Genotypic Variation and Molecular Genetics of Flavonol Glycoside Biosynthesis in Soybean Leaves

May 2015

Shaokang DI

Genotypic Variation and Molecular Genetics of Flavonol Glycoside Biosynthesis in Soybean Leaves

A Dissertation Submitted to

the Graduate School of Life Environmental Sciences

the University of Tsukuba

in Partial Fulfillment of the Requirements

Foe the Degree of Doctor in Philosophy in Agricultural Sciences

(Doctoral Program in Advanced Agricultural Technology and Sciences)

Shaokang DI

TABLE OF CONTENTS

CHAPTER 5

5. General discussion

Summary

Acknowledgments

References

CHAPTER 1

Introduction

1.1 Introduction of flavonoids

Flavonoids is a kind of low molecular secondary metabolites and belong to a large family of natural substances with variable phenolic structures (Koes *et al.* 1994). They are found in wide range of plants species such as fruit, vegetables, grains, bark, roots, stems and flowers, (Middleton *et al.* 1998).

Flavonoids have been studied many years and their potential benefits for human health have been revealed. They have been proved to counteract the growth of several types of cancer through multiple mechanisms (Maggioni *et al.* 2014), against coronary heart disease (de Groot and Rauen. 1998) and antibacterial, antiviral and anti-inflammatory effects (Bueno-Silva *et al.* 2013).

The presence of flavonoids in plant tissues also play important roles, such as providing flower pigmentation to attract pollinators (Winkel-Shirley. 2001a, b), involvement in pollen germination (Tanaka and Ohmiya. 2008), protection against insect predation and defense against microbes (Bidart-Bouzat and Kliebenstein. 2008), protecting plants from UV radiation, thus replacing mycosporine-like amino acids detected in algae (Guo *et al*. 2008; Agati *et al*. 2013), acting as signal molecules in biological communication in the rhizosphere (Cesco *et al.* 2012; Weston and Mathesius. 2013) and possible plant energy escape valves (Hernandez and Breusegem. 2012).

1.2 Flavonoid structure and classification

Flavonoids are benzo-γ-pyrone derivatives consisting of phenolic and pyrane rings

(Fig. 1) and characterized by a planar structure because of a double bond in the central aromatic ring (Nijveldt *et al*. 2001; Heim *et al.* 2002). flavonoids differ in the arrangements of hydroxyl, methoxy, and glycosidic side groups, and in the conjugation between the A- and B- rings. One of the best-described flavonoids, quercetin, is a member of this group. Quercetin is found in abundance in onions, apples, broccoli, and berries. The second group is the flavanones, which are mainly found in citrus fruit. An example of a flavonoid of this group is narigin. Flavonoids can be divided into various classes on the basis of their molecular structure (Rice-Evans *et al.* 1996). The major groups of flavonoids are exhibited in Fig. 2.

1.3 Pathway of flavonoid biosynthesis

Molecular aspects of flavonoid pathway have been extensively investigated and reviewed (Winkel-Shirley. 2001a, b; Grotewold. 2006a; Saito *et al.* 2013). Many genes have been identified within the flavonoid biosynthetic pathway. Plant phenolics are products of the shikimate, phenylpropanoid, flavonoid, anthocyanin, and lignin pathways. The shikimate pathway produces the aromatic amino acids, including phenylalanine, which can be further modified through the sequential of elongation and cyclization steps to form the flavonoids (Fig. 3). In general, flavonoids are sub-classified into several families of fifteen-carbon molecules including flavonol, flavone, flavanone, flavan-3-ol, isoflavone and anthocyanidin (Ferreyra *et al.* 2012). As a first step, phenylalanine gets converted into coumaroyl-CoA by a number of enzymatic reactions involving PAL, cinnamate 4-hydroxylase (C4H) and 4-coumarate:CoA ligase (4CL). The CHS and CHI convert coumaroyl-CoA into naringenin and onward, the pathway diverges forming flavanones, dihydroflavonols, leucoanthocyanins, anthocyanidins, and flavan-3-ols through a series of enzymatic steps shown in Fig.3. These molecules can produce a series of products, flavones and isoflavones, flavonols, anthocyanins, and proanthocyanidins

(condensed tannins) existing in the forms of monomers, dimers, and polymers (Chan *et al.* 2009; Pati *et al.* 2009; Abeynayake *et al.* 2012; Hancock *et al.* 2014).

1.4 Flavonol glycosides genes for flavonoid biosynthesis in soybean

Leaves of soybean (*Glycine max* (L.) Merr.) contain a variety of flavonol glycosides (FGs), derivatives of quercetin and kaempferol (Buttery and Buzzell. 1973). Also HPLC analysis of the soybean leaves showed that different cultivars have different kind and content of flavonol glycosides (Table 1) (Rojas Rodas *et al.* 2014). Buttery and Buzzell (1975) proposed four flavonol glycoside alleles, *Fg1* (β(1-6)-glucoside present), *Fg2* (α(1-6)-rhamnoside present), *Fg3* (β(1-2)-glucoside present), and *Fg4* (α(1-2)-rhamnoside present). These alleles are defined by the ability to bind glucose or rhamnose at either 2″- or 6″-position to glucose that is bound to the 3-position of flavonols (Fig. 4). Later, Buzzell and Buttery (1992) reported a new allele of the *Fg2* locus, resulting in a series of alleles, *Fg2-a*, *Fg2-b* and *fg2*. *Fg3* and *Fg4* are linked with a recombination frequency of 12 % in the molecular linkage group C2 (chromosome 6) (Buzzell. 1974).

Glycosyltransferases (GTs) catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules, forming glycosidic bonds (Lairson *et al.* 2008). GTs are encoded by large multigene families, and classified into at least 96 families (GT1-GT96, http://www.cazy.org/GlycosylTransferases.html). The family 1 glycosyltransferase, referred to as UDP glycosyltransferases (UGTs), comprise the largest group in plants. UGTs catalyze the transfer of a glycosyl moiety from UDP sugars to a wide range of acceptor molecules including flavonoids (Yonekura-Sakakibara and Hanada. 2011).

Rojas Rodas *et al.* (2014) performed genetic analysis using RILs derived from a cross between soybean cultivars with gray pubescence, Koganejiro and Kitakomachi. FGs of Koganejiro had rhamnose at the 6″-position of glucose or galactose that is bound to the

3-position of kaempferol, whereas FGs of Kitakomachi were devoid of the rhamnose. Presence of the 6″-rhamnose was controlled by a single gene. They cloned a candidate gene *GmF3G6″Rt* in the molecular linkage group O (chromosome 10). The recombinant *GmF3G6″Rt* protein converted UDP-rhamnose and kaempferol 3-*O*-glucoside or kaempferol 3-*O*-galacoside to kaempferol 3-*O*-rutinoside or kaempferol 3-*O*-robinobioside, proving that *GmF3G6″Rt* encodes flavonol 3-*O*-glucoside (1→6) rhamnosyltransferase and corresponds to the *Fg2* gene. Thus, either glucose or galactose was attached at the 3-position of kaempferol in contrast to the report of (Buttery and Buzzell. 1975) in which only glucose was attached to the 3-position. In addition, FGs having rhamnose at the 4″-position of 3-*O*-galactose, kaempferol 3-*O*-rhamnosyl-(1→4)-[rhamnosyl-(1→6)-galactoside] and kaempferol 3-*O*-rhamnosyl-(1→4)-[glucosyl-(1→6)-galactoside] have been identified in the leaves of soybean (Murai *et al.* 2013; Rojas Rodas *et al.* 2014), suggesting the existence of flavonol 3-*O*-glucoside (1→4) rhamnosyltransferase in soybean. Accordingly, the schematic diagram on the genetic control of FG biosynthesis proposed by (Buttery and Buzzell. 1975) should be revised. Overall, glucose or galactose is attached to the 3-position of kaempferol or quercetin in the biosynthesis of FGs in soybean leaves. To the glucose or galactose, glucose can be attached to 2″- or 6″-positions and rhamnose can be attached to 2″-, 4″- or 6″-positions, resulting in a wide variety of FGs (Rojas Rodas *et al.* 2014).

In 1976, Buttery and Buzzell found that plants with the *Fg1Fg3* alleles have a lower rate of photosynthesis, lower leaf chlorophyll concentration, lower leaf weight, and lower seed yield and (Buzzell and Buttery l. 1976). Further, *Fg1* and *Fg3* control waviness of leaf margins (Buttery and Buzzell. 1998) and inhibition of stomatal development (Liu-Gitz *et al.* 2000).

1.5 Aim of this study

Preliminary experiments suggested that a Japanese landrace Nezumisaya and a Canadian cultivar Harosoy, both having gray pubescence, had a distinct FG composition in leaves (Iwashina *et al.* unpublished results) and using different cultivars cross to identify the FGs is a possible method.

The first objective of this study was to identify other genes for FG biosynthesis, locate them in the soybean genome and verify their function using crosses between different germplasm. The second objective was to evaluate agronomic importance of FG genes by developing near-isogenic lines (NILs) for FG genes. The third objective was to evaluate variation of FGs in leaves among various germplasm of soybean and *G.soja*. By understand the biosynthesis pathway of FGs in soybean leaves and the physiological effects in soybean, we can improve the quality and yield of soybean cultivars and make more profitable to agriculture.

