGGCATGAAGAAGTTGCACAAGAAGTTTGATGCGTTGTTAACGA   760

Bay          GCATGATTGAGGAGCATGTTGCTTCTAGTCACAAGAGAAAGGGCAAGCCCGATTTCTTAGACATGGTAATGGCTCA   836
E023-H-12    GCATGATTGAGGAGCATGTTGCTTCTAGTCACAAGAGAAAGGGCAAGCCCGATTTCTTAGACATGGTAATGGCTCA   836
Clark-w1     GCATGATTGAGGAGCATGTTGCTTCTAGTCACAAGAGAAAGGGCAAGCCCGATTTCTTAGACATGGTAATGGCTCA   836

Bay          TCATAGTGAGAACTCCGATGGGGAGGAACTATCGCTCACCAACATCAAGGCACTACTCTTGAACCTATTCACCGCA   912
E023-H-12    TCATAGTGAGAACTCCGATGGGGAGGAACTATCGCTCACCAACATCAAGGCACTACTCTTGAACCTATTCACCGCA   912
Clark-w1     TCATAGTGAGAACTCCGATGGGGAGGAACTATCGCTCACCAACATCAAGGCACTACTCTTGAACCTATTCACCGCA   912

Bay          GGCACCGATACATCTTCAAGTATAATAGAGTGGTCCTTAGCCGAGATGTTGAAGAAGCCCAGCATAATGAAGAAGG   988
E023-H-12    GGCACCGATACATCTTCAAGTATAATAGAGTGGTCCTTAGCCGAGATGTTGAAGAAGCCCAGCATAATGAAGAAGG   988
Clark-w1     GGCACCGATACATCTTCAAGTATAATAGAGTGGTCCTTAGCCGAGATGTTGAAGAAGCCCAGCATAATGAAGAAGG   988

Bay          CTCATGAAGAAATGGACCAAGTCATAGGAAGGGATCGCCGTCTCAAAGAATCTGACATACCAAAGCTTCCCTACTT   1064
E023-H-12    CTCATGAAGAAATGGACCAAGTCATAGGAAGGGATCGCCGTCTCAAAGAATCTGACATACCAAAGCTTCCCTACTT   1064
Clark-w1     CTCATGAAGAAATGGACCAAGTCATAGGAAGGGATCGCCGTCTCAAAGAATCTGACATACCAAAGCTTCCCTACTT   1064

Bay          CCAAGCCATTTGCAAAGAGACCTATAGAAAGCACCCTTCAACACCCCTAAACCTGCCTCGAATCTCATCTGAACCG   1140
E023-H-12    CCAAGCCATTTGCAAAGAGACCTATAGAAAGCACCCTTCAACACCCCTAAACCTGCCTCGAATCTCATCTGAACCG   1140
Clark-w1     CCAAGCCATTTGCAAAGAGACCTATAGAAAGCACCCTTCAACACCCCTAAACCTGCCTCGAATCTCATCTGAACCG   1140

Bay          TGCCAAGTGAATGGTTACTACATTCCCGAGAACACTAGGCTGAATGTGAACATTTGGGCCATAGGAAGAGACCCTG   1216
E023-H-12    TGCCAAGTGAATGGTTACTACATTCCCGAGAACACTAGGCTGAATGTGAACATTTGGGCCATAGGAAGAGACCCTG   1216
Clark-w1     TGCCAAGTGAATGGTTACTACATTCCCGAGAACACTAGGCTGAATGTGAACATTTGGGCCATAGGAAGAGACCCTG   1216

Bay          ATGTGTGGAACAATCCTTTGGAGTTTATGCCCGAGAGGTTTTTGAGTGGGAAGAATGCCAAAATTGACCCACGTGG   1292
E023-H-12    ATGTGTGGAACAATCCTTTGGAGTTTATGCCCGAGAGGTTTTTGAGTGGGAAGAATGCCAAAATTGACCCACGTGG   1292
Clark-w1     ATGTGTGGAACAATCCTTTGGAGTTTATGCCCGAGAGGTTTTTGAGTGGGAAGAATGCCAAAATTGACCCACGTGG   1292

Bay          GAATGATTTTGAGCTTATTCCATTTGGTGCTGGGAGGAGGATTTGTGCAGGGACTAGGATGGGGATTGTGTT----   1364
E023-H-12    GAATGATTTTGAGCTTATTCCATTTGGTGCTGGGAGGAGGATTTGTGCAGGGACTAGGATGGGGATTGTGTT----   1364
Clark-w1     GAATGATTTTGAGCTTATTCCATTTGGTGCTGGGAGGAGGATTTGTGCAGGGACTAGGATTTTGAGCTTATTCCAT   1368

Bay          --------------------------------------------------GGTTCACTACATTTTGGGCACTTTGGT   1391
E023-H-12    --------------------------------------------------GGTTCACTACATTTTGGGCACTTTGGT   1391
Clark-w1     TTGGTTCACTACATTTTGGGCTTATTCCATTTTTGAGCTTATTCCATTTGGTTCACTACATTTTGGGCACTTTGGT   1444

Bay          GCATTCGTTTGATTGGAAGCTACCCAATGGGGTGAGGGAGTTAGACATGGAGGAGTCCTTTGGGCTTGCCTTGCAA   1467
E023-H-12    GCATTCGTTTGATTGGAAGCTACCCAATGGGGTGAGGGAGTTAGACATGGAGGAGTCCTTTGGGCTTGCCTTGCAA   1467
Clark-w1     GCATTCGTTTGATTGGAAGCTACCCAATGGGGTGAGGGAGTTAGACATGGAGGAGTCCTTTGGGCTTGCCTTGCAA   1520

Bay          AAAAAGGTTCCACTTGCTGCTTTGGTTACCCCTAGGTTGAACCCAAGTGCTTACATTTCTTAGAATTGGTTGGGTT   1543
E023-H-12    AAAAAGGTTCCACTTGCTGCTTTGGTTACCCCTAGGTTGAACCCAAGTGCTTACATTTCTTAGAATTGGTTGGGTT   1543
Clark-w1     AAAAAGGTTCCACTTGCTGCTTTGGTTACCCCTAGGTTGAACCCAAGTGCTTACATTTCTTAGAATTGGTTGGGTT   1596

Bay          CGAATATTCACCAGCTATGTTCTCTAGCCTTATTTTGTTGTCCAATGATTTTGTGGCTGTGGCTACATAAATAAGT   1619
E023-H-12    CGAATATTCACCAGCTATGTTCTCTAGCCTTATTTTGTTGTCCAATGATTTTGTGGCTGTGGCTACATAAATAAGT   1619
Clark-w1     CGAATATTCACCAGCTATGTTCTCTAGCCTTATTTTGTTGTCCAATGATTTTGTGGCTGTGGCTACATAAATAAGT   1672

Bay          AATGTTTGGGTTG                                                                  1632
E023-H-12    AATGTTTGGGTTG                                                                  1632
Clark-w1     AATGTTTGGGTTG                                                                  1685
```

Figure 2.9 Alignment of *F3'5'H* gene in Bay, E023-H-12 and Clark-*w1*

One base substitution in E023-H-12 and 65 bp-indel in Clark-*w1* are shown in red fond. Primers of indel analysis were underlined.

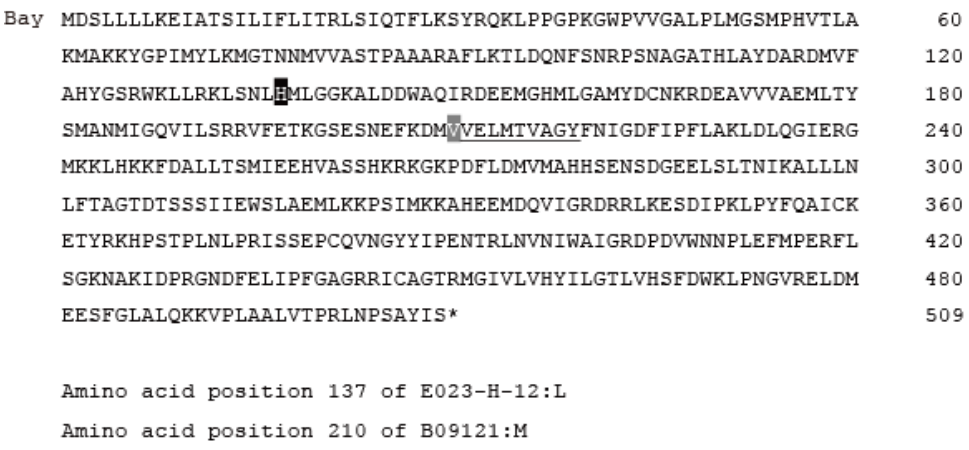

Figure 2.10

Amino acid sequence of the *F3'5'H* gene in soybean cultivar Bay

Amino acid substituted in E023-H-12 is highlighted in black. Amino acid that uniquely substituted in B09121 is highlighted in gray. Putative substrate recognition site 2 (SRS2) is underlined.

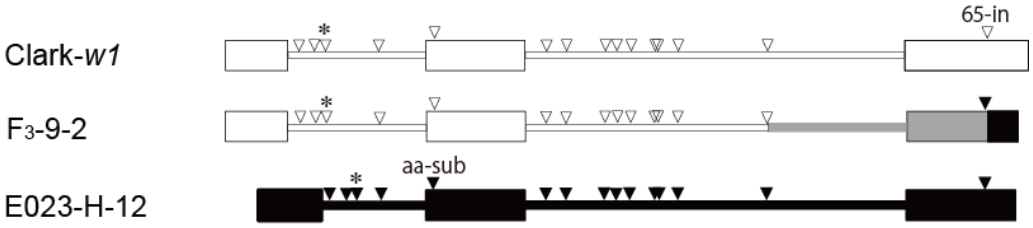

Figure 2.11

Gene structure of the *F3'5'H* gene of soybean lines Clark-*w1*, E023-H-12 and F_3-9-2 (an F_3 plant fixed with purple flowers derived from a cross between E023-H-12 and Clark-*w1*).

Exons and introns are indicated by thick boxes and narrow boxes, respectively (exons and introns of Clark-*w1* is shown in white boxes whereas those of E023-H-12 are shown by black boxes). Nucleotide polymorphisms are indicated by triangles (nucleotides identical with Clark-*w1* are shown by white triangles whereas those identical with E023-H-12 in black triangles). A large indel of 181 nucleotides in the first intron is indicated by asterisks. 'aa-sub' indicates a SNP associated with light purple flower in E023-H-12. '65-in' indicates a 65-bp insertion associated with white flower in Clark-*w1*. Gene fragment of F_3-9-2 probably derived from Clark-*w1* is indicated in white whereas fragment probably derived from E023-H-12 is shown in black. Fragment whose origin in unknown is shown in gray.

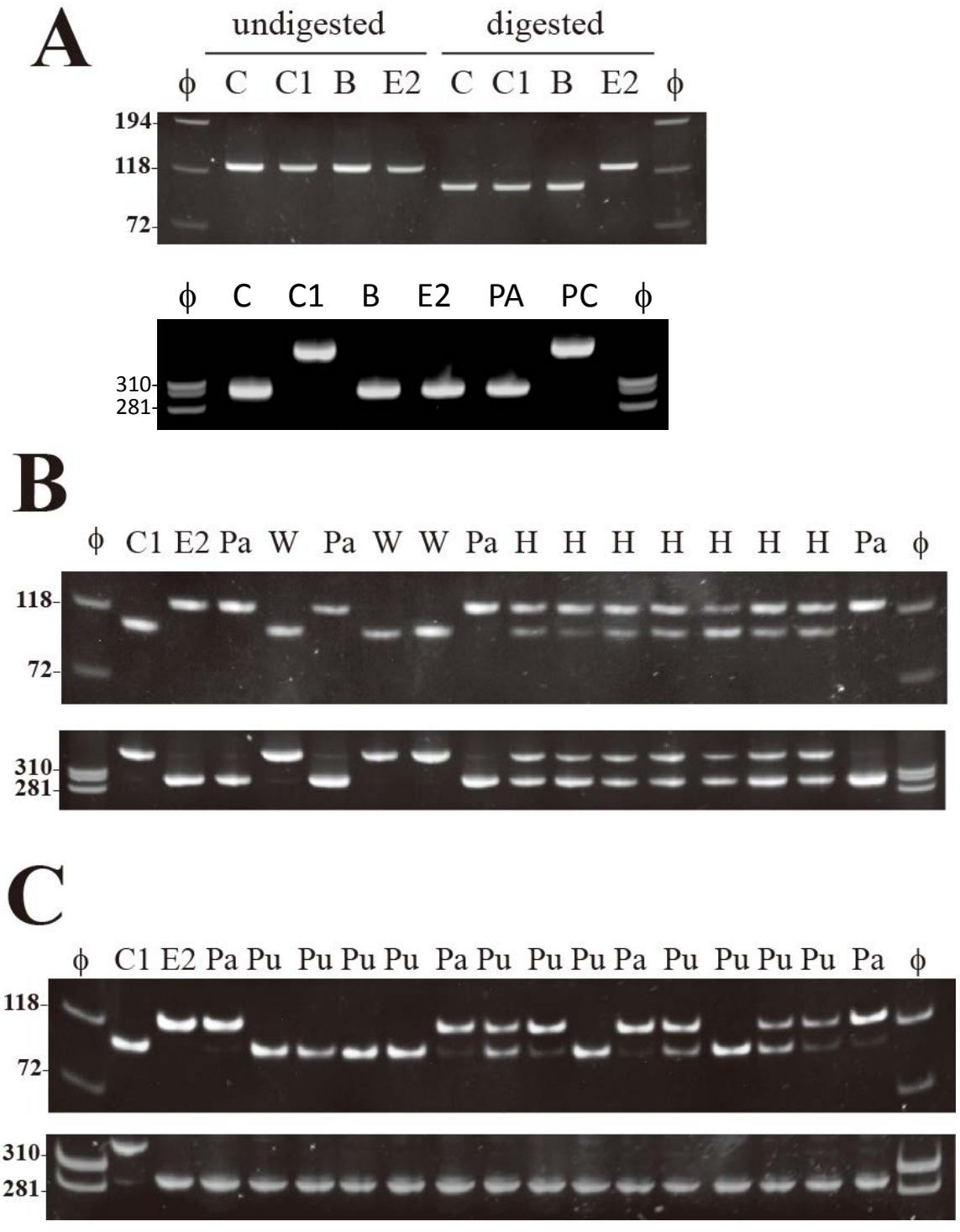

Figure 2.12

Results of dCAPS and indel analyses of *F3'5'H* gene in soybean and *Glycine soja*

(A) Results of dCAPS (upper panel) and indel analyses (lower panel) of flower color variants

PCR products amplified with dCAPS primers were digested by *Bsr*GI and the digests were separated on an 8% polyacrylamide gel. PCR products amplified with indel primers were similarly

electrophoresed. ϕ, molecular marker ϕx174/*Hae*III; C, Clark; C1, Clark-*w1*; B, Bay; E2, E023-H-12; PA, PI 424008A; PC, PI 424008C. The migration of size markers is shown to the left of the gel

(B) Results of dCAPS analysis (upper panel) and indel analysis (lower panel) of an F_2 population derived from a cross between E023-H-12 and Clark-*w1*

ϕ, ϕx174/*Hae*III; C1, Clark-*w1*; E2, E023-H-12; Pa, F_2 plants fixed for light purple flower; W, F_2 plants fixed for white flower; H, F_2 plants segregating for flower color

(C) Results of dCAPS analysis (upper panel) and indel analysis (lower panel) of F_3 population derived from an F_2 plant with purple flowers

ϕ, ϕx174/*Hae*III; C1, Clark-*w1*; E2, E023-H-12; Pa, F_3 plants with light purple flowers; Pu, F_3 plants with purple flowers

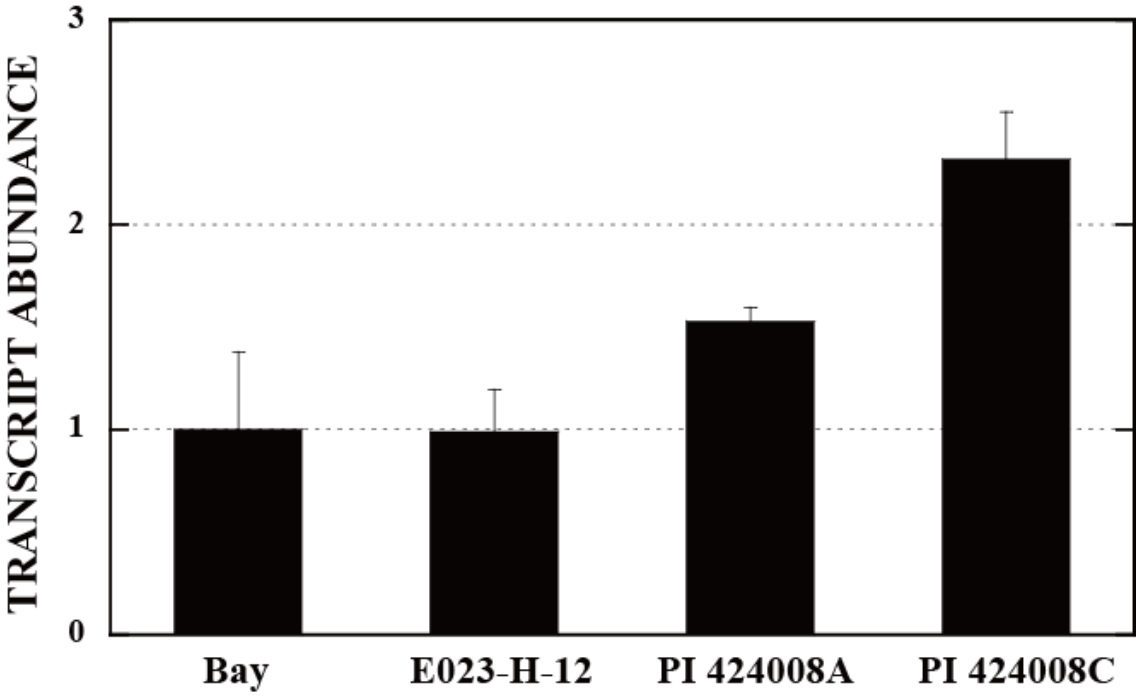

Figure 2.13

Expression of *F3'5'H* gene relative to cultivar Bay in flower petals of soybean and *Glycine soja*

Transcript levels were standardized to *actin 1* transcript. The means and SDs from 3 biological repeats are exhibited.