Table 1: Flavonol glycosides distribution in leaves of different soybean germplasm (in percentage). (Unpublished)

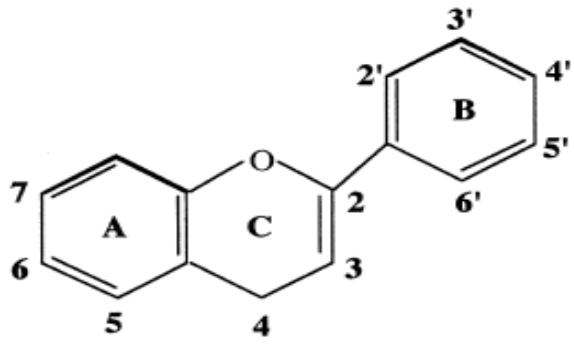

Figure 1.

Nuclear structure of flavonoids (Heim *et al.* 2002).

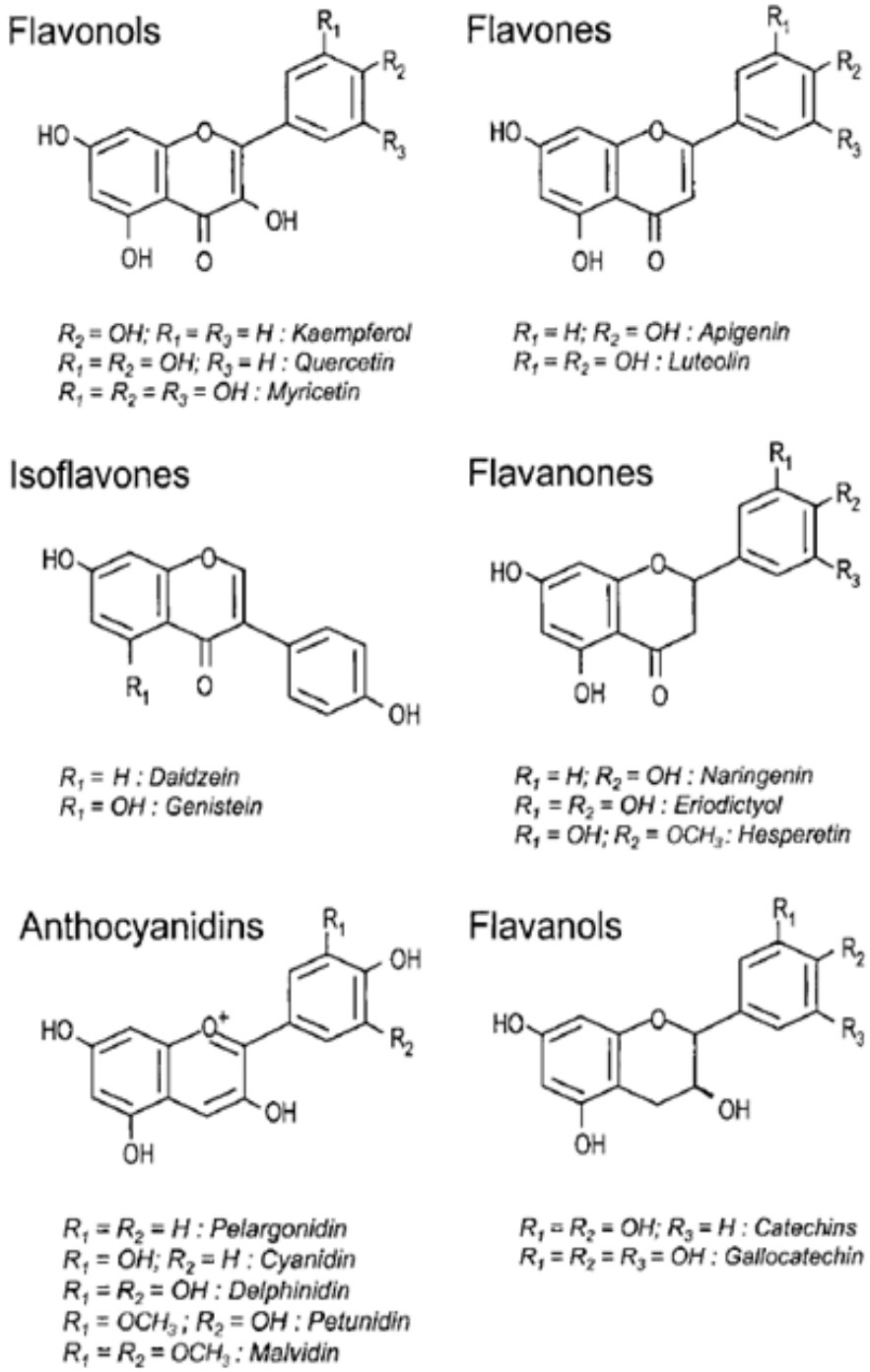

Figure 2.

Classification of flavonoid groups (Manach *et al*. 2004).

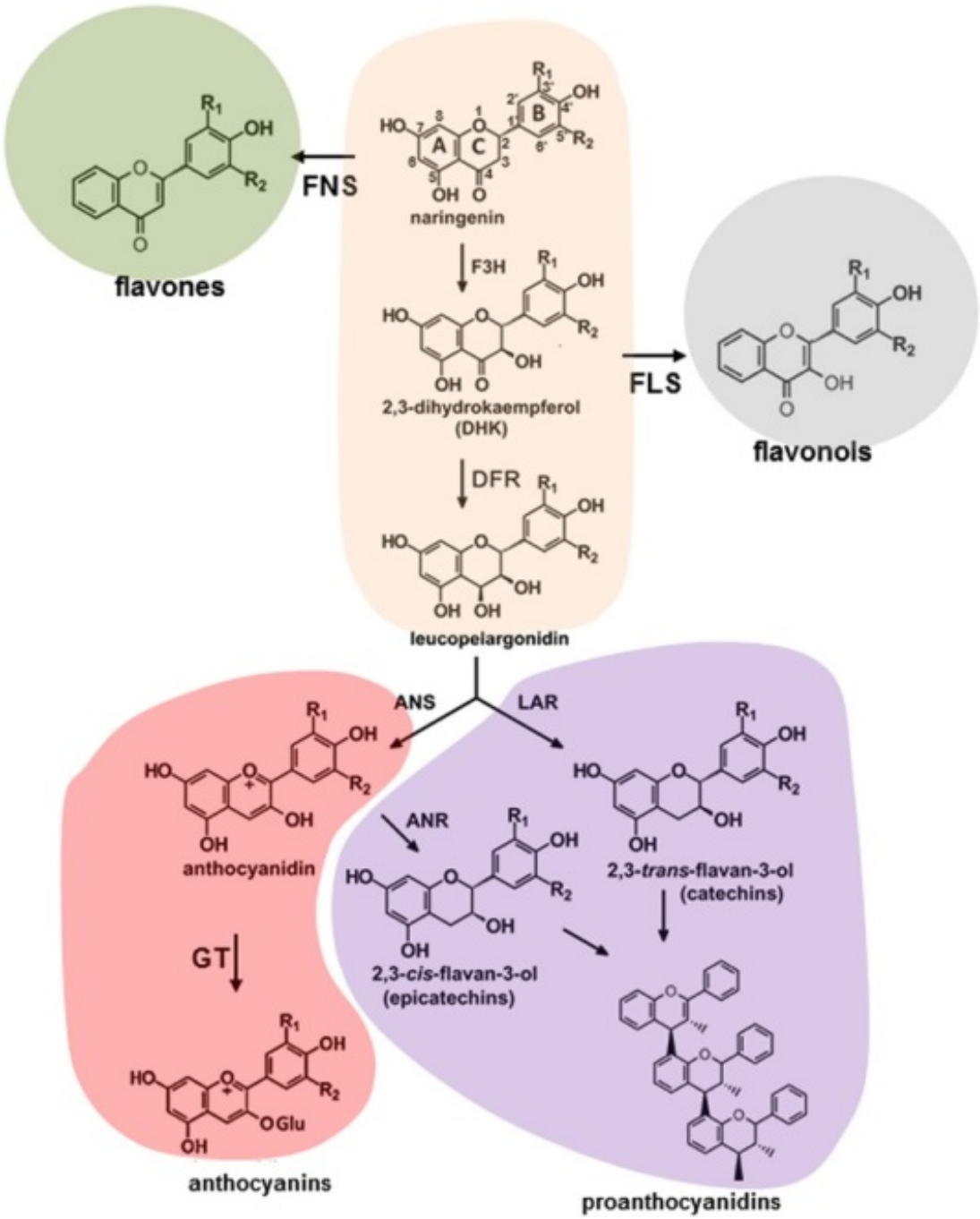

Figure 3.

Schematic representation of flavonoid pathway in plants (Mouradov and Spangenberg. 2014).

ANR, anthocyanidin reductase; ANS, anthocyanidin synthase; DFR, dihydroflavonol 4-reductase; F3H, flavonoid 3′-hydroxylase; FLS, flavonole synthase; FNS, flavone synthase; GT, glucosyltransferase; LAR, leucoanthocyanidin reductase.