CHAPTER 3 Molecular genetic analysis of seed coat variegation

Background

Seed coat color of soybean is primarily controlled by five genes, *I*, *T*, *O*, *R* and *W1* (Palmer *et al*., 2004). Nagai and Saito (1923) designated *R*/*r* responsible for black/brown seed coat and *M* responsible for black concentric mottling on brown seed coat. But the relation between *M* and *R* was not clear. Later, genetic analysis suggested that *M* was allelic with *R* and *r*, thus the *R* locus consists of multiple alleles of *R* (black), *r* (brown) and *r-m* (black spots and/or concentric streaks on brown seed) (Bernard and Singh, 1969) (Fig. 3.1). The *R* gene was assigned to classical linkage group II (molecular linkage group K, chromosome 9) (Nagai and Saito, 1923; Weiss, 1970). Recently, Zabala and Vodkin (2014) and Gillman *et al*. (2011) suggested that *R* locus regulate the late stages of anthocyanin metabolic pathway affecting the pigmentation (Fig. 1.2). The *r-m* allele produces seeds that are highly variable not only among seeds in individual plants but also across generations. Progeny tests revealed that brown revertant sublines or black revertant sublines descended from plants with variegated seeds gave rise to stripe or striped + black-seeded plants, suggesting that revertants produced from *r-m* allele are not stable but interconvert among three seed pigmentation patterns (Chandlee and Vodkin, 1989). Unstable pigmentation presumed to be caused by a transposable element inserted in the *R* gene (Chandlee and Vodkin, 1989).

Transposable elements (TEs) are DNA fragments that can move one position of the genome to another, occasionally resulting in disruption of gene function, or

variegation of tissues when inserted in genes responsible for pigmentation. TEs existed in almost organisms, including bacteria, yeast, plants and mammals. Generally, they can divide into class I retrotransposons and class II DNA-based transposons.

In soybean, eleven transposable elements, *Tgm1*, fragmental elements *Tgm2* to *Tgm7*, *Tgm-Express1*, *Tgmt**, *Tgm9* and *Tgm10*, all belonging to the CACTA-family class II transposable elements, have been cloned (Toda *et al*., 2002; Zabala and Vodkin, 2005; Zabala and Vodkin, 2008; Xu *et al*., 2010). In *Glycine soja*, an element *Tgs1* was cloned from an accession having variegated flowers (Takahashi *et al.*, 2012).

Characteristic features of the CACTA-superfamily include terminal inverted repeats (TIR) and a 3 bp target site duplication upon integration (Gierl *et al*., 1989). Integration of *Tgm-Express1*, *Tgmt**, *Tgm9* or *Tgs1* into genes responsible for flavonoid biosynthesis is associated with flower or pubescence color variation (Table 3.1). A 5.7 kb element, *Tgm-Express1* was inserted into the second intron of *Wp* gene encoding flavanone 3-hydroxylase (Zabala and Vodkin, 2005). A 20.5 kb element, *Tgmt** was integrated into the first intron of the *T* gene encoding a flavonoid 3′-hydroxylase (Toda *et al*., 2002; Zabala and Vodkin, 2008). A 20.5 kb element *Tgm9* was inserted into the second intron of the *W4* gene encoding dihydroflavonol 4-reductase 2 (Xu *et al*., 2010). *Tgmt** and *Tgm9* are nearly identical (Xu *et al*., 2010). A 3.9 kb element *Tgs1* was inserted in the first exon of the *W1* gene encoding a flavonoid 3′5′-hydroxylase in *G. soja* line B00146-m having variegated flowers (Zabala and Vodkin, 2007; Takahashi *et al*., 2012). A 3.5 kb element *Tgm1* was

inserted in the exon of a *lectin* gene of a soybean cultivar having a recessive allele of the locus (Vodkin *et al*., 1983). *Tgs1* and *Tgm1* shared high similarity in nucleotide sequence, element size, and motifs of TIR and subterminal regions (STR). In addition, both lacked a functional transposase and were integrated into exons. These results suggest that they comprise a subgroup (Takahashi *et al.*, 2012).

Genetic analysis, genome information and gene expression assays suggested that loss of function of a MYB transcription factor gene, Glyma09g36990, might be responsible for brown seed coats (Gillman *et al*., 2011). This study was conducted to identify a possible causal genetic factor responsible for seed coat variegation in soybean.

3.1 Materials and Methods

Plant Materials

US cultivar Clark (*i-ii-i RR TT W1W1*) with yellow seed coat and black hilum, and a near-isogenic line (NIL) of Clark 63 with *i* and *r-m* alleles, L72-2040 (*ii r-mr-m TT W1W1*) were used (Fig. 3.1). Clark 63 differs Clark by the *rxp* and *Rps1* allele. Seeds of the NILs were provided by the USDA Soybean Germplasm Collection. The NIL was developed by backcrossing the variegated seed trait five times from PI 91073 into L67-7384, a Clark 63 NIL with *i* and *r* alleles (Bernard *et al*., 1991).

Molecular Cloning

For transposon cloning, genomic DNA of Clark and L72-2040 were extracted from leaves using CTAB method. To clone the entire gene of Glyma09g36990,

primers (a) GAACTCTCAGGAGAGGGTGTGGTAA and (b) ACAGTGACAACGACAACAGTCATAA were used to clone upstream fragment of Glyma09g36990 including the transposable element. Primers (c) TAACTGTTATCATCACCGTC and (d) TAGAATTTGAGGTCCCTTGTGAA were used to clone the downstream. The two fragments overlap to each other covering entire coding region of Glyma09g36990.

The PCR mixture contained 2.5 μl 10 × Advantage 2 Buffer, 2.0 μl dNTP Mixture (2.5 mM), 50 ng genomic DNA, 0.5 μl of each primer (10 μM), 0.5 μl 50 × Advantage 2 Polymerase (5 U/μl) and adjust a final volume of 25 μl. Initial denaturation 1 min at 95^{o}C, 40 cycles of 30 sec denaturation at 95^{o}C, 1 min annealing at 63^{o}C and 10 min extension at 68^{o}C. A final 12 min extension at 68^{o}C completed the program and reserved at 4^{o}C. The PCR was performed in an Applied Biosystems 9700 thermal cycler. PCR products were loaded on a 0.8 % agarose gel containing crystal violet and visualized under fluorescent light. The PCR fragment of about 14 kb was extracted from the gel and the entire DNA fragment was TA-cloned into pCR-XL-TOPO vector (Invitrogen) and transformed into the TOP10 electrocompetent cells supplied in the kit. The PCR fragments of about 1 and 1.4 kb were cloned into the pCR2.1 vector (Invitrogen).

Primers of (a) and (d) were used to amplified the entire gene of Glyma09g36990 in Clark. The PCR mixture contained 0.5 μg of genomic DNA, 10 pmol of each primer, 5 pmol of dNTPs and 1 unit of ExTaq in 1 x ExTaq Buffer supplied by the manufacturer (Takara Bio) in a total volume of 25 μl. A 30 sec denaturation at 94^{o}C was followed by 30 cycles of 30 sec denaturation at 94^{o}C, 1 min

annealing at 59°C and 2 min extension at 72°C. A final 7 min extension at 72°C completed the program. The PCR amplicons were cloned into the pCR2.1 vector.

The study of Todd and Vodkin (1993) revealed that extraction of RNA from seed coat of cultivars with black or brown seeds is difficult. So, total RNA was extracted from leaves using the TRIZOL Reagent for transposase cloning. To synthesize the cDNA, reverse transcription was conducted using the Superscript III First-Strand Synthesis System. The transposase cDNA was amplified by PCR using two combinations of primers. Primers P1 and P3 were used to clone ORF2 while P2 and P3 were for ORF1 (P1-TTCTGAGCGCTTGTTATACT, P2-AACTTTGAGAGTATGACCGA and P3-TTAATGTTGAGACATGAATTGT).

The PCR mixture contained 0.5 µg of the cDNA, 10 pmol of each primer, 10 pmol of nucleotides and 1 unit of Advantage 2 Polymerase Mix in 1 x Advantage 2 PCR Buffer in a total volume of 25 µl. A 1 min denaturation at 95°C was followed by 40 cycles of 30 sec denaturation at 95°C, 1 min annealing at 60°C and 4 min extension at 68°C. A final 6 min extension at 68°C completed the program. A long PCR fragment (about 4 kb) was cloned into pCR-XL-TOPO vector whereas shorter fragments were cloned into a pCR 2.1 vector.

TOPO Cloning

1.0 µl pCR-XL-TOPO vector and 4.0 µl gel-purified Long PCR product were incubated at room temperature for 30 minutes. After incubation, 1 µl of the 6 × TOPO cloning stop solution was added.

One Shot eletroporation

For the short PCR fragments (less than 4 kb), we used chemically competent cells (method same as section 2.1). For the long PCR products (longer than 4 kb), we performed the electroporation using *BIO-RAD* GENE PULSER.

a. Add 2 μl of TOPO cloning mixture to one tube of electrocompetent cells
b. Transfer the mixture of DNA and competent cells into an electroporation cuvette (0.1 cm)
c. Electrical conditions are 600 Ω, 10 μF and 1.8 kV for pulsing 10 sec.
d. Add 450 μl of SO.C medium and mix well
e. Transfer the solution to a 15 ml snap-cap tube and shake for 1 hour at 37°C
f. Spread on LB plate with 50 μg/ml kanamycin
g. Incubate the plate overnight at 37°C or 25°C until colonies appeared

Sequencing Analysis

Nucleotide sequences were determined using an ABI3130 Genetic Analyzer (Appplied Biosystems). Sequencing primers are shown in Table 3.2. Four plasmid clones were sequenced for genomic clones whereas six or seven clones were sequenced for cDNA clones. Most parts of the transposable element and the surrounding gene fragments were sequenced with BigDye terminator version 3.1 (Applied Biosystems) using a protocol recommended by the manufacturer. But, the 5′ and 3′ STR were sequenced with dGTP BigDye terminator version 3.0 (Applied Biosystems). Nucleotide sequences and amino acid translations were analyzed with the BLAST program (Altschul *et al.*, 1997). Survey of the genome sequence of a

soybean cultivar Williams 82 was performed with the soybean genome database (Phytozome, http://www.phytozome.net/soybean.php). Sequence alignment was performed with ClustalW (http://clustalw.ddbj.nig.ac.jp/index.php?lang=ja) using default settings. Gene prediction was performed with the GENSCAN software (http://genes.mit.edu/GENSCAN.html). Conserved domains in polypeptides were investigated with NCBI Conserved Domain Search (Marchler-Bauer *et al.*, 2011) at http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi.

Estimation of Transposon Footprint

The PCR amplicons of about 1 kb and 1.4 kb that were generated together with the 14 kb band in L72-2040 were deduced to be generated after excision of the transposable element. The PCR fragments were cloned into pCR2.1 vector. Nine clones each derived from the 1 kb band and 1.4 kb band were sequenced.

3.2 Results

PCR Amplicon Polymorphism

As expected, PCR of Clark generated a band of about 1 kb. However, L72-2040 produced three bands using upstream primers of Glyma09g36990 (Fig. 3.2). The three bands were about 14 kb, 1.4 kb and 1 kb. The results suggest that a transposable element of about 13 kb inserted in Glyma09g36990 of L72-2040. The two short bands were presumed to derive when the element excised from the gene of Glyma09g36990.

Gene Structure

Glyma09g36990 of Clark consisted of three exons and two introns in accordance with the previous report (Gillman *et al*., 2011) (Fig. 3.3). Compared with Clark, L72-2040 had an insertion of 13060 bp close to the second exon, 16 SNPs and three insertions (4 bp, 7 bp and 6 bp) close to the third exon. The fragment had CACTA motif at both ends and imperfect 30 bp TIR (Table 3.3). It had inverse repetition of short sequence motifs close to the 5′ and 3′ ends; 11 repeats in the 5′ region and 17 repeats in the 3′ region (Table 3.4). Furthermore, the fragment had a duplication of three nucleotides (ATG) at the site of integration that is characteristic of a CACTA-family transposable element (Fig. 3.4). These results indicated that the fragment belongs to a CACTA-family transposable element (Gierl *et al*., 1989). We designated the element as *Tgm11*.

Tgm11 had high overall nucleotide similarity of 93 % with *Tgs1*, an active non-autonomous element of about 3.8 kb that was inserted in the flavonoid 3′5′-hydroxylase gene of a *G. soja* accession having variegated flowers (Takahashi *et al*., 2012). Although differing in size, the regions close to the 5′ and 3′ ends were similar. *Tgm11* was also similar to *Tgm1*, an inactive non-autonomous element of about 3.5 kb inserted in the lectin gene with 88 % of overall similarity (Vodkin *et al*., 1983). Among the CACTA-superfamily transposable elements of soybean and and *G. soja*, the TIR of *Tgm11* had the highest similarity with *Tgs1* (96 % for 5′ TIR and 100 % for 3′ TIR) among CACTA-family TEs of soybean and *G. soja,* followed by *Tgm1* (96 % for 5′ TIR and 86 % for 3′ TIR) (Table 3.3). Motifs in the subterminal repeats of *Tgm11* were similar to *Tgs1* and *Tgm1* (Table 3.5).

GENSCAN and BLAST analyses suggested that *Tgm11* might have two ORFs (ORF1 and ORF2) similar to *Tgm9*, maize element *En*/*Spm* and *Antirrhinum majus* element *Tam1* (Masson *et al*., 1989; Nacken *et al*., 1991; Xu *et al*., 2010). RT-PCR using primer combination of P2 and P3 generated fragments of 4.4, ≈1.8 (two bands), ≈1.1 (two bands) and 0.8 kb. In addition, RT-PCR using primer combination of P1 and P3 generated a fragment of about 2.5 kb (Fig. 3.5 A).

Seven clones derived from 2.5 kb fragment consisted of two types generated by alternative splicing: one with 2404 bp long encoding 711 amino acids and the other was 2576 bp long encoding 680 amino acids (Fig. 3.5 B). The former and the latter transcripts were designated as *GmTNP3a* and *GmTNP3b*, respectively. Compared with *GmTNP3a*, *GmTNP3b* had a three base (one amino acid) deletion and retained the 8th intron resulting in premature termination of translation. *GmTNP3a* had amino acid identity of 43 % with *GmTNP1*, a soybean transposase encoded by *Tgm9* (Xu *et al*., 2010) (Fig. 3.6). GmTNP3a had a pfam0317 domain belonging to TNP1-like transposase 23 superfamily similar to GmTNP1 (Xu *et al*., 2010) (Fig. 3.6). The cDNA sequence of *GmTNP3a* was quite similar to the corresponding genome sequence of *Tgm11* (> 99 %). However, two consecutive nucleotides were deleted and seven nucleotides were substituted resulting in four amino acid alterations compared with *Tgm11*.

Seven clones derived from 4.4 kb fragment consisted of five types with variable lengths in 3′ untranslated region (total of 4393, 4465, 4486, 4583 or 4661 bp) (Fig. 3.5 B). All these variants encoded a polypeptide consisting of 954 amino acids, and were classified into two groups designated as *GmTNP4a* and *GmTNP4b*

having four amino acid alterations (Fig. 3.7). *GmTNP4a* had amino acid identity of 66 % with *GmTNP2*, a transposase encoded by *Tgm9* (Xu *et al*., 2010) (Fig. 3.7). GmTNP4a and GmTNP4b had a pfam0292 domain belonging to TNP2-like transposase 21 superfamily similar to GmTNP2 (Xu *et al*., 2010) (Fig. 3.7). The polypeptides of *GmTNP4a* and *GmTNP4b* had overall similarity with *GmTNP2*, except for a deletion consisting of 108 consecutive amino acids in the upstream of pfam0292 domain in *GmTNP4a*. The cDNA sequences of *GmTNP4a* and *GmTNP4b* were quite similar to the corresponding region of *Tgm11* (> 99 %). However, four nucleotides (two amino acids) and eight nucleotides (six amino acids) were substituted in *GmTNP4a* and *GmTNP4b*, respectively.

Seven clones derived from the ≈1.1 kb fragments and six clones from the 0.8 kb fragment were sequenced. Clones derived from the ≈ 1.1 kb bands were classified into three types (941, 1037 or 1212 bp) encoding 227 amino acids (Fig. 3.5 B). Clones derived from 0.8 kb bands were 733bp encoding 103 amino acids. All these short fragments might be derived by alternative splicing of *GmTNP4a* or *GmTNP4b* (Fig. 3.5 B).

Eight clones derived from ≈ 1.8 kb bands were classified into 2 types (1666 or 2030 bp). The nucleotide sequence of these transcripts had many substitutions compared with *Tgm11*. Moreover, the transcripts had stop codons in all reading frames. They were presumed to derive from other region in the genome.

Estimation of Transposon Footprint

Sequencing from PCR bands of 1 kb and 1.4 kb revealed that these PCR

products correspond to fragments of Glyma09g36990 around the site of *Tgm11* integration (Fig. 3.4). Six clones derived from the 1.4 kb fragment (clones 1-1 to 1-6) and all nine clones (clones 2-2 to 2-10) from the 1 kb fragment had similar lengths (1,015 to 1,035 bp). Among these 15 clones, six had the sequence identical with Clark, suggesting that a precise excision without a footprint might have occurred. Three clones were devoid of several bases including the duplicated site ATG that found in Clark. The other six clones had extra nucleotides compared with Clark. In contrast, three clones derived from the 1.4 kb fragment were 1403, 1443 and 1445 bp long. They retained about 140 bp of 5' end and about 270 bp of 3' end of *Tgm11*. *Tgm11* may have been excised from the middle of STR region leaving behind part of fragments in the three clones.

Accession numbers

Sequence data was deposited in the DDBJ Data Libraries under accession numbers AB934275 (*Tgm11*), AB934276 (*GmTNP3a*), AB934277 (*GmTNP3b*), AB934278 (*GmTNP4a*), AB934279 (*GmTNP4b*), AB934280 (genome sequence of Glyma09g36990 from Clark) and AB934281 (genome sequence of Glyma09g36990 from L72-2040).

3.3 Discussion

R locus responsible for seed coat color consists of multiple alleles of *R* (black), *r* (brown) and *r-m* (black spots and/or concentric streaks on brown seed) (Bernard and Singh, 1969). Genetic analysis, genome sequence and gene expression assay

suggested that a MYB transcription factor gene, Glyma09g36990 might be the responsible for brown seed coat (Gillman *et al*., 2011). In our study, a DNA fragment of about 13 kb was inserted in the second intron of Glyma09g36990 in a NIL having *r-m* allele (Fig. 3.3). The fragment had the characteristics of a CACTA-superfamily transposable element: CACTA motif at both ends, imperfect 30-bp TIR, inverse repetition of short sequence motifs around the 5′ and 3′ ends, and duplication of three nucleotides at the site of integration, indicating that it belongs to a CACTA-family transposable element. We designated the element as *Tgm11*. A transposable element *TgmR** of 13,021 bp long was recently cloned from the same isoline, and it may be essentially same with *Tgm11* (Zabala and Vodkin, 2014).

CACTA elements of soybean were classified into two classes based on overall nucleotide sequence, 5' and 3' TIR and STR motifs. One class consists of *Tgm1* and *Tgs1* while the other class includes *Tgm9*, *Tgmt** and *Tgm-EXPRESS 1* (Takahashi, *et al*., 2012). The present study revealed that *Tgm11* was also similar to *Tgm1* and *Tgs1*, based on its nucleotide sequence, 5′ and 3′ TIR and motif of STR. According to the 80-80-80 rule (Wicker *et al*., 2007), these three elements may comprise a family. We designated the family as *Tgm1* family. Among the family, *Tgm11* was integrated into an intron whereas *Tgm1* and *Tgs1* were inserted into exons.

RT-PCR suggested that *Tgm11* had two ORFs (ORF1 and ORF2) similar to other plant CACTA-family elements. RT-PCR produced various transcripts including *GmTNP3* (*GmTNP3a* and *GmTNP3b*) and *GmTNP4* (*GmTNP4a* and *GmTNP4b*) probably generated by alternative splicing. GmTNP3a had amino acid identity of 43% with GmTNP1 expressed by soybean active autonomous element *Tgm9*. GmTNP4a

and GmTNP4b had amino acid identity of 66% with GmTNP2 that was similar to both TNP2 encoded by *Tam1* and TNPD encoded by *En*/*Spm* (Masson *et al*., 1989; Nacken *et al*., 1991; Xu *et al*., 2010). We postulate that the excision mechanism of *Tgm11* may be similar as element *En/Spm* (Masson *et al.*, 1989; Nacken *et al*., 1991; Gierl *et al*., 1988; Gierl *et al*., 1989; Frey *et al*., 1990). After alternative splicing, *Tgm11* may produces two transposases GmTNP3 and GmTNP4. GmTNP3 may be a kind of DNA-binding protein, recognizing certain repeats and binding to the both 3' and 5' STRs of *Tgm11* (Gierl *et al*., 1988; Grant, *et al*., 1990) (Fig. 3.8). Whereas GmTNP4 acting like TNPD, may binds to GmTNP3, interacts with the TIRs of *Tgm11* and pulls the ends together for transposition (Gierl *et al*., 1989; Frey *et al*., 1990) (Fig. 3.8). Recipient site was postulated to occur either within or at some distance of the GmTNP3-binding domain. Moreover, GmTNP4 may excise the element *Tgm11* from gene Glyma09g36990 acting as an endonuclease.

The cDNA sequence of *GmTNP3a* and *GmTNP4a* was very similar (> 99%) to the corresponding regions of *Tgm11*, but some nucleotides were substituted resulting in amino acid alterations. Transposase genes of *Tgm11* may be silenced similar to the *Tgmt** (Zabala and Vodkin, 2008) by DNA methylation, a defense mechanism that deactivate possibly harmful mobile elements (Lisch, 2009). GmTNP3 and GmTNP4 proteins may be produced by other element(s) having high similarities to *Tgm11* existing elsewhere in the genome.

Wicker *et al*. (2007) defines a transposable element as autonomous if it appears to encode all the domains that are typically necessary for its transposition without implying that the element is either functional or active. Based on the

definition *Tgm11* might be an active autonomous element, though transposase genes might be silenced. Active Non-autonomous element *Tgs1* and *Tgm1* might be deletion derivatives of autonomous elements such as *Tgm11*. The above results are consistent with the presumption of Gillman *et al*. (2011) that loss of function of Glyma09g36990 might be responsible for brown seed coat color. Reverse genetic experiments should be conducted to prove that Glyma09g36990 corresponds to the *R* gene.