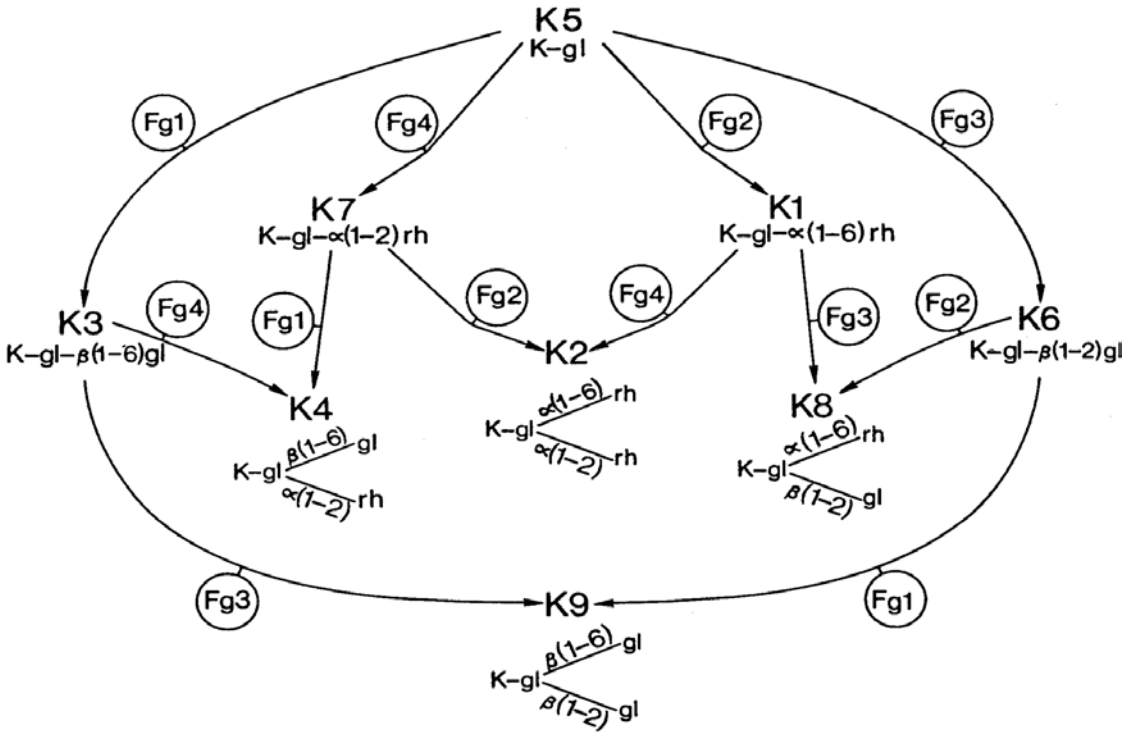

Figure 4.

Schematic presentation of flavonol glycoside genes (Buttery and Buzzell. 1975).

CHAPTER 2

Linkage mapping, molecular cloning and functional analysis of FG gene segregated in Nezumisaya × Harosoy population

2.1 Background

Preliminary experiments suggested that a Japanese landrace Nezumisaya and Harosoy, both having gray pubescence, had a distinct FG composition in leaves (Iwashina *et al* unpublished results). Two genes are involved in Nezumisaya × Harosoy F_6 RILs. Linkage mapping revealed that one of the gene is responsible for attachment of glucose to the 6″-position which located in the molecular linkage group I (chromosome 20) between Satt623 and Sat_420. The other gene is responsible for attachment of glucose to the 2″-position.

The objective of this study is to locate, identify, clone, sequence, verify the function of candidate gene responsible for attachment of glucose to the 2″-position and develop NILs for the target gene to evaluate agronomic characters in planta.

2.2 Material and Methods

2.2.1 Plant Materials

Nezumisaya crossed with harosoy and selfing to develop recombinant inbred lines to F_6 generation. (NH population) Nezumisaya with yellow hilum, yellow seed coats, gray pubescence and purple-blue flowers (*IIttW1W1w2w2w3w3W4W4WmWmWpWp*) was crossed with a Canadian cultivar Harosoy with yellow hilum, yellow seed coats, gray pubescence and purple flowers (*IIttW1W1W2W2 w3w3W4W4WmWmWpWp*). Flowers of Nezumisaya were emasculated one day before opening and pollinated with

Harosoy in 2004. Hybridity of F_1 plants was ascertained based on purple flowers. A total of 120 RILs of F_6 generation were developed without any selection by the single-seed descent method from the F_1 plant. Seeds were planted at the National Institute of Crop Science, Tsukuba, Japan (36º06'N, 140º05'E) on June 9, 2011. N, P, and K were applied at 3.0, 4.4, and 8.3 g m^{-2}, respectively. Plants were individually planted with spaces between rows and plants of 70 and 10 cm, respectively. On average, nine plants were grown for each parent and RIL (Rojas Rodas *et al.* unpublished).

2.2.2 SSR analysis

Genomic DNA of the parents and the F_2 plants was isolated from trifoliolate leaves by the CTAB method (Murray and Thompson. 1980). A total of 462 SSR markers developed by USDA (Song *et al.* 2004) and by the Kazusa DNA Research Institute (Hisano *et al.* 2007) was used for screening of polymorphisms between the parents. The PCR mixture contained 20 ng of genomic DNA, 2.25 pmol of primer, 625 pmol of nucleotides, and 0.125 unit of ExTaq in 1 x ExTaq buffer supplied by the manufacturer (Takara Bio, Ohtsu, Japan) in a total volume of 5 μL. The initial 15 min denaturation at 95º C was followed by 35 cycles of 1 min denaturation at 92º C, 1 min annealing at 46 ºC, and 1 min extension at 68º C. A final extension at 68º C for 5 min completed the program. PCR was performed in an Applied Biosystems 9700 thermal cycler (Applied Biosystems, Foster City, CA). The PCR products were separated in 8 % nondenaturing acrylamide gels, and the fragments were visualized by staining with ethidium bromide. (Rojas Rodas *et al.* 2014).

2.2.3 Linkage mapping

The linkage map was constructed using AntMap software version 1.2 (Iwata and Ninomiya. 2006). The parameter were: grouping method – all combinations; grouping criterion – LOD; threshold value – 3; map function – Kosambi; the designation of the linkage groups followed by Cregan *et al*. (1999). All the marker were tested by Chi-square test for segregation in 1:1 ratio.

2.2.4 QTL analysis (Unpublished)

2.2.5 Molecular cloning

Total RNA was extracted from trifoliolate leaves (200 mg) of Nezumisaya and Harosoy using the TRIZOL reagent (Invitrogen) according to the manufacturer's instructions. cDNA was obtained by reverse transcription of total RNA using the superscript Ⅲ First-Stand Synthesis System (Invitrogen) and an oligo d (T) primer according to the manufacturer's instructions. The full-length cDNA was cloned by end-to-end PCR from Nezumisaya and Harosoy using a pair of PCR primers (Table 2) designed based on the genome sequence of US cultivar Williams 82 deposited in the soybean genome database (Phytozome, http://www.phytozome.net/soybean.php). The PCR mixture contained 0.5 μg of cDNA, 10 pmol of each primer, 5 pmol of nucleotides 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^{o}C and 1 min extension at 72^{o}C. A final 7 min extension at 72^{o}C completed the program. The 5` upstream region which about 1.8 kb was amplified from Nezumisaya and Harosoy using GenomicWalker (Invitrogen) according to the manufacturer's instructions. Primer using for nested PCR was shown in Table 1. The PCR product was cloned into

pCR 2.1 vector (Invitrogen) and sequenced.

2.2.6 Sequencing analysis

Nucleotide sequences of both strands were determined with the BigDye terminator cycle method using an ABI3100 Genetic Analyzer (Applied Biosystems). Nucleotide sequences and the putative amino acid translations were analyzed with the BLAST program (Altschul *et al*. 1997). Intron/exon structure of the gene was estimated based on the comparison between the cDNA sequence and the corresponding genome sequences of Williams 82 deposited in the soybean genome database. Sequence alignment was performed with ClustalW (http://clustalw.ddbj.nig.ac.jp/top-j.html) using default settings. The amino acid sequences were aligned using ClustalW and the alignment was used to construct a phylogenetic tree using the neighbor-joining method with MEGA5 version 5.2.2 (http://www.megasoftware.net/) (Tamura *et al*. 2011). Bootstrap test was performed for1000 replication

2.2.7 dCAPS analysis

A pair of PCR primers (Table 2) was designed to detect the two-base substitution. A mismatched base G was incorporated in the forward primer to the produce a *Kpn*I site in the amplified product Harosoy. The substitution within the restriction site absence in Nezumisaya to generate 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. After an initial 30 sec denaturation at 94 ^{o}C, there were 30 cycles of 30 sec denaturation at 94 ^{o}C, 1 min annealing at 56 ^{o}C and 1 min extension at 72 ^{o}C. A final

7 min extension at 72 ^{o}C completed the program. The amplified products were digested with *Kpn*I, and the digests were separated on an 8 % nondenaturing polyacrylamide gels. After electrophoresis, the gel was stained with ethidium bromide and the DNA fragments were visualized under UV light.

2.2.8 In vitro characterization of recombinant GmF3G2″Gt

The entire coding region of *GmF3G2"Gt* was amplified from cDNA of Harosoy and Nezumisaya by PCR using the KOD -Plus- DNA polymerase (Toyobo) with high PCR fidelity using primers containing restriction enzyme sites of *Sac*I and *Xho*I (Table 2). The PCR mixture contained 30 ng of genomic DNA, 10 pmol of each primer, 5 pmol of nucleotides, 2 mM of $MgSO_4$ and 0.5 unit of KOD –Plus- in 1 x KOD –Plus- Buffer supplied by the manufacturer in a total volume of 25 μl. After an initial 2 min denaturation at 94^{o}C, there were 40 cycles of 15 sec denaturation at 94^{o}C, 30 sec annealing at 59^{o}C and 1 min extension at 68^{o}C. A final 7 min extension at 68^{o}C completed the program. The PCR amplicon was digested by *Sac*I and *Xho*I, and was cloned into the pCold ProS2 vector (Takara Bio) digested with the same restriction enzymes. The plasmid was sent to RIKEN for functional analysis. GmF3G2"Gt proteins were expressed and purified as described previously (Yonekura et al. 2014). Enzyme assays and MS and MS/MS analyses were conducted as described previously (Rojas Rodas *et al*. 2014; Yonekura et al. 2014). The ESI source was operated in negative ionization mode.