Table 3.1 List of identified full-length transposable elements of soybean and *Glycine soja*

Name	Length	Insertion gene	Citation
Tgs1	3.9 kb	first exon of *W1* gene	Takahashi, *et al.*, 2012
Tgm1	3.5 kb	exon of a *lectin* gene	Vodkin, *et al.*, 1983
Tgm9	20.5 kb	second intron of *W4* gene	Xu M, *et al.*, 2010
*Tgmt**	20.5 kb	first intron of *T* gene	Zabala and Vodkin, 2008
Tgm-EXPRESS 1	5.7 kb	second intron of *Wp* gene	Zabala and Vodkin, 2005

Table 3.2 PCR primers used for sequencing

Target fragment for sequencing	Forward primer (5′-3′)	Reverse primer (5′-3′)
Glyma09g36990 of Clark	TGATCAGATTGCACAAGCTTTTGG TGACAGAGGTGAAAATCATA ACATGGACAGCATAAAGGTT TAAGATAATGTTCTCACAAGA	ATCAGGGACCATCTACATCGACA TCTACTGTCATGACCTTCAT TGGCCAATATACAAACCTGT
Glyma09g36990 of L72-2040 including transposable element (*Tgm11*)	TGATCAGATTGCACAAGCTTTTGG GTACCGATGTGGAAAGTAGT AAATGTTGAGGTAGGTGACA ATCAGTGAAATACTGCATGA AGTGGAGACCAATATATGTT AGCTCCCTGTTTATCCAACA GTTTGATGGTTATACCGGCT ATTAGTTACCTCCCATGAAG TAGCCATTTGGCTGTGAAAA GGTCTTGTGCTTTTGTATAG CCATGATGTCAAAGAATGCA CCATCAGACCTTGGATGTAA AGTCCAAGCCTAAGATTTCT CGATCTCCCATTTCTACATC CCTTAACACATGAATCTACC AATGTGTCACAAATGAGCCA TGCGTATGGATTCTTACGAA TCATCTGTCATCGACATCCA GTTCAATAAGAGAGCCAAAC AGGTGGTGTTTTGTTTGATTT CCATGGAGAGAGTTAGAAAT GGTCTCAATGGTGTGAAACT ACGATTGCAGTGAAACTGTT GTAGCATCCACTATTTACTC TACGTCCATTTGTATGTTGG ATAATACAGCTTGGGGTGTA	TGTTTTGTTGTTCTTCTTTG GTCAAAGATCGATCTACTGC TCAGCCTAACATAAAATGAA ACCCTAAAATTTGAATCACT GACATTCAGTAGTAAAGAAG GATGAGTCTTGATTTGAGTG TGTGGCGGAGGACAGGGTT ATGAAAATACTGCATTCTAAGA TTCGTAAGAATCCATACGCA TTCTGAGCGCTTGTTATACT GTGAGATTTTCATGCATCGA TGCCCCTATGTATGATACAT GTTGTCTATGATGATAACGG GAATACTCGTCAAGATCTAG AACAAGCTCACTTGTATGTG TGGTCAGTGGTTTTACATGG TTCCATATTGTTTGTTAGCC AGAAAACTGGTGACATGACA TTCAGCTGCTCGTGTCAGCA AATTGCACATGGTGGTCAAT ATGGTCGTCATTTTGCCTAA TTCACTGATCCTAGACCATT TTATATGTTGTGTGGTCTGAT GCAGGTTATTTTTGGGATTTA CGTATCATATAACATCGCTT TGGCCAATATACAAACCTGT
Transposase gene of *GmTNP3* and *GmTNP4*	AGTTGACATCCAATGCATGA TCAAGTCCTTTCAAAGCACT GACCATTTGCCAATGTGCA TAACCTGCAATGGCTTCAGA ACAACCCACATTTGTATGAG ATTGAGTTGCTTCCTCTAGT AGATGCTTCAGGGATATCAA ATGTTACACCAGTAGAAGGA ACCTCTTACTCTTTCGTCAT TAACCGCTCAAATTCCCATA CTTGAGTTTGTACCGTGATA	GGAGTATCAGAAGATTCATG AGGTGTTGTGGTATCTTCC GATTTTCCAGCATATGGGAA TGCATTCTTTGACATCATGG CAAGACTTTCCTAGATTGGT CGAAGACCCTTGTGATCAA TGATATGGGAGGATATTCAG TAGAGACCCTTCATGGGAG

Table 3.3 Comparison of 5′ and 3′ terminal inverted repeats of *Tgm11* in soybean line L72-2040 with CACTA-family transposable elements of soybean and *Glycine soja*

Transposon	5′ terminal inverted repeat (5′-3′)	Identity to *Tgm11*	3′ terminal inverted repeat (5′-3′)	Identity to *Tgm11*
Tgm11	CACTATTACAAAAAGTAGTTTCAACATCGG	100%	CACTACTAGAAAATAAGGTTTTAACATCGG	100%
Tgs1	CACTATTACAAAAAGTAGTTTTAACATCGG	96%	CACTACTAGAAAATAAGGTTTTAACATCGG	100%
Tgm1†	CACTATTACAAAAAGTAGTTTTAACATCGG	96%	CACTATTAGAAAATATGTTTTTTACATCGG	86%
Tgm6	CACTATTAGAAAATATGTTTTCTACATCGA	76%	CACTACTACAAAAAGCAGTTTTAACATCGA	86%
Tgm9	CACTACTAGAATAATGTTTTTTTTACGACGT	66%	CACTACTACAAATAAAGCTTTTTAAGTCGG	76%
Tgm-express 1	CACTACTAAAAAAATCTGTTTTTTACGACGC	66%	CACTACTACAAAAGAGGTTTTTTAAGTCGG	76%
*Tgmt**	CACTACTAGAATAATGTTTTTTTTACGACGT	66%	CACTACTACAAATAAAGCTTTTTAAGTCGG	76%

† 5’ and 3’ terminal inverted repeats were interchanged

Table 3.4 Subterminal repeats of *Tgm11* in soybean line L72-2040

Location	Subterminal repeats (5′-3′)	Nucleotide position (bp)
5′ subterminal region	TTAACATCGGTTTTTTCAAAACCGATGTTAA	35-66
	TTAACATCGGTTTTCCAAAAAACCGATGTTAA	106-137
	TTAACATCGGTTTTTTGGAAAACCGATGTTAA	147-178
	TTAACAGCGGTTTTTCAAAAACCGATGTTAA	188-218
	TTAACATCGGTTTTCCAAAAAACCGATGTTAA	230-261
	TTAACATCGGTTTTTCAAAAACTGATGTTAA	272-302
	TTAACATCGGTTTTCCAAAAACCGATGTTAA	314-344
	TTAACATCGGTTTTTTGGAAAATCGATGTTAA	355-386
	TTAACATCGGTTTTTTAAAAAACTGATGTTGTAA	396-429
	TTAACATCGGTTTTTTAAAAAACTGATGTTGTAA	437-470
	TTAACATCGGTTTTTTAAAAAACTGATGTTGTAA	478-511
Consensus sequence	TTAACATCGGTTTT	
3′ subterminal region	AACATCGGTCTTTAAAAAAACTGATGTG	12148-12175
	AACATCAGTTTTTTTAAAAACTGATGTT	12189-12216
	TTAACATCGGTTTTTAAAAAAACTGATGTTGT‡	12241-12272
	CACATCAGTTTTTAAAAAAACTGATGTG	12284-12311
	AACATCAGTTTTTTAAAAACCGATGTC	12325-12351
	TTAACATCGGTTATTTTAAAAACCGATGTTAC‡	12363-12394
	AACATCAGTTTTTTAAAAATCGATGTT	12406-12432
	AACATCGCCTATTTAAATAACCGATGTT	12479-12506
	AACATCAGATTTTTAAAAAACTGATGTG	12520-12547
	AACATCAGTTTTTAAAAAACCGATGTC	12561-12587
	AAAATCAGTTTTTTAAAAAACCGATGTT	12642-12669
	AACATCGGTTTTTTAAAAAACCGATGTT	12702-12729
	AACATCAGTTTTTTAAAAAACCGATGTT	12743-12770
	AACATCGGTTTTTAAAAAACCGATGTT	12784-12810
	AACATCAGTTTTTTAAAAAACCGATGTT	12824-12851
	AACATCGCTTATTTAAATAACCGATGTT	12896-12923
	TTAACATCGGTTTTGAAAAACCGATGTTAA‡	12935-12964
Consensus sequence	AACATCAGTTTTT	

‡Similarity was estimated with the motif of 5′STR

Table 3.5 Alignment of subterminal repeat motifs among CACTA family transposable elements in soybean and *Glycine soja*

Element	Subterminal repeat motif (5′-3′)
Tgm11 (5′)	TTAACATCGGTTTT
Tgm11 (3′)	AACATCAGTTTTT
Tgs1 (5′)	TTAACATCGGTTTTT
Tgs1 (3′)	AACATCAGTTTTTT
Tgm1	TTAACATCGGTT
Tgm9	TCTAAGACGGTT
*Tgmt**	TCTAAGACGGTT
Tgm-EXPRESS 1	TCTAAGACGGTT

Clark
(*i-ii-i RR*)

L72-2040
(*ii r-mr-m*)

L67-3469
(*ii RR*)

L66-17
(*ii rr*)

Figure 3.1

Seeds of soybean cultivar Clark, L72-2040 and Clark mutant lines (L67-3469 and L66-17) having different alleles at *R* locus

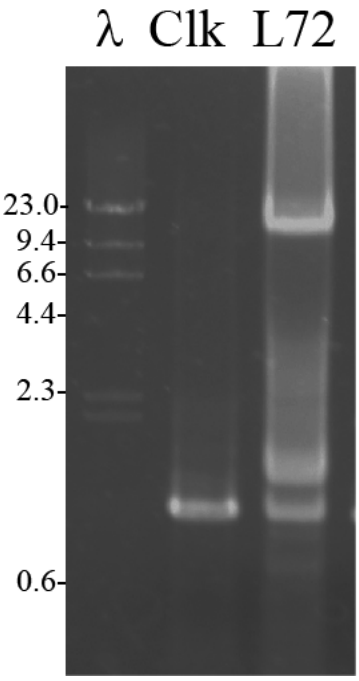

Figure 3.2

PCR products generated from Clark and L72-2040

The migration of the size marker (kb) is shown to the left of the gel. **λ**: molecular markerλ/*Hind* III; **Clk**: Clark; **L72**: L72-2040

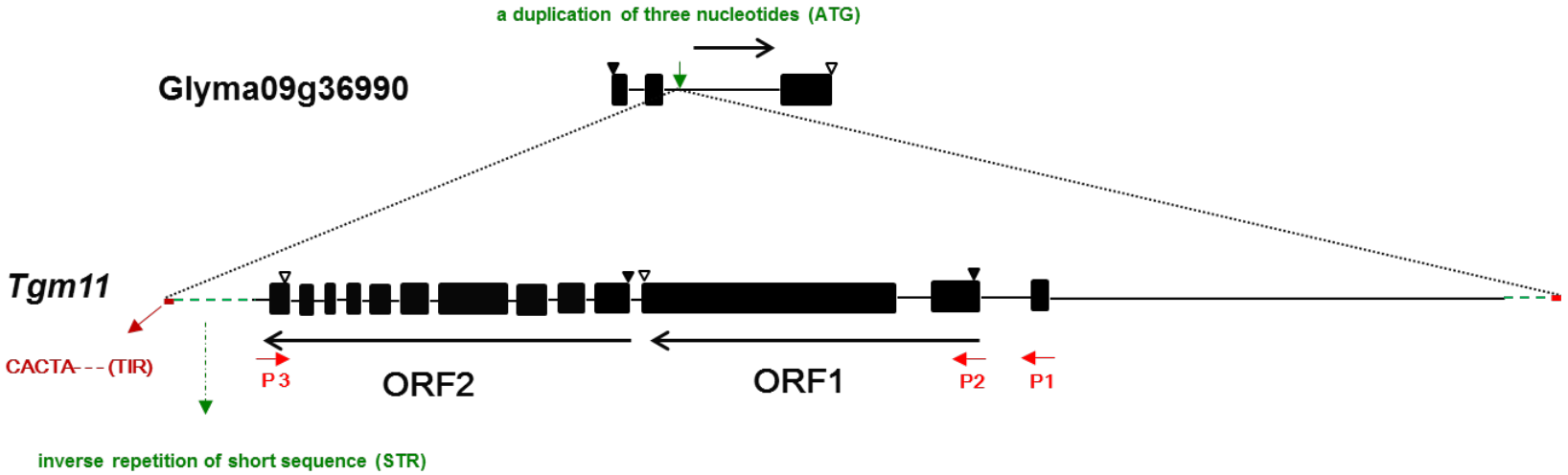

Figure 3.3

Gene structure of a soybean transposable element *Tgm11* inserted into MYB transcription factor gene Glyma09g36990

Exons and introns are indicated by black boxes and solid black lines, respectively. Start codon and stop codon are shown by black and white triangles, respectively. Orientation of open reading frames is indicated by black arrows. P1, P2 and P3 were primers for cloning of transposases.

Figure 3.4

Schematic presentation of integration of *Tgm11* into Glyma09g36990 and the footprints left behind by its excision from the gene

Tgm11 insertion is shown by a white triangle. DNA sequences around the site of the integration from Clark, L72-2040 and the corresponding regions of six DNA clones (clones 1-1 to 1-6) derived from PCR band of 1.4 kb and nine clones (clones 2-2 to 2-10) derived from PCR band of 1 kb were aligned. Footprint nucleotides added by the *Tgm11* excision are shown in bold. Nucleotides representing the target site duplication are shown in red.

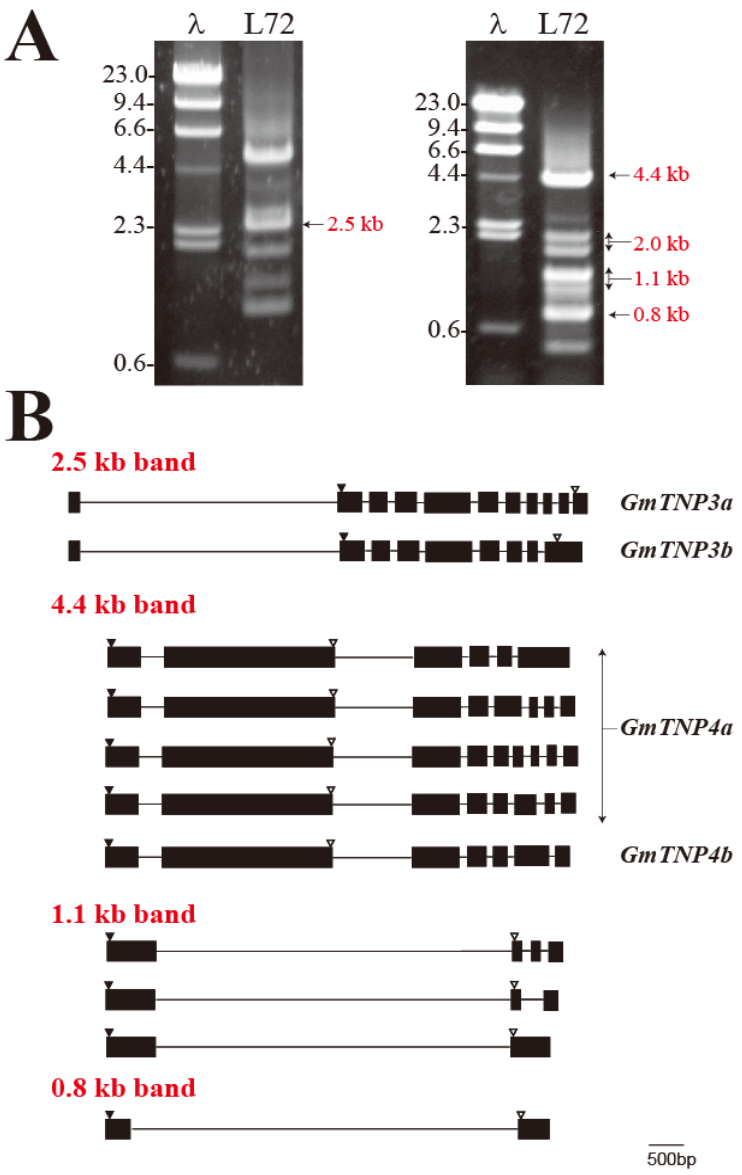

Figure 3.5

Expression and alternative splicing of transposase gene in soybean transposable element *Tgm11*

(A) Expression of transposase genes detected in L72-2040 by RT-PCR using different set of PCR primers. λ: molecular marker λ/*Hind* III; **L72**: L72-2040. The migration of the size marker (kb) is shown to the left of the gels. The left panel was obtained by PCR primers P1 and P3 whereas the right panel was obtained by primers P2 and P3.

(B) Alternative splicing of transposase genes. Exons and introns are indicated by black boxes and solid black lines, respectively. Start codon and stop codon are shown by black and white triangles, respectively.

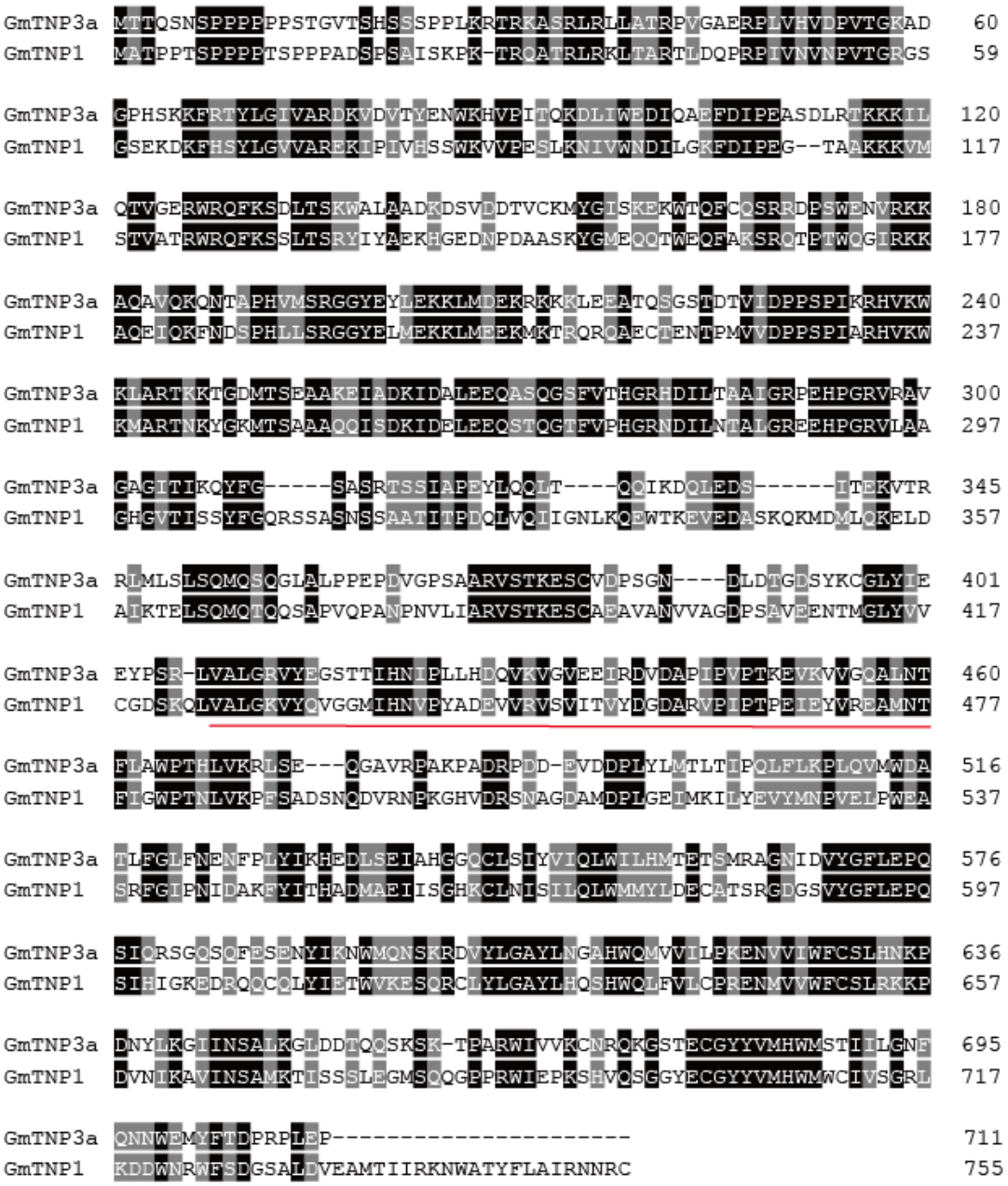

Figure 3.6

Amino acid alignment of soybean transposase, *GmTNP3a* and *GmTNP1* expressed in plants carrying transposable elements *Tgm11* and *Tgm9*, respectively

Identical amino acids are highlighted in black. Similar amino acids with conserved substitutions are highlighted in gray. Dashes represent gaps introduced to improve the alignment. The motif for pfam0317 is red underlined.