2.2.9 Quantitative real-time PCR

Quantitative real-time PCR cDNA was conducted by reverse-transcription of 5 μg of total RNA of the parents at R6 stage from three replicated leaves using the

Superscript III First-Strand Synthesis System and an oligo d (T) primer according to the manufacturer's instruction. Primer was designed at the 5`upstream of the genes (Table 2).Analysis was carried out in a StepOnePlus Real-Time PCR system. The expression level of soybean *actin* gene (GenBank accession number: J01298) was used as a control and compared by the t-test using Statistica03J (StatSoft). The initial 30 sec denaturation at 94^{o}C was followed by 40 cycles of 3 sec denaturation at 94^{o}C, 30 sec annealing at 58^{o}C and 30 sec extension at 72^{o}C. A final 7 min extension at 72 ^{o}C completed the program (Rojas Rodas *et al*. 2014).

2.2.10 Functional analysis using NILs (Unpublished)

2.3 Results

2.3.1 HPLC and genetic analyses

The HPLC chromatogram of Nezumisaya and Harosoy results indicated that FGs of Nezumisaya had a glucose at the 2" position of glucose or galactose that is bound to the 3-position of kaempferol, whereas the glucose was absent in Harosoy. Conversely, FGs of Harosoy had a rhamnose at the 4" position of glucose or galactose that is bound to the 3-position of kaempferol, whereas the glucose was absent in Nezumisaya (Fig. 5)

Among the total of 91 RILs, 21 RILs had peaks of the Harosoy-type, and 23 RILs had peaks of the Nezumisaya-type. 17 RILs had a peak combination designated as 'type 3' that lacked three peaks, F1, F3 and F4 from the Harosoy-type. Further, 28 RILs had a peak combination designated as 'type 4' having peaks of a mixture of both cultivars in addition to a unique peak F13 (Fig. 5). Base on the results of the HPLC chromatogram of NH RILs F6 population (Rojas Rodas *et al.* unpublished) Two RILs

had unique peak distribution, suggesting that these RILs might have heterozygous allele(s). Excluding the RILs, the segregation fitted to 1:1:1:1 ratio (χ^2=2.82, P= 0.42), suggesting that two genes control the FG pattern.

2.3.2 SSR analysis

Among the 1025 SSR markers tested, 146 exhibited polymorphism between the parents and distinctly segregated in the RILs. A total of 99 markers fell into 28 linkage groups spanning 2172 cM (Rojas Rodas *et al.* unpublished).

2.3.3 Linkage mapping

Linkage mapping revealed that one gene which responsible for attachment of glucose to the 6″-position was located in the molecular linkage group I (chromosome 20) between Satt623 and Sat_420, in which Harosoy is dominant and Nezumisaya is recessive. Which had been found the genes *Fg1* (Rojas Rodas *et al.* unpublished). The other that responsible for attachment of glucose to the 2″-position was located in the molecular linkage group C2 (chromosome 6) between Satt307 and Sat_202, in which Nezumisaya is dominant and Harosoy is recessive.

2.3.4 QTL analysis (Unpublished)

2.3.5 Molecular cloning

Survey of the genome sequence of a US cultivar Williams 82 suggested that existence of a gene similar to GT gene, Glyma06g43380 between Satt307 and Sat_202 (Fig. 6). The entire coding region of Glyma06g43380 was amplified by PCR and cloned. Sequencing analysis revealed that ORF of Glyma06g43380 is 1380 bp

long encoding 459 amino acids. We designated the gene as *GmF3G2”Gt*. *GmF3G2"Gt* belongs to the family 1 glycosyltransferase, and it was designated as UGT79B30 by the UGT Nomenclature Committee (Mackenzie *et al*. 1997). BLAST analysis suggested that it had a 55 % amino acid similarity with *In3GG* gene of morning glory encoding anthocyanin 3-*O*-glucoside-2”-*O*-glucosyltransferase (Morita *et al*. 2005) (Fig. 7). Comparison with the genome sequence of Williams 82 suggested that *GmF3G2”Gt* had one intron. Eight nucleotides were polymorphic between Harosoy and Nezumisaya; six single-nucleotide substitutions and one two-nucleotides substitution. The polymorphism resulted in four amino acid substitutions between the cultivars (Fig. 7). The cDNA of Nezumisaya and Harosoy were designated as *GmF3G2”Gt-a* and *GmF3G2”Gt-b* respectively. The 5`upstream region of about 1.8kb was cloned of Nezumisaya and Harosoy by GenomeWalker Kit. Nucleotide sequence of 5`upstream region, exons and introns of Harosoy was identical with those from William 82. Contrast, Nezumisaya had many indels and substitutions in the 5`upstream region compare with Harosoy (Fig. 8).

2.3.6 dCAPS analysis

Bands with molecular size of about 230 bp were amplified with the dCAPS primers in Nezumisaya and harosoy (Fig. 9). *Kpn*I digestion generated a band of about 213 bp in Harosoy, whereas the PCR amplicon of Nezumisaya was unaffected. Banding pattern of the RILs was classified into Nezumisaya type, Harosoy type and heterozygous type having both bands. In total, 52 RILs had bands of Nezumisaya type, 40 RILs had bands of Harosoy type and 2 RILs had both bands (Fig. 9).

Banding patterns were co-segregated with FG patterns; RILs with FGs of Nezumisaya type and Type4 had bands of nezumisaya type, whereas RILs with FGs

of Harosoy type and Type3 had bands of only harosoy type (Fig. 9).

2.3.7 Gene expression

Survey of soybean genome sequence revealed the existence of a gene similar to *GmF3G2″Gt*, Glyma12g14050 with nucleotide similarity of 93 % in chromosome 12 (molecular linkage group H) and it was difficult to design primers that specifically amplify fragment of *GmF3G2″Gt* based on the nucleotide sequence of the coding region. So, we designed primers for real-time PCR based on sequences of 5' untranslated region distinct between the genes (Table 2). At the R6 stage, transcript level of *GmF3G2″Gt* in leaves of Harosoy was 14.6 % of that in Nezumisaya (t = 3.60*) (Fig. 10).

2.3.8 Functional analysis using NILs (Unpublished)

2.3.9 In vitro characterization of recombinant GmF3G2″Gt

The GmF3G2″Gt recombinant protein of Nezumisaya (GmF3G2″Gt-a) and Harosoy (GmF3G2″Gt-b) were expressed in E. coli as a His/ProS2 fusion and purified. After cleavage of the His/ProS2 tag, GmF3G2″Gt proteins were used for enzymatic assays.

GmF3G2″G-a and GmF3G2″G-b converted kaempferol 3-O-glucoside to kaempferol 3-O-sophoroside as confirmed by comparison of retention time, UV spectra and MS/MS ionization with the standard compound (Fig 13). GmF3G2″Gt-a showed a broad activity for kaempferol/quercetin 3-O-glucoside/galactoside derivatives.

However, GmF3G2″Gt-a did not glucosylate kaempferol 3-*O*-rhamnosyl-(1 →4)-

[rhamnosyl-(1 → 6)-glucoside] and 3-*O*-rhamnosyl-(1 → 4)-[glucosyl-(1 →6)-glucoside]. GmF3G2″Gt-a had a higher preference for UDP-glucose than UDP-galactose, with only 3 % activity relative to that for UDP-galactose. No UGT activity was detected for UDP-arabinose and UDP-glucuronic acid. GmF3G2″Gt-b also showed similar substrate specificity (Table S1). Accordingly, GmF3G2″Gt-a and GmF3G2″Gt-b were defined as flavonol 3-*O*-glucoside/galactoside (1→2) glucosyltransferases.

2.4 Discussion

Soybean cultivars, Harosoy and Nezumisaya have gray pubescence and primarily deposit kaempferol derivatives in leaves. However, structure of FG components was distinct. HPLC analysis and comparison with authentic samples suggested that FGs having glucose at the 2”-positon of glucosr or galactose thai is bound to the 3-positon of kaemoferol were present in Nezumisay, whereas FGs of Harosoy were devoid of 2”-glucose. Conversely, FGs having glucose at the 6”-positon of glucosr or galactose thai is bound to the 3-positon of kaemoferol were present in Harosoy whereas FGs of nezumisaya were absent. Apigenin 7-O-glucoside almost exclusively deposits in the cytoplasm of gray pubescence (Iwashina *et al*. 2006), suggested that the peak F8 may have been derived from pubescence on leaves.

From the genetic analysis of NH RILs, One of the genes may be responsible for attachment of glucose to the 2”-position, and probably encodes flavonol 3-O-glucoside (1-2) glucosyltransferase. Nezumisaya may have a dominant, and Harosoy may have a recessive allele of the genes. The gene was mapped in molecular linkage groupC2 (chromosome 6) between Satt307 to Sat_202. Judging from relative location with SSR markers, its position was similar to that of *Fg3* reported by Buzzell (1974).