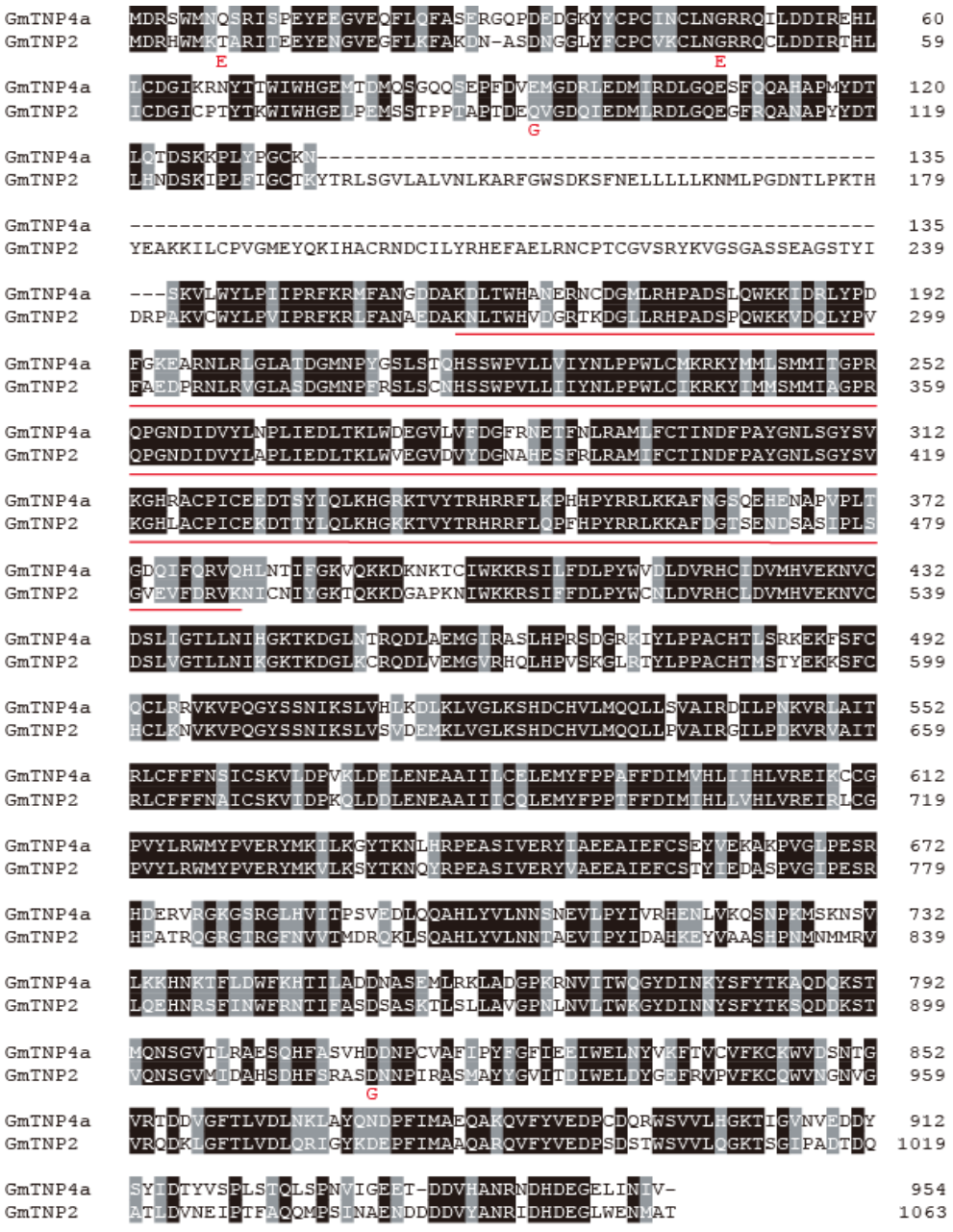

Figure 3.7

Amino acid alignment of soybean transposase, *GmTNP4a* and *GmTNP2* expressed in plants carrying transposable elements *Tgm11* and *Tgm9*, respectively

Identical amino acids are highlighted in black. Similar amino acids with conserved substitutions are highlighted in gray. Dashes represent gaps introduced to improve the alignment. The motif for pfam0292 is red underlined. Amino acids polymorphic in *GmTNP4b* are shown in red font under alignment.

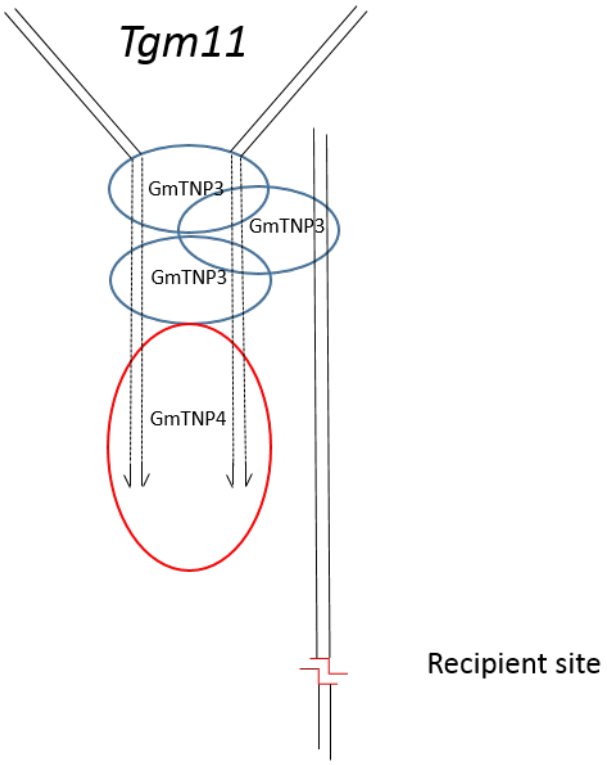

Figure 3.8

A hypothetical model of *Tgm11* transposition

STRs of *Tgm11* are shown by dotted line and TIR was in solid line with arrow. Protein of GmTNP3 and GmTNP4 are indicated by blue and red circle, respectively. Staggered cleavage at the recipient site is in red cutting edge.

CHAPTER 4 Analysis of flavonoids in sprout of soybean near-isogenic lines and mutants

Background

Flavonoids representing a large class of secondary metabolites gain a lot of attentions these years, because of their antioxidant properties, including chalcones, flavones, flavonols, flavonoids, anthocyanins and condensed tannins (Brenda, 2001). Some plants also synthesize special forms such as isoflavones in legume plants and stilbenes in red wine. Isoflavonoids were one of important compounds almost exclusively existing in soybean beneficial to human cancer (Dellapenna, 1999). The consumption of soy-based food has been increased worldwide these years, because of abundance of isoflavonoids in soybean seed (Winkel-Shirley, 2001; Phommalth *et al*., 2008). Anthocyanins, as antioxidants, were identified to prevent vascular diseases, protect vitamins and increase the serum antioxidant capacity in fresh food (Murkovic *et al*., 2000; Youdim *et al*., 2000; Kong *et al*., 2003; Prior, 2003).

Flavonoids comprised of a multiple family deriving from Coumaroyl-CoA and Malonyl-CoA. In the flavonoid biosynthetic pathway, there are two kinds of genes, structural and regulatory genes affecting the pigmentation of flower, seed coat or pubescence in soybean. The *W1* gene encoding flavonoid 3'5'-hydroxylase has pleotropic effects on flower color, seed coat color and hypocotyl color. Generally, soybean cultivars having purple or white flowers have purple or green hypocotyls (Takahashi and Fukuyama, 1919; Groose and Palmer, 1991). *W1* gene and *W4* gene affect flower colors and hypocotyl colors. Whereas, *T* gene affects pubescence colors

and seed coat colors. Furthermore, the genes in the biosynthetic pathway affect each other, because of competition of substrates (Fig. 1.2). Therefore, mutations will influence flavonoid components in various tissues.

Nozzolillo examined eighty-eight species of legume seedlings for anthocyanins (Nozzolillo, 1973). Cyanidin glycosides was the most common component in 44 pigmented-species and malvidin glucoside was predominant in red pigmentation seedlings (Nozzolillo, 1973). Later, another report indicated that delphindin, petunidin and malvidin were responsible to the visible color in the hypocotyls. Moreover, compared to the concentration of delphinidin, malvidin was 40 to 60-fold and petunidin was 4-fold higher, respectively (Peters *et al*., 1984).

Seedlings growth in continuous dark environment affects both the levels and distribution of the isoflavones (Terrence, 1991). Under darkness, the concentration of daidzein, genistein and their derivatives are higher in cotyledons. But the level in root tip and hypocotyl are reduced dramatically (Terrence, 1991). Cotyledon containing less isoflavones compared to hypocotyl (Eldridge *et al*., 1983; Murphy *et al*., 2002; Ribeiro *et al*., 2006; Yuan *et al*., 2009). Hypocotyl of soybean containing high amounts and concentration of isoflavones was considered to be the most valuable resource of soybean (Eldridge *et al*., 1983; Tsangalis *et al*., 2004; Kim *et al*., 2007; Yuan *et al*., 2009). However, isoflavones cannot be detected or at dramatically low level in seed coat (Eldridge *et al*., 1983; Ribeiro *et al*., 2006).

We expect to find some unique components beneficial to human health and understand the relationship between components and genes in near-isogenic lines and mutants of anthocyanin biosynthetic pathway. Twenty-three soybean lines

including near-isogenic lines and mutants were planted, extracted and analyzed by HPLC analysis of flavonoid.

4.1 Materials and Methods

Plant material

Twenty-two soybean lines and one *G. soja* accession kw4 including NILs and mutants with different flower color or seed coat were used for flavonoid analysis (Table 4.1).

Preparation of Seedlings

Seeds of the NILs and the mutants shown in Table 4.1 were planted in a plastic tray filled with vermiculate. We conducted two treatments with same materials in a phytotron for light treatment (14 h light/10 h dark, 28°C) and under continuous darkness in an incubator at 28°C. After five days, hypocotyls about 5 cm long were cut and kept in an ice container. Experiments were repeated in 3 replications.

Extraction of flavonoids

Hypocotyl and cotyledon were separated and cut into small pieces using razor blades. A total of 0.5 g hypocotyl or cotyledon pieces were collected and soaked in 5ml MeOH containing 0.1% (v/v) HCl, respectively. Samples were mixed well and kept at 4°C for 48 h. Extracts were filtered with disposable filtration units (Maishoridisc H-13-5, Tosoh) and kept at -20°C.

HPLC analysis

Quantitative analysis of the extracts was performed with the Agilent 1100 HPLC System (Agilent Technologies) using Inertsil ODS4, 5 μm 6.0×150 mm (GL Science Inc.) for anthocyanin analysis. L-column 2ODS [I.D.6.0×150 mm (Chemicals Evaluation and Research Institute)] was used for flavonoid analysis. Injection volume was 10 μl.

Anthocyanin was separated with phosphoric acid/acetic acid/acetonitrile/water (3:8:6:83) at a flow rate of 1.0 ml/min for 40min. For flavonol, dihydroflavonol and isoflavone two solvents were used. Solvent A: 0.2% phosphoric acid; Solvent B: phosphoric acid/acetonitrile/water (0.2:30:70) and the gradient was 0 min, B 20%; 18 min, B 50%; 25 min, B 70%; 31 min, 50%; 34 min, B 20% at a flow rate of 1.0 ml/min for 40 min. Detection wavelength of anthocyanin was at 530 nm, flavonol at 350 nm, dihydroflavonol at 280 nm and isoflavone at 260 nm.

4.2 Results

Phenotype of seedlings

Seedlings grown under light had green cotyledons and different pigments of hypocotyls (Fig. 4.1 A). Seedlings grown in dark had yellow cotyledons and white hypocotyls (Fig. 4.1 B). As expect, plants having purple, purple-blue or magenta flower colors had purple hypocotyls under light condition. The lines having light purple or pale flower colors had light purple hypocotyls. The plants having white or near white flower colors had green hypocotyls. However, Clark-*w1* was differed

from these white-flowered and had bronze hypocotyl (Fig. 4.1 A).

HPLC Analysis of Anthocyanin

In HPLC analysis of anthocyannins, a dominant peak A1 was observed around 17 min in addition to two weak peaks in purple hypocotyls (Fig. 4.2). Light purple hypocotyl had a relatively weak peak of A1 compared with purple hypocotyl (Fig. 4.2). No major peaks were detected in green hypocotyls (Fig. 4.2). However, Clark-*w1* differed from these lines producing a unique peak of A2 around 13 min (Fig. 4.2). A1 and A2 were identified, based on comparison with authentic specimens, as malvidin 3, 5-di-*O*-glucoside and peonidin 3, 5-di-*O*-glucoside, respectively. In cotyledons and hypocotyls grown in dark, there was no specific peak consistent with observations.

These results suggested that anthocyanin could only be detected in hypocotyls under light environment. Malvidin 3, 5-di-*O*-glucoside was the most predominant components in purple hypocotyls. Peonidin 3, 5-di-*O*-glucoside was responsible for the bronze color of hypocotyl in Clark-*w1*.

HPLC Analysis of Isoflavone

There was no significant difference between dark and light environment or cotyledon and hypocotyl. One predominant peak appeared at 29.8 min in all test samples. The concentration of 29.8-min peak was dramatically high in hypocotyl, about 90 mAU under dark and about 100 mAU under light (Fig. 4.3).

Hypocotyls of cultivar Bay and Bay mutants had two unique peaks at 22.8

and 24.0 min under light environment, which were absent or at low level in other samples (Fig. 4.3 Bay-H-L).

HPLC Analysis of Dihydroflavonol

In the HPLC analysis of dihydroflavonol, there are two dominant peaks at 11.4 min and at 29.8 min. Hypocotyls of Harosoy produced a special peak at 14.7 min under light condition (Fig. 4.4 Harosoy-H-L). The 14.7 min peak was also observed in other lines and mutants except Clark-*sb*, Clark-*w1*, Clark-*w4*, Clark-*W3w4*, kw4, RIL138, Bay-*w4*, Bay-*w4-lp* and NIL-*wp*. The results suggested that component of 14.7 min peak might be related to purple hypocotyl. It is necessary to identify the composition.

However, some other peaks were detected in the cotyledon of kw4, which were absent in other lines. Two peaks at 8.5 min and at 15.9 min were existed under dark and another peak at 6.1 min under light (Fig. 4.4 kw4-C-D and kw4-C-L).

HPLC Analysis of Flavonol Glycoside

There was no major peak in hypocotyl and cotyledon under dark or light environment (Fig. 4.5).

4.3 Discussion

In cotyledons, green hypocotyls and hypocotyls grown in dark, there was no specific peak corresponding to anthocyanin in consistent with observations. Under

light environment, seedlings have purple, light purple, green and bronze hypocotyl colors in accordance with previous study (Groose and Palmer, 1991). The phenotypic results suggested that pigmentation depends on the interaction of pigmentation genes. Under *W1* genotype, soybean genotype with *W3W4* has dark purple flowers and purple hypocotyls, *W3w4* has dilute purple or purple throat flowers and light purple hypocotyls, *w3W4* has purple flowers and purple hypocotyls, and *w3w4* has near white flowers and green hypocotyls. Under *w1* genotype, soybean has tawny pubescence colors (*T*-) having bronze hypocotyls and gray pubescence colors (*tt*) having green hypocotyls.

Malvidin 3, 5-di-*O*-glucoside was the most abundant components existing in purpled hypocotyl in agreement with a previous report (Nozzolillo, 1973). Based on phenotype and HPLC results, we postulate component of peonidin 3, 5-di-*O*-glucoside was responsible for bronze hypocotyl in Clark-*w1*. Anthocyanins have many effects for human health such as preventing vascular diseases and protecting vitamins. So, hypocotyls having pigments were considered to be more valuable.

HPLC analysis of isoflavone, dihydroflavonol and flavonol glycoside showed some unique peaks in specific lines. Identification of each peak components is under way. We expect we could find some unique components in mutants or NILs. Though components were unidentified, structure and abundance of flavonoids differed among hypocotyl and cotyledon or under light and dark environment. Several reports suggested that hypocotyl was the most valuable tissue of soybean as containing variety types and high concentration of isoflavone (Eldridge *et al*., 1983; Kim *et al*., 2007; Tsangalis *et al*., 2004; Yuan *et al*., 2009). In our study,

hypocotyl contained more peaks corresponding to flavonoid components. Qualitative and quantitative analysis of these peaks will be necessary to identify components beneficial to human health.

Table 4.1 Plant materials of near-isogenic lines and mutants

NO.	Cultivar		Flower Color	Hypocotyl Color	Seed Coat Color
1	Harosoy	-	Purple	Purple	Yellow
2	Harosoy-*w2*	-	Purple-blue	Purple	Yellow
3	Harosoy-*wm*	T235	Magenta	Purple	Yellow
4	Clark	-	Purple	Purple	Yellow
5	Clark-*sb*	-	Slight blue	Green	Yellow
6	Clark-*i*	L67-3469	Purple	Purple	Black
7	Clark-*ir*	L66-17	Purple	Purple	Brown
8	Clark-*irm*	L72-2040	Purple	Purple	Brown with black stripe or spot
9	Clark-*w1*	L63-2373	White	Bronze	Yellow
10	Clark-*w4*	L68-1774	Near white	Green	Yellow
11	Clark-*W3w4*	L70-4422	Dilute purple	Light purple	Yellow
12	RIL138	-	White	Green	Yellow
13	T369	-	Pale	Light purple	Yellow
14	T321	-	Dilute purple	Light purple	Yellow
15	kw4	-	Near white	Green	Black
16	Bay	-	Purple	Purple	Yellow
17	Bay-*w4*	222-A-3	Near white	Green	Yellow
18	Bay-*S10*	E-023-H-12	Light purple	Light purple	Yellow
19	Bay-*S20*	E-013-C-1	Dark purple	Purple	Yellow
20	Bay-*w4-lp*	E-30-D-1	Light purple	Light purple	Yellow
21	Tachinagaha	-	Purple	Purple	Yellow
22	NIL-*Wp*	-	Purple	Purple	Yellow
23	NIL-*wp*	LD05-15019	Pink	Light purple	Yellow

Figure 4.1

Hypocotyl color of light-grown seedlings

A: under light environment; B: under dark. From left to right (Clark, T369, RIL138 and Clark-*w1*).

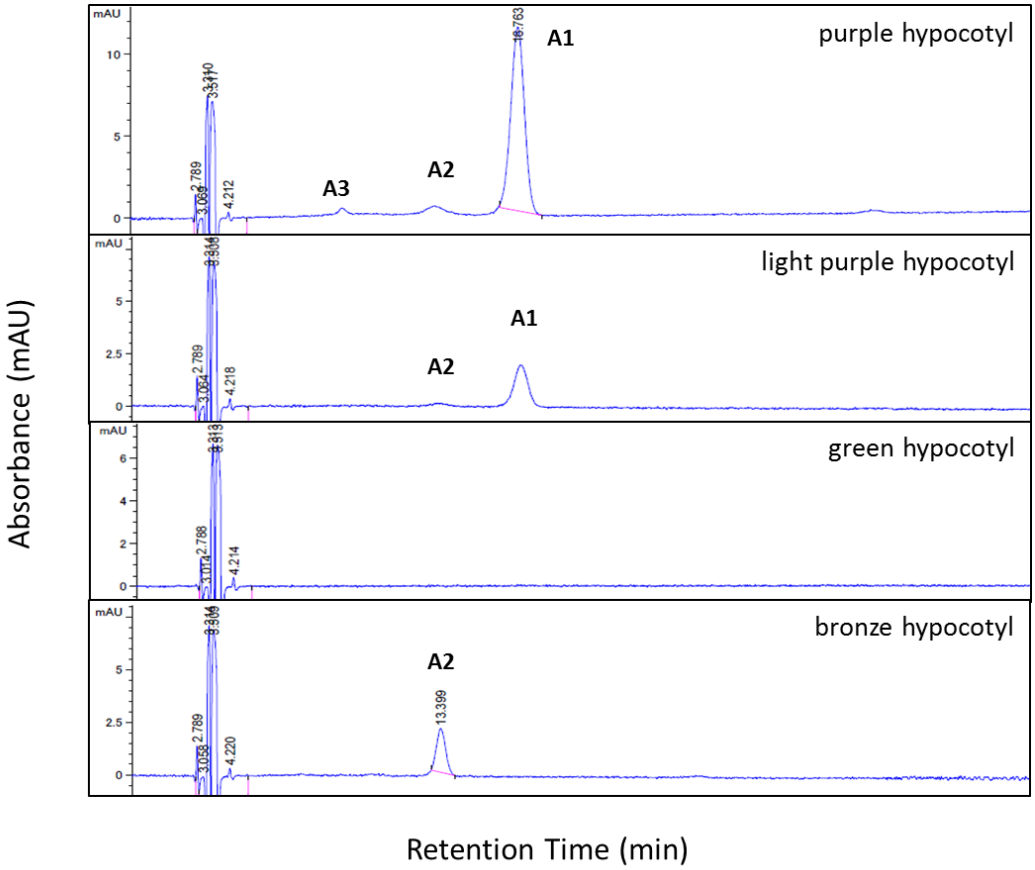

Figure 4.2

HPLC chromatogram of anthocyanins extracted from hypocotyls grown under light conditions having various colors

Total of 500 mg of hypocotyls was extracted with 5 ml of MeOH containing 0.1% (v/v) HCl. Eluents: Phosphoric acid/ Acetic acid / Acetonitrile / water (3:8:6:83) for 40min. Flow-rate: 1.0 ml/min. Injection: 10 µl. Detection wavelength: 530 nm. A1: malvidin 3, 5-di-*O*-glucoside, A2: peonidin 3, 5-di-*O*-glucoside.

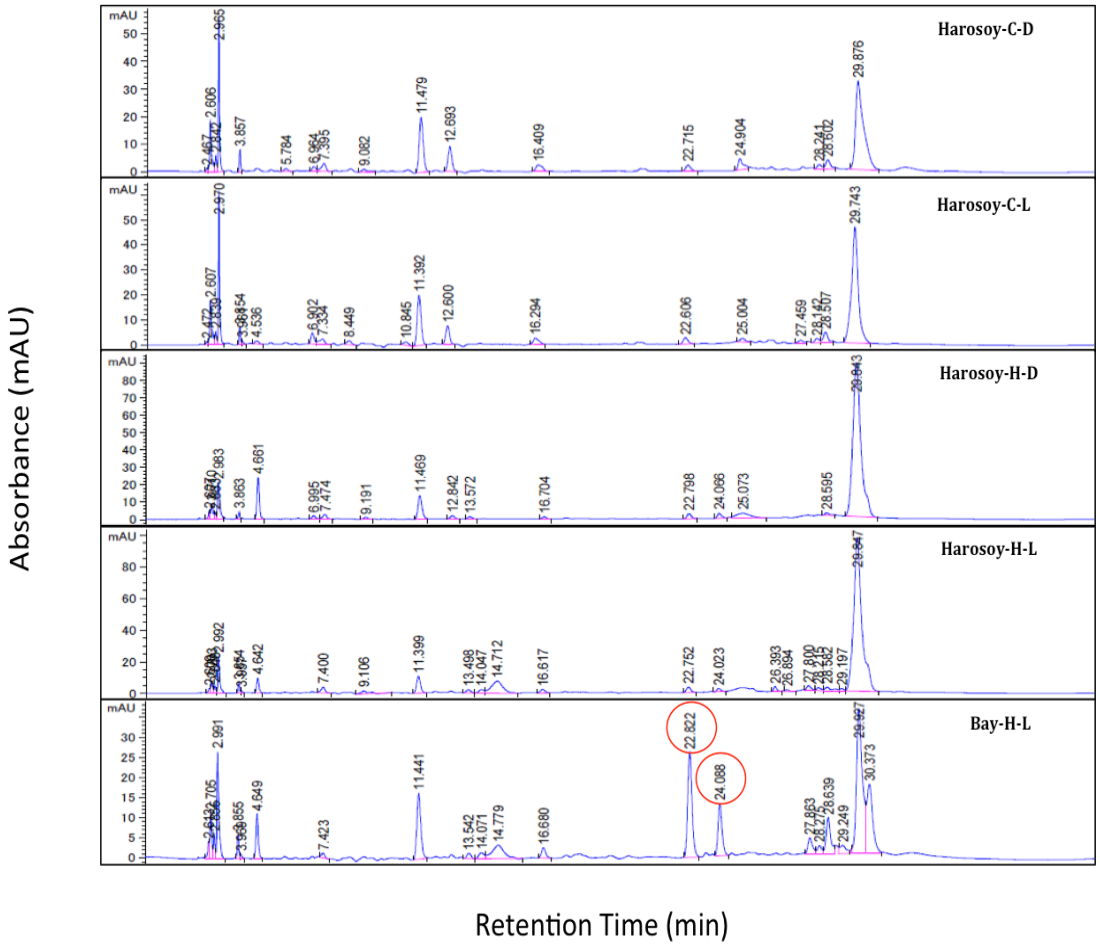

Figure 4.3

HPLC chromatogram of isoflavones

C-D: cotyledon grown in dark, C-L: cotyledon grown under light, H-D: hypocotyl grown in dark, H-L: hypocotyl grown under light. Total of 500 mg tissues was extracted with 5 ml absolute MeOH. Solvent A: 0.2% phosphoric acid; Solvent B: phosphoric acid/acetonitrile/water (0.2:30:70) and the gradient was 0 min, B 20%; 18 min, B 50%; 25 min, B 70%; 31 min, 50%; 34 min, B 20%. Flow-rate: 1.0 ml/min. Injection: 10 µl. Detection wavelength: 260 nm.

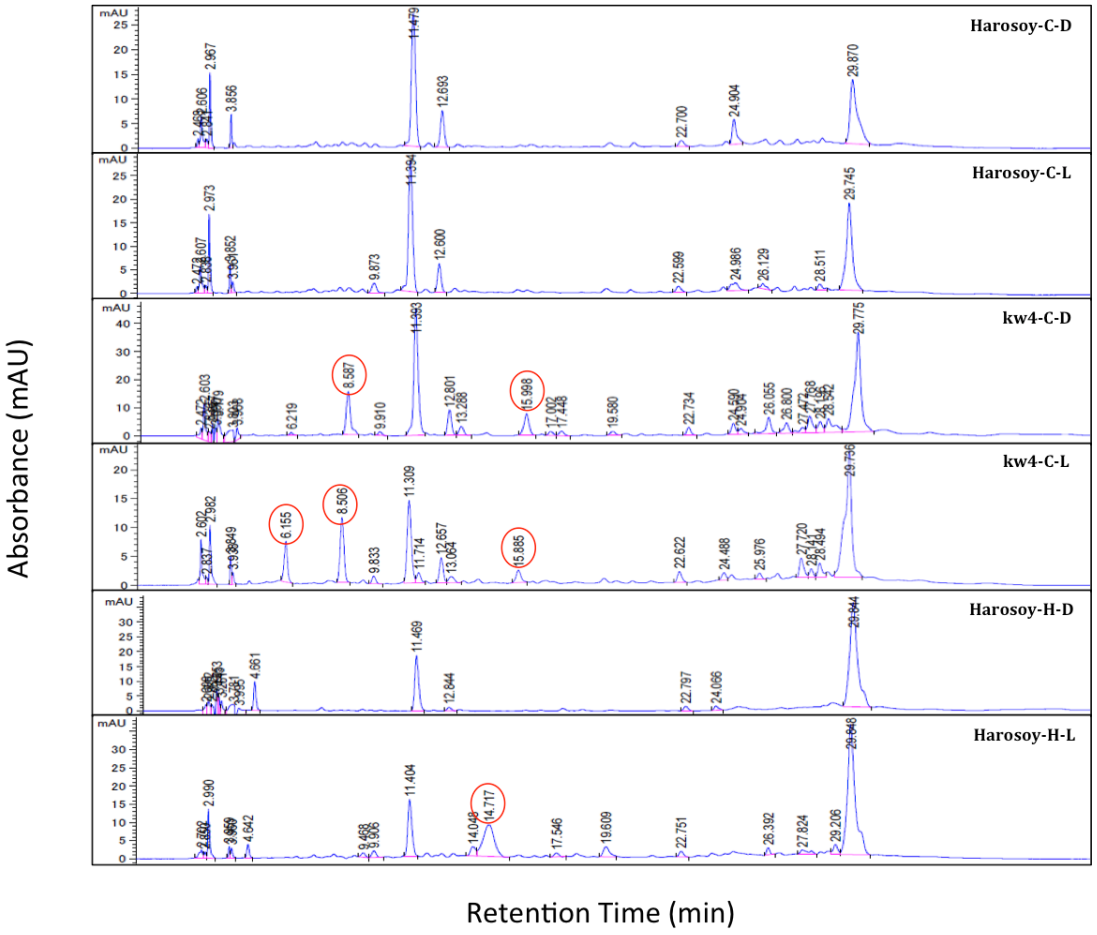

Figure 4.4

HPLC chromatogram of dihydroflavonol

C-D: cotyledon grown in dark, C-L: cotyledon grown under light, H-D: hypocotyl grown in dark, H-L: hypocotyl grown under light. Total of 500 mg tissues was extracted with 5 ml absolute MeOH. Solvent A: 0.2% phosphoric acid; Solvent B: phosphoric acid/acetonitrile/water (0.2:30:70) and the gradient was 0 min, B 20%; 18 min, B 50%; 25 min, B 70%; 31 min, 50%; 34 min, B 20%. Flow-rate: 1.0 ml/min. Injection: 10 µl. Detection wavelength: 280 nm.