QTLs analysis indicate that existence of peak components existened F1, F3, F4 and F13 were mapped in linkage group I (chromosome 20), near the marker GMGLPSI2, which was similar as the previous results (Rojas Rodas *et al*. 2014).The components of peak F1, F3 and F4 are only existened in Harosoy and all had 6”-glucose, so this gene is similar like Fg2. And existence of peak components from F2, F5, F6, F9, F10 and F11 were mapped in linkage group C2 (chromosome 6), near the marker SATT202. Peak F9 and F10 are only existened in Nezumisaya and all had 6”-glucose, so this gene is similar like Fg2. For the peak F5 and F6, they all existed in both Harosoy and Nezumisaya, and it is the upstream products of F9 and F10. Maybe because of the function of FG gene, F5 and F6 transform into F9 and F10. (Table 3)

Survey of the genome sequence of a US Williams 82 suggested that a gene similar to GT gene, Glyma06g43880 existed between Satt307 and Sat_202. We cloned the cDNA clone and designated as *GmF3G2”Gt.* The coding region to the *GmF3G2”Gt* is 1380 bp long encoding 459 amino acids. *GmF3G2”Gt* belongs to the 1,2/1,6 branch-forming glucosyltransferase gene cluster. *GmF3G2”Gt* of Nezumisaya had amino acid similarity of 55% either the *InGGT* of morning glory encoding 3-O-gluscoside (1-2) glucosyltransferase that also attaches glucose to the 2”-position of glucose that is bound to the 3-position of anthocyanidins (Morita *et al*. 2005). Genome sequence of Williams 82 suggest existence of a gene Glyma12g14050 having 93% of nucleotide identity with Glyma06g43880 in chromosome 12, However, the gene responsible for attachment of glucose to the 2”-position was not mapped to chromosome 12 in the current as well as previous study (Buzzell. 1974). So Glyma12g14050 may not be responsible for attachment of glucose to the 2”-positon in experimental materials analyzed so far.

In the coding region of *GmF3G2"Gt*, eight nucleotides were polymorphic between

Harosoy and Nezumisaya, resulting in four amino acid substitutions. The dCAPs marker to discriminate the two-base substitution was co-segregated with FG patterns, suggesting that *GmF3G2″Gt* might correspond to a flavonol 3-*O*-glucoside (1→2) glucosyltransferase (*Fg3*) gene. The recombinant *GmF3G2″Gt-a* protein converted UDP-glucose and kaempferol 3-*O*-glucoside or kaempferol 3-*O*-robinobioside to kaempferol 3-*O*-sophoroside or kaempferol 3-*O*-glucosyl-(1→2)-[rhamnosyl-(1→6)-galactoside], proving that *GmF3G2″Gt* encodes flavonol 3-*O*-glucoside (1→2) glucosyltransferase (analyzed by RIKEN). The transcript level of *GmF3G2″Gt* in leaves of Harosoy was about 15 % of that in Nezumisaya. Indels and substitutions in the 5' upstream region may be responsible for the differences in expression level. Promoter assays may be necessary to determine which polymorphism is critical for gene expression. GmF3G2″Gt protein may have a threshold for catalysis and that of Harosoy may be lower than the threshold.

The nucleotide sequence of the 5′ upstream region, exons and intron of *GmF3G2″Gt* from Harosoy was identical with that from Williams 82. The HPLC chromatogram of leaf methanol extracts of Williams 82 was same with that of US cultivar Clark having allelic combination of *fg1Fg2fg3Fg4* (Buzzell and Buttery. 1998), suggesting that Williams 82, Clark and Harosoy might have a recessive allele at the *Fg3* locus. The results were consistent with our assumption that Nezumisaya has a dominant and Harosoy has a recessive allele for the *Fg3* gene. Based on the HPLC profiles, Harosoy may have the allelic combination of *Fg1Fg2fg3Fg4* whereas Nezumisaya have the combination of *fg1Fg2Fg3Fg4*. It is consistent with the previous report that Harosoy has allelic combination of either *Fg1Fg2fg3* or *fg1Fg2fg3* depending on the line (Buzzell and Palmer. 1985).

Cloning of *Fg1* encoding flavonol 3-*O*-glucoside (1-6) glucosyltransferase and

Fg4 encoding flavonol 3-*O*-glucoside (1-2) rhamnosyltransferase may help clarity the mechanisms of site-specificity (at 2”- or 6”-position) and suger specificity (UDP-glucose or UDP-rhamnose) in the attachment of sugar to suger in flavonols, and molecular evolution of the genes. In addition, another gene having function of flavonol 3-*O*-glucoside (1-4) rhamnosyltransferase remains to be cloned. Cloning of the entire FG gene set may be necessary to get the entire picture of FG biosynthesis in soybean.

…… (Unpublished data)

Table 2. PCR primers used in this study.

Purpose	Target gene	Forward primer (5′-3′)	Reverse primer (5′-3′)
Cloning of cDNA for sequencing	*GmF3G2"Gt*	AATCACGTCCTTTGCACATA	CAAAACCATTGCCATTCATG
Sequencing of cDNA	*GmF3G2"Gt*	AGGCTGTGCACTATTGTACT CCATATATGGATTACATTGG TTCTTGCACACCCTTCTGTA	TCACAATGCTCACAGCTTTA CACTTCCAAAGCAACAATAA GACAGAGCTAGCAGTACAAT
Cloning of 5' upstream region[a]	*GmF3G2"Gt*		GTGTGTGTCTCTGGTGTAGTAATCATG CTCACACAAAACAGAGGTTCAATACGAGG
Sequencing of 5' upstream region	*GmF3G2"Gt*	ACAACGGTCATATATGCATG ACAACGGTCATTTCTGCATG GTCAACAATGAAGGACATGG	TGGCGGAGAGGAAGAAGATA CTCTCTGAAATGGATCCCAA TAGAAATGCACGTAGTAGGT TCCACCACAATATGCTATGT
dCAPS analysis	*GmF3G2"Gt*	CCAAACTCACGAGGCTGGT	TCCTCCAAATCTAAAGTTGG
	GmF3G6"Gt	CCGGAGCTCATGTCTTGTGAAGTTGTGAAC	GCCGAAGCTTCTAAGCCATAGACTTCAACT
Cloning for functional analysis[b]	*GmF3G2"Gt*	GCCGGAGCTCATGAAATCACGTCCTTTGC	GCCGCTCGAGTCAAATTCCCTCTACAATCTC
Real-time PCR	*GmF3G2"Gt*	GATCTCCGTTACTCTCGTTT	CATGGCTTTGTGGAAACAAG
	actin	GTCCTTTCAGGAGGTACAACC	CCACATCTGCTGGAAGGTGC

[a]Gene-specific primers for genome walking.

[b]Restriction sites incorporated in primers, *Sac*I site in forward primer and *Xho*I site in reverse primer are underlined.

Table 3. QTLs responsible for areas of HPLC peak F1-F13 observed in a F6 RILs population cross between Nezumisaya and Harosoy (Unpublished)

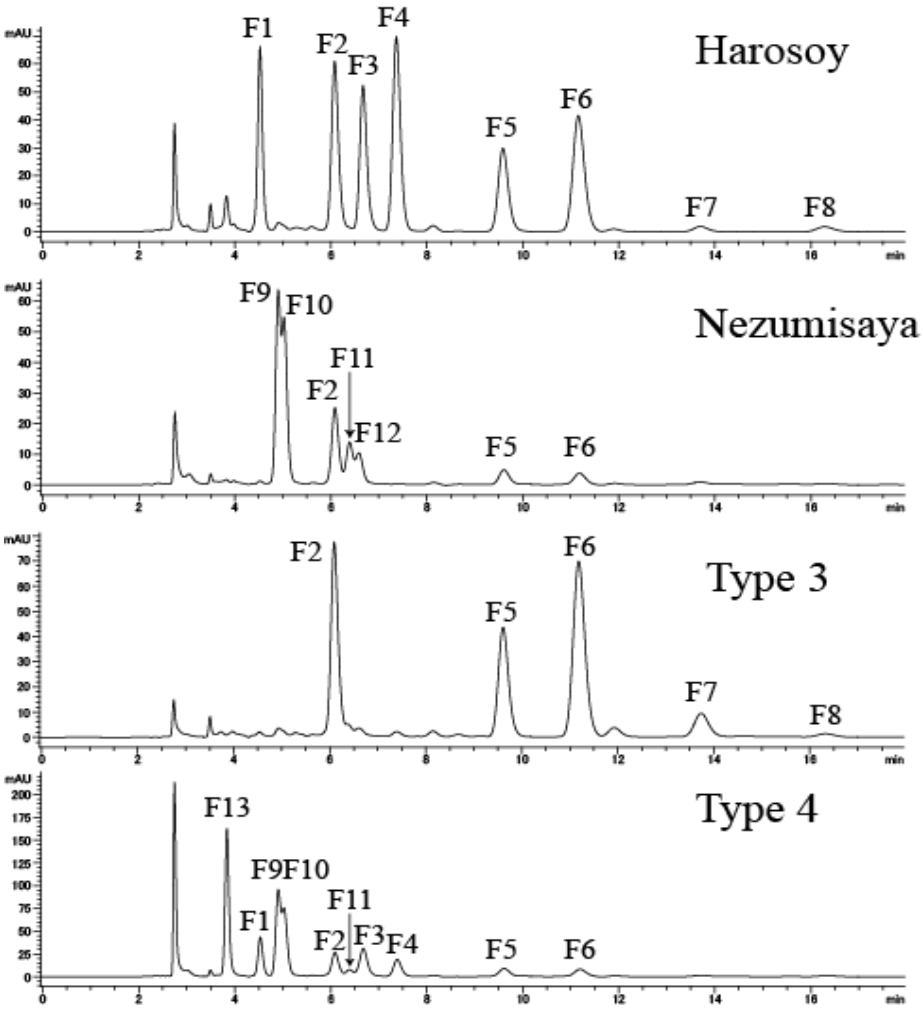

Figure 5

HPLC chromatogram of MeOH extracts from leaves of soybean cultivars Nezumisaya, Harosoy and recombinant inbred lines derived from a cross of the cultivars, and chemical structure of flavonol glycosides corresponding to HPLC peaks. (A) HPLC chromatogram. 100 mg of trifoliolate leaves was extracted with 1 ml of MeOH. Eluent: phosphoric acid/acetonitrile/H_2O (0.2:18:82). Flow-rate: 1.0 ml/min. Injection: 10 μl. Detection: 350 nm.