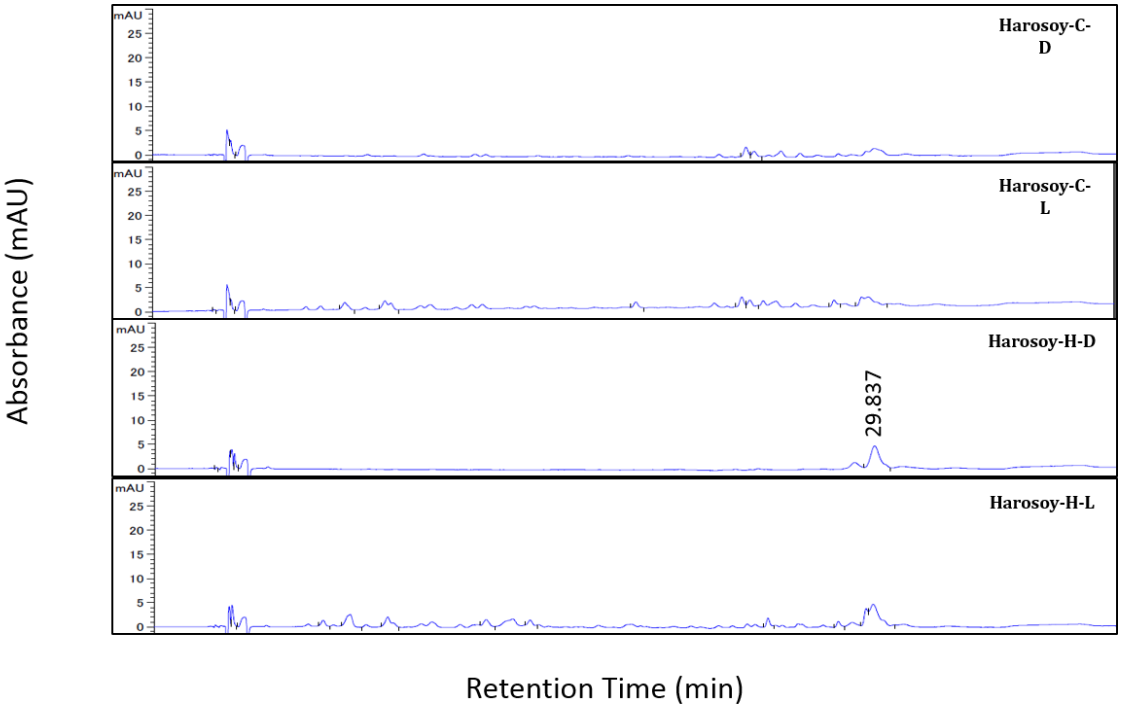

Figure 4.5

HPLC chromatogram of flavonols

C-D: cotyledon grown in dark, C-L: cotyledon grown under light, H-D: hypocotyl grown in dark, H-L: hypocotyl grown under light. Total of 500 mg tissues was extracted with 5 ml absolute MeOH. Solvent A: 0.2% phosphoric acid; Solvent B: phosphoric acid/acetonitrile/water (0.2:30:70) and the gradient was 0 min, B 20%; 18 min, B 50%; 25 min, B 70%; 31 min, 50%; 34 min, B 20%. Flow-rate: 1.0 ml/min. Injection: 10 µl. Detection wavelength: 350 nm.

CHAPTER 5 General discussion

Soybean plants display diverse phenotypes for their flower color, seed coat color, hypocotyl color and pubescence color. *W1* gene encoding flavonoid 3' 5'-hydroxylase has a pleiotropic effect on pigmentation of flower, hypocotyl and seed coat. Generally, soybean cultivars having purple or white flower color have purple or green hypocotyl (Takahashi and Fukuyama, 1919; Groose and Palmer, 1991). Seed coat color is also controlled by *W1* gene (Palmer *et al*., 2004). The flower and seed coat color variants are likely the result of mutation affecting transcription, expression or function of various genes responsible for anthocyanin biosynthesis (Iwashina *et al*., 2007; Iwashina *et al*., 2008; Takahashi *et al*., 2010; Gillman *et al*., 2011; Zabala and Vodkin, 2014).

Previous studies revealed that mutation of *F3'5'H* gene or *DFR2* gene lead to near white, pale, dilute or light purple flowers (Hartwig and Hinson, 1962; Xu *et al*., 2010; Takahashi *et al*., 2010). In our study, genetic analysis and molecular cloning revealed that flower colors of 222-A-3, E30-D-1 and kw4 were also controlled by *W4* locus encoding DFR2, whereas E023-H-12 was controlled by *W1* locus encoding F3'5'H. Amino acid polymorphism (i.e., truncation) or null expression was associated with these flower color variation.

Soybean has three copies of *DFR* genes, *DRF1*, *DFR2* and *DFR3* (Yang *et al*., 2010). The function of *DFR2* may be partially supplemented by the activity of other *DFR* genes depending on environmental conditions. In kw4, *DFR2* gene was not expressed in flower petal. The transcript level of Clark-*w4* and 222-A-3 was

substantially lower than that of Bay, probably because of nonsense-mediated mRNA decay, surveillance mechanisms to eliminate aberrant mRNA transcripts that contain premature stop codons (Chang *et al*., 2007). E30-D-1 and E023-H-12 had single amino acid substitution might have affected transcript abundance and/or gene functions resulting in less anthocyanin contents and light concentration of flower color. Transgenic experiments may be necessary to ascertain the functional of single nucleotide polymorphism. *DFR2* gene and *F3'5'H* gene of soybean controls intensity and distribution of pigmentation in flower petals. Mutation of these two gene results in unique flavonoid composition and a wide variety of flower color patterns, from light purple, dilute purple and pale to near white.

Based on the beneficial effects of flavonoids, we expect to find some mutants and NILs with unique components or high levels of flavonoids beneficial to human health. So, we used twenty-three lines including flower color, seed coat color mutants and near-isogenic lines for evaluation of flavonoid. Based on HPLC results, some lines had specific peaks corresponding to isoflavone and dihydroflavonol. Qualitative and quantitative analysis of each peak are under way. Based on these results, we will carry out experiments to investigate interaction between flavonoid biosynthetic genes and flavonoid components. For example, Clark-*w1* has specific component of peonidin 3, 5-di-*O*-glucoside which was absent in other lines. We are planning to use Clark, Clark-*w1*, Clark-*t*, Clark-*i t w1*, Harosoy, Harosoy-*w1*, Harosoy-*T* and Williams 82 to investigate the relationship between anthocyanin components of malvidin 3, 5-di-*O*-glucoside and peonidin 3, 5-di-*O*-glucoside and genes (*w1* and *T*). Moreover, we will identify more anthocyanin variants understanding of the entire

genetic control of anthocyanin biosynthesis to provide more information for breeding science. Investigate the regulation and expression of anthocyanin biosynthesis genes in the metabolic pathway to improve agronomic and nutritional values in soybean.

A single plant having purple flowers was found in the F_2 population generated from a cross between E023-H-12 and Clark-*w1*. Genetic analysis, gene cloning, dCAPS and indel analysis suggested that the F_2 plant had the heterozygous allele *W1w1-lp2*. The nucleotide sequence of the genomic fragment of an F_3 plant fixed with purple flowers was identical with that of Clark-*w1* from the first exon to the second intron (Fig. 2.11). However, the nucleotide sequence of the third exon was identical with E023-H-12. Intragenic recombination may have occurred in the gray region (Fig. 2.11). The chimeric nature of the gene strongly suggests that outcrossing may not be responsible. Purple-flowered plants were also generated in the F_2 and F_3 populations derived from a cross between plants with light purple flowers (*w1-lp* allele) and plants with white flowers (*w1* allele) (Takahashi*, et al.*, 2010). The frequent occurrence of intragenic recombination in populations derived from a cross with plants having *w1* allele suggests that existence of tandem repeats in the 65-bp insertion is possibly responsible for the high frequency of intragenic recombination.

In our study, we also cloned a large CACTA transposon (*Tgm11*) from L72-2040 having *r-m* allele responsible for seed coat variegation. *Tgm11* as the known active autonomous transposon will be a useful tool for gene engineering and gene mutation. We can also use *Tgm11* as a tool for gene tagging cloning unknown genes

and production of new alleles for understanding functional roles of genes. Therefore, *Tgm11* could be a valuable element for breed science and agriculture.

Summary

This study was carried out to provide more genetic basis and chemical information of anthocyanin biosynthetic pathway for improving agronomic and nutritional properties in soybean. Results are summarized as follow:

1. We investigated the genetic and molecular mechanism of flower color variants. Flower color of Clark-*w4*, T321, T369, 222-A-3, E30-D-1 and kw4 were controlled by the *W4* gene encoding DFR2. Whereas, flower color of E023-H12, PI 424008A and PI 424008C were controlled by *W1* locus encoding F3′5′H.
   i. In Clark-*w4*, base substitution of the first nucleotide of the fourth intron of *DFR2* gene abolished the 5′ splice site, resulting in the retention of the 4th intron (nonfunctional polypeptide).
   ii. In 222-A-3, a single-base deletion of *DFR2* gene probably produced a nonfunctional polypeptide consisting of 24 amino acids.
   iii. The *DFR2* gene of kw4 was not expressed. Thus, loss of function of *W4* gene leads to near white flower colors.
   iv. Flower color of E30-D-1 was controlled by a new allele *w4-lp*.
   v. In T369, expression of *DFR2* gene was 2.3 times that of purple flowers.
   vi. Flower color of E023-H12 was controlled by a new allele *w1-lp2*.
   vii. The entire sequence of *F3'5'H* gene in PI 424008C was identified with Clark-*w1*. Expression level of *F3'5'H* gene was higher in the flower petals of PI 424008A (1.5 times of Bay) and PI 424008C (2.3 times of Bay).

2. We identified a possible causal genetic factor responsible for seed coat variegation in L72-2040, a Clark 63 NIL with the *r-m* allele.
   i. A DNA fragment of 13,060 bp was inserted in the second intron of Glyma09g36990 in L72-2040.
   ii. The fragment had the CACTA motif at both ends, imperfect terminal inverted repeats (TIR), inverse repetition of short sequence motifs close to the 5′ and 3′ ends and duplication of three nucleotides at the site of integration indicating that it belongs to a CACTA-superfamily.
   iii. Existence of transposon footprints, mottling of seed coats and existence of ORFs for transposase genes suggest that *Tgm11* might be an active autonomous element.
   iv. Overall nucleotide sequence, motifs of TIR and subterminal repeats were similar to those of *Tgs1* and *Tgm1*, suggesting that these elements comprise a subfamily. Active non-autonomous element *Tgs1* and *Tgm1* might be deletion derivatives of autonomous elements such as *Tgm11*.

3. We analyzed the components of flavonoids in 23 anthocyanin mutants and near-isogenic lines.
   i. Malvidin 3, 5-di-*O*-glucoside was the predominant components responsible for purple pigmentation in hypocotyl.
   ii. Peonidin 3, 5-di-*O*-glucoside was responsible for the bronze hypocotyl.
   iii. No major peak corresponding flavonols was detected.

iv. In isoflavones analysis, Bay and Bay mutants had two unique peaks in hypocotyl under light environments.

v. In dihydroflavonol analysis, kw4 had two unique peaks in cotyledon under dark and anther peak under light environments.

Abbreviations

CHI	chalcone isomerase
CHS	chalcone synthase
DFR	dihydroflavonol 4-reductase
F3H	flavanone 3-hydroxylase
F3′H	flavonoid 3′-hydroxylase
F3′5′H	flavonoid 3′5′-hydroxylase
FLS	flavonol synthase
ANS	anthocyanidin synthase
ANR	anthocyanidin reductase
LAR	leucoanthocyanidin reductase
UF3GT	UDP-flavonoid 3-*O*-glucosyltransferase
LDOX	leucoanthocyanidin dioxygenase
G. soja	*Glycine soja*
HPLC	high performance liquid chromatography
NIL	near-isogenic line
SNP	single nucleotide polymorphism
TE	transposable element
TIR	terminal inverted repeat
STR	subterminal region
ORF	open reading frame

Acknowledgements

I would like to express my sincere gratitude to my supervisor Prof. Takahashi of National Institute of Crop Science and Tsukuba University, for providing me opportunity to study in Japan. His guidance and encouragement of experiment really helped me a lot in these three years. I really appreciate it.

I would like to express my grateful acknowledges to Prof. Otobe, Prof. Tanaka and Prof. Ohsawa for critical review of the manuscript and valuable comments.

I would like to express thanks to my laboratory members and friends, Tito-san, Felipe-san, Di-san, Auchithya-san and Iizumi-san, for their support and friendship. And kindly acknowledgements to Dr. Kyoko Toda of National Institute of Crop Science, for the help of HPLC analysis. Thanks to Dr. Yoshinori Murai and Dr. Tsukasa Iwashina at Tsukuba Botanical Garden for the help of identified flavonoids.

I want to gratitude to Tsukuba University and National Institute of Crop Science, for accepting me and providing the opportunity that I can study in Japan.

A very special thanks to Prof. Q.Y. Wang of Jinlin University and Prof. D.H. Xu of Japan International Research Center for Agricultural Sciences.

Finally, I want to express thanks to my parents, my husband and my friends, thank you for their kindly support and encouragement.