F1, kaempferol 3-*O*-rhamnosyl-(1→4)-[glucosyl-(1→6)-galactoside];

F2, kaempferol 3-*O*-rhamnosyl-(1→4)-[rhamnosyl-(1→6)-galactoside];

F3, kaempferol 3-*O*-glucosyl-(1→6)-galactoside;

F4: kaempferol 3-*O*-glucosyl-(1→6)-glucoside;

F5, kaempferol 3-*O*-rhamnosyl-(1→6)-galactoside;

F6, kaempferol 3-*O*-rhamnosyl-(1→6)-glucoside;

F7, kaempferol 3-*O*-glucoside;

F8, apigenin 7-*O*-glucoside;

F9, kaempferol 3-*O*-glucosyl-(1→2)-[rhamnosyl-(1→6)-galactoside];

F10, kaempferol 3-*O*-glucosyl-(1→2)-[rhamnosyl-(1→6)-glucoside];

F11, kaempferol glycoside;

F12, kaempferol 3-*O*-glucosyl-(1→2)-glucoside; F13, unidentified kaempferol glycoside.

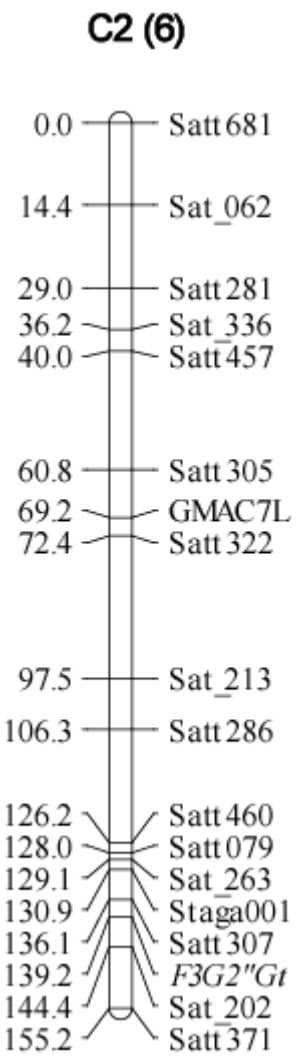

Figure 6.

Linkage mapping of a gene (*F3G2"Gt*) responsible for attachment of glucose to the 2″-position of glucose or galactose that is bound to the 3-position of kaempferol using recombinant inbred lines derived from a cross between soybean cultivars Nezumisaya and Harosoy. The name of the linkage groups is indicated at the top followed by the chromosome number in parenthesis. Distances (cM) of markers from the top of the linkage group are shown on the left.

A

B

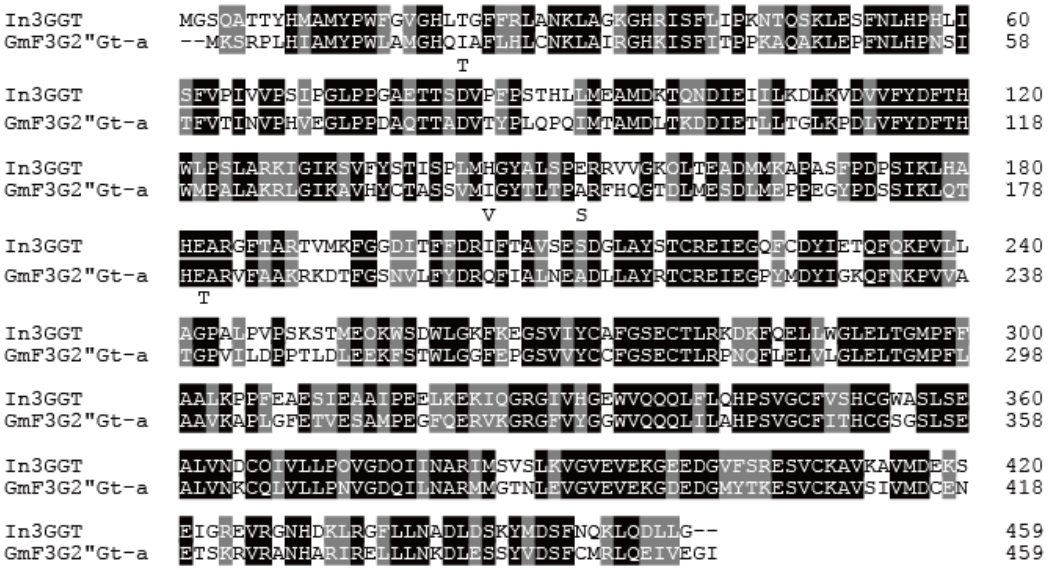

Figure 7

Gene structure and amino acid alignment of *GmF3G2"Gt*. (A) Intron-exon structure of *GmF3G2"Gt* gene. (B) Amino acid alignment of soybean *GmF3G2"Gt-a* and the morning glory *In3GG* encoding anthocyanin 3-*O*-glucoside (1→2) glucosyltransferase (accession number: Q53UH4). Identical amino acids are indicated in black, similar amino acids in gray. Four amino acids that differed in *GmF3G2"Gt-b* are exhibited below the sequence of Nezumisaya.

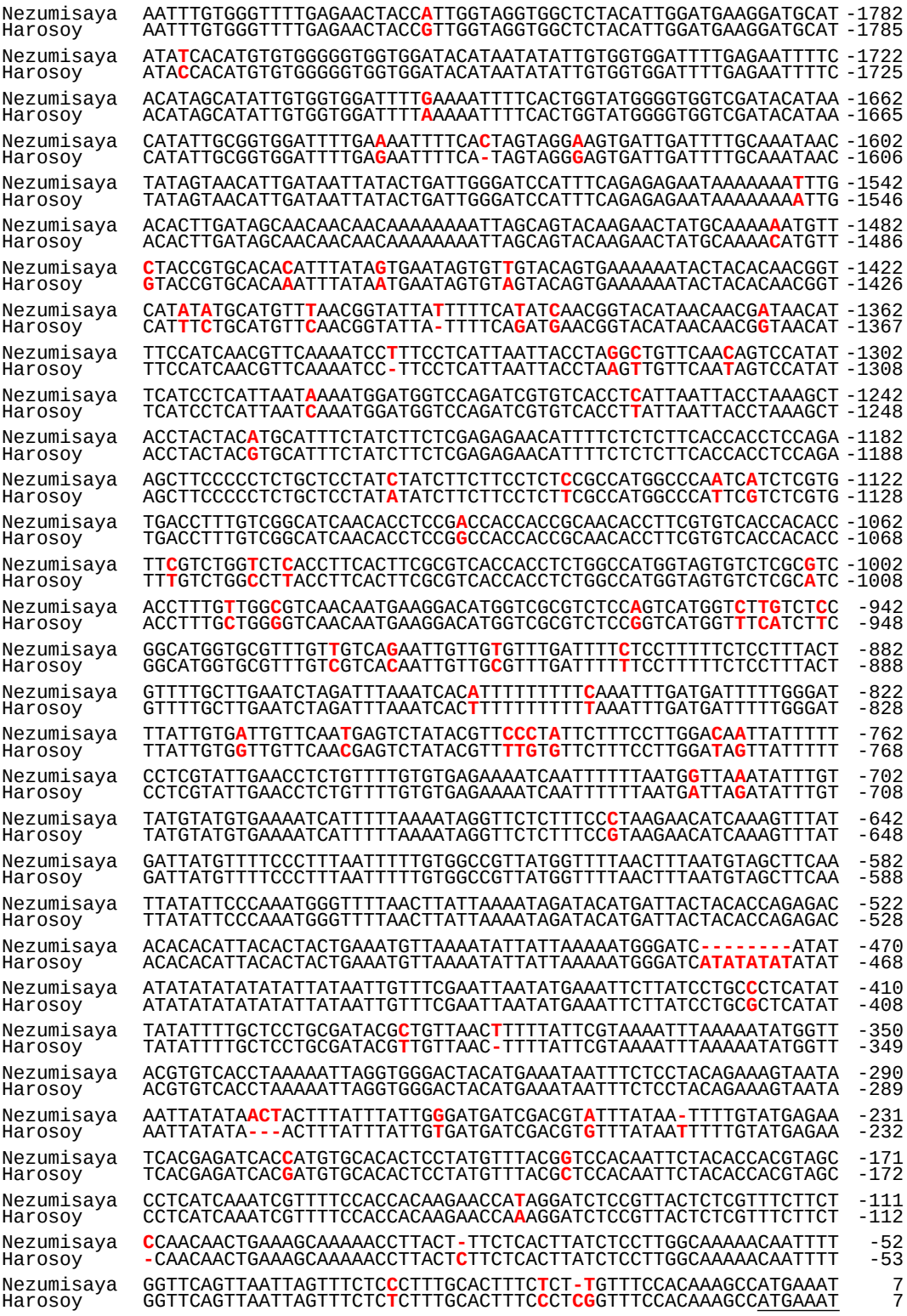

Figure 8

Alignment of the 5’ upstream region of *GmF3G2"Gt* gene in soybean cultivars Harosoy and Nezumisaya. Polymorphic nucleotides are shown in red font. Coding sequence is underlined.

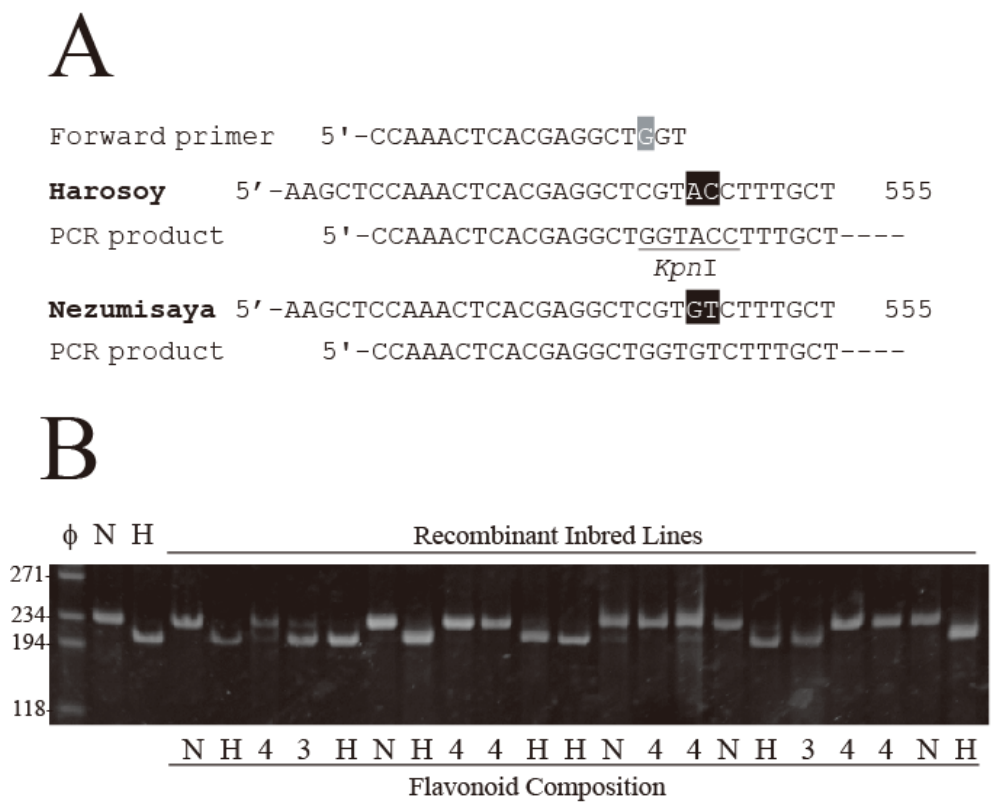

Figure 9

Outline and results of dCAPS analysis of *GmF3G2"Gt* in soybean. (A) Schematic presentation of dCAPS analysis. Partial nucleotide sequences around the region polymorphic between Harosoy and Nezumisaya are exhibited. A mismatched base in the forward primer is highlighted in gray. Polymorphic nucleotides are highlighted in black. (B) Results of dCAPS analysis of the parents and the recombinant inbred lines derived from a cross between Nezumisaya and Harosoy. PCR products amplified with dCAPS primers were digested by *Kpn*I and the digests were separated on an 8 % polyacrylamide gel. ϕ, molecular marker ϕx174/*Hae*III; N, Nezumisaya; H, Harosoy. FG pattern of the recombinant inbred lines is exhibited below the gel. H, Harosoy-type; N, Nezumisaya-type, 3, type 3; 4, type 4. The migration of size markers (bp) is shown to the left of the gel.

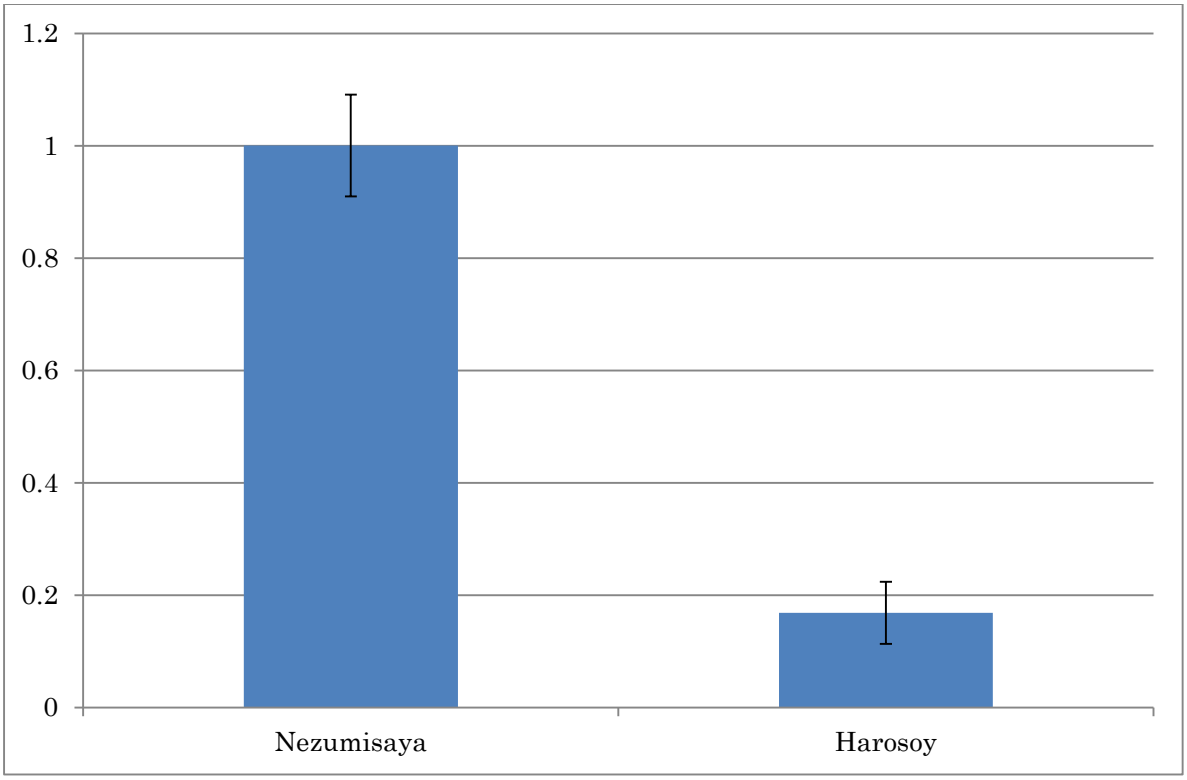

Figure 10

Relative gene expression (mean±SD) of *GmF3G2"Gt* in leaves of Nezumisaya and Harosoy at R6 stage.

Figure 11

Agronomic characters (mean±SD) in NILs for *Fg3*, F7-50 and F8-100. **(Unpublished)**

Figure 12

Agronomic characters (mean±SD) in NILs for *Fg1*, F8-17 and F8-98. **(Unpublished)**

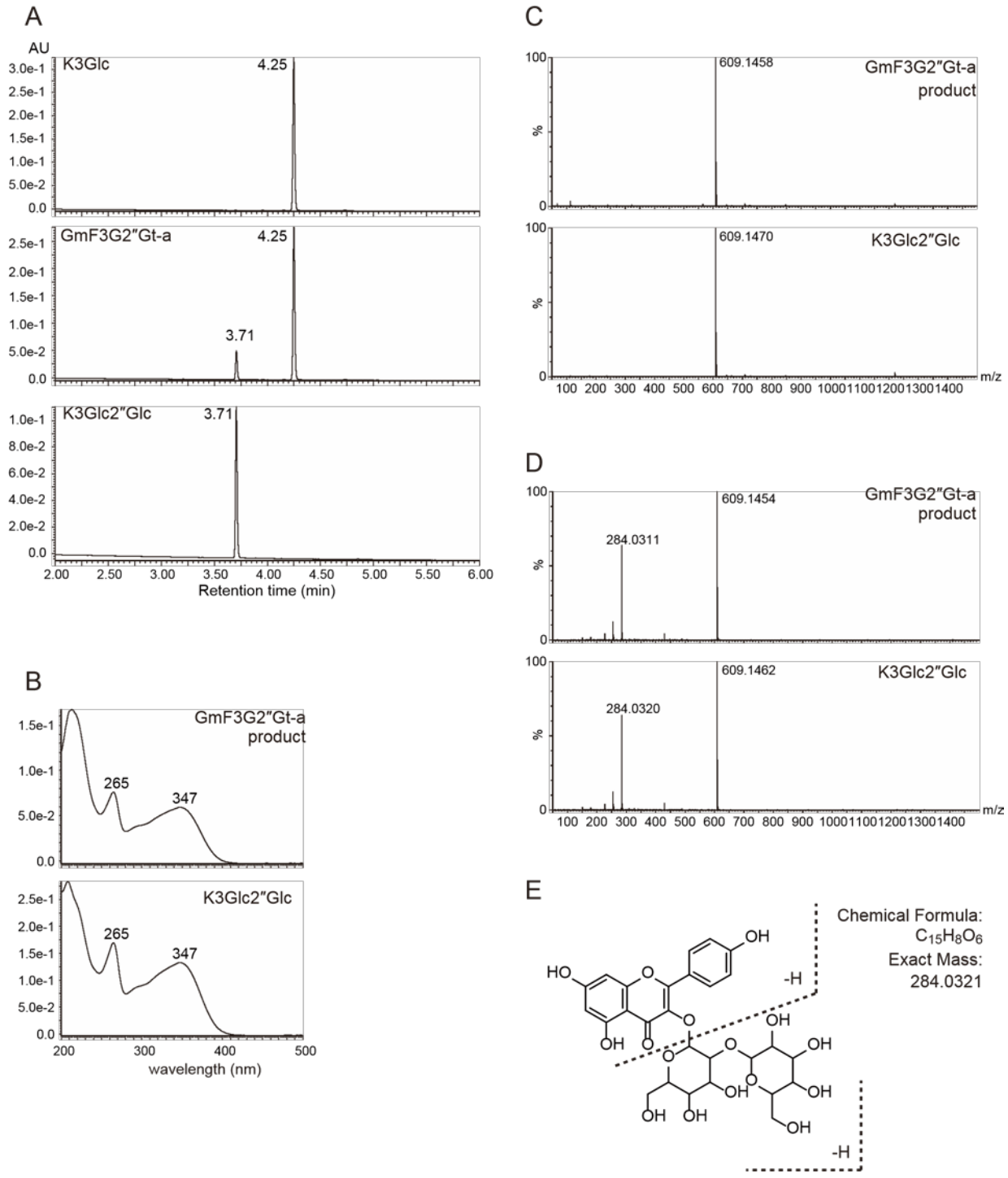

Figure 13

Identification of reaction product of GmF3G2"Gt-a (cultivar Nezumisaya). (A) Elution profiles of the standards (kaempferol 3-*O*-glucoside and kaempferol 3-*O*-sophoroside) and reaction product of GmF3G2"Gt-a protein. (B) UV spectra of the standard