References

Altschul S.F., Madden T.L., Schaffer A.A., Zhang J.H., Zhang Z., Miller W., Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

Anai T. 2012. Potential of a mutant-based reverse genetic approach for functional genomics and molecular breeding in soybean. Breeding Sci. 61:462-467.Berger M., Rasolohery C. A., Cazalis R., Dayde´ J. 2008. Isoflavone accumulation kinetics in soybean seed cotyledons and hypocotyls: distinct pathways and genetic controls. Crop Sci. 48:700-708.

Bernard R., Nelson R., Cremeens C. 1991. USDA soybean genetic collections: isoline collection. Soyb. Genet. Newsl. 18:27-57.

Bernard R., Singh B.B. 1969. Inheritance of pubescence type in soybeans: glabrous curly, dense, sparse, and puberulent. Crop Sci. 9:192-197.

Brenda W.S. 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol. 126:485-493.

Buzzell R.I., Buttery B.R., Mactavish D.C. 1987. Biochemical Genetics of Black Pigmentation of Soybean Seed. J. Hered. 78:53-54.

Chandlee J.M., Vodkin L.O. 1989. Unstable expression of a soybean gene during seed coat development. Theor. Appl. Genet. 77:587-594.

Chang Y.F., Imam J.S., Wilkinson M.E. 2007. The nonsense-mediated decay RNA surveillance pathway. Annu. Rev. Biochem. 76:51-74.

Chen Y.W., Nelson R.L. 2004. Identification and characterization of a white-flowered wild soybean plant. Crop Sci. 44:339-342.

Dellapenna D. 1999. Nutritional genomics: Manipulating plant micronutrients to improve human health. Science. 285:375-379.

Dixon R.A., Sumner L.W. 2003. Legume natural products: understanding and manipulating complex pathways for human and animal health. Plant Physiol. 131:878-885.

Dixon R.A., Steele C.L. 1999. Flavonoids and isoflavonoids – a gold mine for metabolic engineering. Trends in Plant Science. 4:394-400.

Eldridge A.C., Kwolek W.F. 1983. Soybean isoflavones: effect of environment and variety on composition. J. Agric. Food Chem. 31:394-396.

Falginella L., Castellarin S.D., Testolin R., Gambetta G.A., Morgante M., Di Gaspero G. 2010. Expansion and subfunctionalisation of flavonoid 3', 5'-hydroxylases in the grapevine lineage. BMC Genomics. 11:562.

Fasoula D.A., Stephens P.A., Nickell C.D., Vodkin L.O. 1995. Cosegregation of purple-throat flower color with dihydroflavonol reductase polymorphism in soybean. Crop Sci. 35:1028-1031.

Forkmann G., Martens S. 2001. Metabolic engineering and applications of flavonoids. Curr. Opin. Biotechnol. 12:155-160.

Frey M., Reinecke J., Grant S., Saedler H., Gierl A. 1990. Excision of the *En/Spm* transposable element of Zea mays requires two element-encoded proteins. EMBO J. 9: 4037–4044.

Gierl A., Lutticke S., Saedler H. 1988. TnpA product encoded by the transposable element *En-1* of Zea mays is a DNA binding protein. EMBO J. 7: 4045–4053.

Gierl A., Saedler H., Peterson P.A. 1989. Maize transposable elements. Annu. Rev. Genet. 23:71-85.

Gillman J.D., Tetlow A., Lee J.D., Shannon J.G., 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.

Gotoh O. 1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. J. Biol. Chem. 267:83-90.

Grant S.R., Gierl A., Saedler H. 1990. *En/Spm*-encoded *tnpA* protein requires a specific target sequence for suppression. EMBO J. 9: 2029-2035.

Groose R.W., Palmer R.G. 1991. Gene action governing anthocyanin pigmentation in soybean. J. Hered. 82: 498–501.

Guo Y., Qiu L.J. 2013. Allele-specific marker development and selection efficiencies

for both flavonoid 3'-hydroxylase and flavonoid 3', 5'-hydroxylase genes in soybean subgenus *soja*. Theor. Appl. Genet. 126:1445-1455.

Hartwig E.E., Hinson K. 1962. Inheritance of flower color of soybeans. Crop Sci. 2:152-153.

Haslam E. 1998. Practical polyphenolics: from structure to molecular recognition and physiological action. Cambridge: Cambridge University Press.10-18.

Holton T.A., Cornish E.C. 1995. Genetics and biochemistry of anthocyanin biosynthsis. Plant Cell. 7:1071-1083.

Iwashina T., Oyoo M.E., Khan N.A., Matsumura H., Takahashi R. 2008. Analysis of flavonoids in flower petals of soybean flower color variants. Crop Sci. 48:1918-1924.

Iwashina T., Githiri S.M., Benitez E.R., Takemura T., Kitajima J., Takahashi R. 2007. Analysis of flavonoids in flower petals of soybean near-isogenic lines for flower and pubescence color genes. J. Hered. 98:250-257.

Kamiya M., Kiguchi T. 2003. Rapid DNA extraction method from soybean seeds. Breeding Sci. 53:277-279.

Kim J.A., Hong S.B., Jung W.S., Yu C.Y., Ma K.H.; Gwag J.G., Chung I.M. 2007. Comparison of isoflavones composition in seed, embryo, cotyledon and seed coat of cooked-with-rice and vegetable soybean (*Glycine max* L.) varieties. Food Chem. 102:738-744.

Kong J.M., Chia L.S., Goh N.K., Chia T.F., Brouillard R. 2003. Analysis and biological activities of anthocyanins. Phytochemistry. 64:923-933.

Lacombe E., Hawkins S., VanDoorsselaere J., Piquemal J., Goffner D., Poeydomenge O., Boudet A.M., GrimaPettenati J. 1997. Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: Cloning, expression and phylogenetic relationships. Plant J. 11:429-441.

Lepiniec L., Debeaujon I., Routaboul J.M., Baudry A., Pourcel L., Nesi N., Caboche M. 2006. Genetics and biochemistry of seed flavonoids. Annu. Rev. Plant Biol. 57:405-430.

Lisch D. 2009. Epigenetic regulation of transposable elements in plants. Annu. Rev.

Plant Biol. 60:43-66.

Marchler-Bauer A., Lu S.N., Anderson J.B., Chitsaz F., Derbyshire M.K., Deweese-Scott C., Fong J.H., Geer L.Y., Geer R.C., Gonzales N.R., *et al.* 2011. CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res. 39:D225-D229.

Masson P., Rutherford G., Banks J.A., Fedoroff N. 1989. Essential large transcripts of the maize spm transposable element are generated by alternative splicing. Cell. 58:755-765.

McClitock B. 1948. Mutable loci in maize. Carnegie Inst Wash Year B. 47:155-169.

Mol J., Grotewold E., Koes R. 1998. How genes paint flowers and seeds. Trends in Plant Sci. 3: 212-217.

Murkovic M., Adam U., Pfannhauser W., 2000. Analysis of anthocyane glycosides in human serum, Fresenius J. Anal. Chem. 366:379-381.

Murphy P.A., Barua K., Hauck C.C. 2002. Solvent extraction selection in the determination of isoflavones in soy foods. J. Chromatogr. B. 777:129-138.

Murray M.G., Thompson W.F. 1980. Rapid isolation of high molecular-weight plant DNA. Nucleic Acids Res. 8:4321-4325.

Nacken W.K.F., Piotrowiak R., Saedler H., Sommer H. 1991. The transposable element Tam1 from *Antirrhinum majus* shows structural homology to the maize transposon En/Spm and has no sequence specificity of insertion. Mol. Gen. Genet. 228:201-208.

Nagai I., Saito S. 1923. Linked factors in soybeans. Jpn. J. Bot. 1:121-136.

Nozzolillo C. 1973. A survey of anthocyanin pigments in seedling legumes. Can. J. Bot. 51:911-915.

Owen F.V. 1928. Inheritance studies in soybeans. III. Seed-coat color and summary of all other mendelian characters thus far reported. Genetics. 13(1):50-79.

Palmer R.G., Pfeiffer T., Buss G., Kilen T. 2004 Qualitative genetics. In: Boerma H, Specht J, editors. Soybeans: Improvement, production, and uses. 3rd ed. Madison, WI: ASA, CSSA, and SSSA; p. 137-233.

Palmer R.G., Groose R.W. 1993. A new allele at the *W4* locus drived from the *w4-m*

mutable alele in soybean. J. Hered. 84:297-300.

Palmer R.G., Hedges B.R., Benavente R.S., Groose R.W. 1989. *w4*-mutable line in soybean. Dev. Genet. 10:542-551.

Peters D.W., Wilcox J.R., Vorst J.J., Nielsen N.C. 1984. Hypocotyl pigments in soybeans. Crop Sci. 24:237-239.

Phommalth S., Jeong Y.S., Kim Y.H., Dhakal K.H., Hwang Y.H. 2008. Effects of light treatment on isoflavone content of germinated soybean seeds. J. Agric. Food Chem. 56:10123-10128.

Prior R.L. 2003. Fruits and vegetables in the prevention of cellular oxidative damage. Am. J. Clin. Nutr. 78:570S-578S.

Rhodes P.R., Vodkin L.O. 1988. Organization of the *Tgm* family of transposable elements in soybean. Genetics. 120:597-604.

Ribeiro M.L.L., Mandarino J.M.G., Carrao-Panizzi M.C., Oliveira M.C.N., Campo C. B. H., Nepomuceno A. L., Ida E. I. 2006. β-Glucosidase activity and isoflavone content in germinated soybean radicles and cotyledons. J. Food Biochem. 30:453-465.

Seitz C., Ameres S., Forkmann G. 2007. Identification of the molecular basis for the functional difference between flavonoid 3'-hydroxylase and flavonoid 3',5'-hydroxylase. Febs Lett. 581:3429-3434.

Shah D.M., Hightower R.C., Meagher R.B. 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.

Shimada S., Takahashi K., Sato Y., Sakuta M. 2004. Dihydroflavonol 4-reductase cDNA from non-anthocyanin-producing species in the Caryophyllales. Plant Cell Physiol. 45:1290-1298.

Stadler D.R. 1973. Mechanism of Intragenic Recombination. Annu. Rev. Genet. 7:113-127.

Takahashi R., Benitez E.R., Oyoo M.E., Khan N.A., Komatsu S. 2011. Nonsense mutation of an MYB transcription factor is associated with purple-blue flower color in soybean. J. Hered. 102:458-463.

Takahashi R., Dubouzet J.G., Matsumura H., Yasuda K., Iwashina T. 2010. A new allele of flower color gene *W1* encoding flavonoid 3'5'-hydroxylase is responsible for light purple flowers in wild soybean *Glycine soja*. BMC Plant Biol. 10:155.

Takahashi R., Githiri S.M., Hatayama K., Dubouzet E.G., Shimada N., Aoki T., Ayabe S., Iwashina T., Toda K., Matsumura H. 2007. A single-base deletion in soybean flavonol synthase gene is associated with magenta flower color. Plant Mol. Biol. 63:125-135.

Takahashi R., Matsumura H., Oyoo M.E., Khan N.A. 2008. Genetic and linkage analysis of purple-blue flower in soybean. J. Hered. 99:593-597.

Takahashi R., Morita Y., Nakayama M., Kanazawa A., Abe J. 2012. An Active CACTA-Family Transposable Element is Responsible for Flower Variegation in Wild Soybean Glycine soja. Plant Genome. 5:62-70.

Takahashi R., Yamagishi N., Yoshikawa N. 2013. A MYB transcription factor controls flower color in soybean. J. Hered. 104:149-153.

Takahashi Y., Fukuyama J. 1919. Morphological and genetic studies on soybean (in Japanese). Hokkaido Agr. Exp. Stn. Rep. 10:1-100.

Graham T.L. 1991. Flavonoid and Isoflavonoid Distribution in Developing Soybean Seedling Tissues and in Seed and Root Exudates. Plant Physiol. 95:594-603.

Toda K., Yang D.J., 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 J.J., Vodkin L.O. 1993. Pigmented soybean (*Glycine max*) seed coats accumulate proanthocyanidins during development. Plant Physiol. 102:663-670.

Tsangalis D., Ashton J.F., Stojanovska L., Wilcox G., Shah N.P. 2004. Development of an isoflavone aglycone-enriched soymilk using soy germ, soy protein isolate and bifidobacteria. Food Res. Int. 37:301-312.

Vodkin L.O., Rhodes P.R., Goldberg R.B. 1983. Ca lectin gene insertion has the structural features of a transposable element. Cell. 34:1023-1031.

Weiss M.G. 1970. Genetic linkage in soybeans: linkage group II and group III. Crop Sci. 10:300-303.

Wicker T., Sabot F., Hua-Van A., Bennetzen J.L., Capy P., Chalhoub B., Flavell A., Leroy P., Morgante M., Panaud O.*, et al.* 2007. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8:973-982.

Williams L.F. 1952. The inheritance of certain black and brown pigments in the soybean. Genetics. 37(2):208-215.

Winkel-Shirley B. 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol. 126:485-493.

Xu M., Brar H.K., Grosic S., Palmer R.G., Bhattacharyya M.K. 2010. Excision of an active CACTA-like transposable element from *DFR2* causes variegated flowers in soybean [*Glycine max* (L.) Merr.]. Genetics. 184:53-63.

Xu M., Palmer R.G. 2005. Genetic analysis and molecular mapping of a pale flower allele at the *W4* locus in soybean. Genome. 48(2):334-340.

Yang K., Jeong N., Moon J.K., Lee Y.H., Lee S.H., Kim H.M., Hwang C.H., Back K., Palmer R.G., Jeong S.C. 2010. Genetic analysis of genes controlling natural variation of seed coat and flower colors in soybean. J. Hered. 101:757-768.

Youdim K.A., Shukitt-Hale B., MacKinnon S., Kalt W., Joseph J.A. 2000. Polyphenolics enhance red blood cell resistance to oxidative stress: In vitro and in vivo, Biochim. Biophys. Acta. 1523:117-122.

Yuan J.P., Liu Y.B., Peng J., Wang J.H., Liu X. 2009. Changes of Isoflavone Profile in the Hypocotyls and Cotyledons of Soybeans during Dry Heating and Germination. J. Agric. Food Chem. 57:9002-9010.

Zabala G., Vodkin L.O. 2005. The *wp* mutation of *Glycine max* carries a gene-fragment-rich transposon of the CACTA superfamily. Plant Cell. 17:2619-32.

Zabala G., Vodkin L.O. 2007. A rearrangement resulting in small tandem repeats in the F3'5'H gene of white flower genotypes is associated with the soybean *W1* locus. Crop Sci. 47:S113-S124.

Zabala G., Vodkin L.O. 2008. A putative autonomous 20.5 kb-CACTA transposon insertion in an F3'H allele identifies a new CACTA transposon subfamily in

Glycine max. BMC Plant Biol. 8:124.

Zabala G., Vodkin L.O. 2014. Methylation affects transposition and splicing of a large cacta transposon from a MYB transcription factor regulating anthocyanin synthase genes in soybean seed coats. Plos One. 9(11): e111959.