(kaempferol 3-*O*-sophoroside) and reaction product of GmF3G2"Gt-a protein. Mass spectra (C) and MS/MS spectra (D) of the standard (kaempferol 3-*O*-sophoroside) and reaction product of GmF3G2"Gt-a protein. E, The MS/MS fragmentation for kaempferol 3-*O*-sophoroside. K3Glc, kaempferol 3-*O*-glucoside; K3Glc2"Glc, kaempferol 3-*a*-sophoroside.

CHAPTER 3

Linkage mapping and molecular cloning of FG gene segregated in T005 × Clark population (Unpublished)

CHAPTER 4

Allelic variation of flavonol glycoside genes (Unpublished)

CHAPTER 5

General Discussion (Unpublished)

Summary

We cloned two FG genes, *Fg3* gene and the third allele of *Fg2* gene. The *Fg3* gene was identified from the RILs developed from a cross between cultivars, Nezumisaya and Harosoy. FGs were separated by HPLC and identified based on comparison with authentic samples. Segregation of the RILs suggested that two genes control FG composition and that one of the genes is responsible for attachment of glucose to the 2″-position. Based on SSR analysis, linkage mapping and genome database survey, we cloned a candidate gene designated as *GmF3G2″Gt*. The coding region of *GmF3G2″Gt* is 1380 bp long encoding 459 amino acids. The *GmF3G2″Gt* recombinant protein of Nezumisaya converted UDP-glucose and kaempferol 3-*O*-glucoside to kaempferol 3-*O*-sophoroside. These results indicate that *GmF3G2″Gt* encodes a flavonol 3-*O*-glucoside (1→2) glucosyltransferase and corresponds to *Fg3* gene. *GmF3G2″Gt* was designated as UGT79B30 by the UGT Nomenclature Committee. To investgate the relationship between FG genes and agronomic characters, we developed two sets of NILs for *Fg1* and *Fg3*. Only one set of NILs for *Fg3* had significant difference between Nezumisaya-type and Harosoy-type in photosynthesis rate, chlorophyll contents, plant height, branch number and node number. Only one set of NILs for *Fg1* had significant differences in chlorophyll contents, branch mount and 100-seed weight.

The third allele of *Fg2* gene was cloned and characterized from F_2 plants derived from a cross between cultivar Clark and *Glycine soja* accession T005 having *Fg2-b* allele. Leaves of Clark having *Fg2-a* allele contained seven FGs, whereas leaves of T005 contained eight FGs. HPLC chromatogram of F_2 plants was classified into Clark-type and

T005-type at a 3:1 ratio, suggesting that a single gene controls FG composition. SSR analysis and linkage mapping revealed that the gene was located in molecular linkage group O (chromosome 10) between Satt123 and Satt592 that was close to the position of *Fg2* gene. The nucleotide sequence of the coding region of *GmF3G"Rt* of T005 was identical with that of Clark except for a SNP at nucleotide position 20. The SNP generated amino acid polymorphism at amino acid position 7. The dCAPS marker to discriminate the SNP co-segregated with the FG composition in the F_2 population. These results suggest that the amino acid substitution might be responsible for FG composition.

Based on HPLC analysis of leaf extracts from 79 Japanese soybean core collections, 80 foreign soybean core collections and 35 other germplasms including soybean and *G.soja*, FGs pattern were classified into 12 different flavonoid types. Flavonoid composition and genetic analysis suggested that the 12 patterns may be explained by combination of alleles of *Fg1* to *Fg5* genes. We propose a new scheme for FG biosynthetic pathway.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my academic advisor Prof. Takahashi for his guidance and encouragement for my experiments and help during my stay in Japan.

I want to give my thanks to Prof. Otobe, Prof. Tanaka and Prof. Ohsawa for their guidance for my experiments and thesis writing.

I would like to express special thanks to Dr. Felipe Rojas Rodas and Dr. Tito O. Rodriguez for their guidance and help; and the laboratory member Ms. Fan Yan, Ms. Auchithya Dissanayaka and Mrs. Eiko Izumi, for their help in lab and field.

I am grateful to Dr. Yoshinori Murai and Dr. Tsukasa Iwashina (Tsukuba Botanical Garden) for their help and identifying the flavonoid samples. And I am also grateful to Dr. Yokenura-Sakakibara (RIKEN) and her group for excellent work for functional analysis.

Finally, I want to thank the scholarship from the China Scholarship Council for the support of this study.

REFFRENCES

Abeynayake S. W., Panter S., Chapman R., Webster T., Rochfort S., Mouradov A. 2012. Biosynthesis of proanthocyanidins in white clover flowers: cross talk within the flavonoid pathway. Plant Physiol. 158 666–678

Agati G., Brunetti C., Di Ferdinando M., Ferrini F., Pollastri S., Tattini M. 2013. Functional roles of flavonoids in photoprotection: New evidence, lessons from the past. Plant Physiol. Biochem. 72 35–45

Mouradov A and Spangenberg G. 2014. Flavonoids: a metabolic network mediating plants adaptation to their real estate Front. Plant Sci. 5: 620.

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 25:3389-3402.

Basten C.J. Weir B.S. and Zeng Z.B. 2001. QTL cartographer, version 2.0. Department of statistics, North Carolina State University, Raleigh, USA

Bidart-Bouzat M. G., Kliebenstein D. J. 2008. Differential levels of insect herbivory in the field associated with genotypic variation in glucosinolates in Arabidopsis thaliana. J. Chem. Ecol. 34 1026–1037

Bueno-Silva. B, Alencar. S.M, Koo. H, Ikegaki. M, Silva. G.V, Napimoga. M.H, Rosalen. P.L. 2013. Anti-inflammatory and antimicrobial evaluation of neovestitol and vestitol isolated from Brazilian red propolis. J. Agric. Food Chem. 61, 4546–4550.

Buttery B.R, Buzzell R.I. 1973. Varietal Differences in Leaf Flavonoids of Soybeans. Crop Sci. 13:103-106.

Buttery, B.R., and Buzzell R.I. 1975. Soybean Flavonol Glycosides - Identification and Biochemical Genetics. Canadian Journal of Botany-Revue Canadienne De Botanique. 53: 219-224.

Buttery, B. R., and Buzzell R. I.. 1976. Flavonol Glycoside Genes and Photosynthesis in Soybeans. Crop Sci. 16: 547-550.

Buzzell, R. I. 1974. Soybean linkage tests. Soybean Genetics Newsletter 1: 11-14.

Buzzell, R. I., and. Buttery B. R. 1974. Flavonol Glycoside Genes in Soybeans. Canadian Journal of Genetics and Cytology. 16: 897-899.

Buttery, B.R. and Buzzell, R.I. 1988. Soybean leaf nitrogen in relation to photosynthetic rate and yield. Canadian Journal of Plant Science, 68, 793-795.

Buzzell, R. I., and B. R. Buttery. 1998. Genetics of leaf waviness in soybean. Soybean

Genetics Newsletter. 25: 23-24.

Buzzell R.I, Buttery B.R. 1992. Inheritance of an Anomalous Flavonol Glycoside Gene in Soybean. Genome. 35:636-638.

Buzzell, R. I., and R. G. Palmer. 1985. Soybean linkage group 1 tests. Soybean Genetics Newsletter. 12: 32-33.

Cesco S., Mimmo T., Tonon G., Tomasi N., Pinton R., Terzano R., 2012. Plant-borne flavonoids released into the rhizosphere: impact on soil bio-activities related to plant nutrition. A review. Biol. Fertil. Soils 48 123–149

Chan K. F., Zhao Y. Z., Chow T. W. S., Yan C. S. W., Ma D. L., Burkett B. A., 2009. Flavonoid dimers as bivalent modulators for P-glycoprotein-based multidrug resistance: structure-activity relationships. Chemmedchem. 4 594–614

Crozier, I. B. Jaganath, M. N. Clifford, 2006. In Plant Secondary Metabolites: Occurrence, Structure and Role in the Human Diet (Eds.: A. Crozier, M. N. Clifford, H. Ashihara), Blackwell Publishing Ltd: Oxford,

Cregan, P. B., T. Jarvik, A. L. Bush, R. C. Shoemaker, K. G. Lark .1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39: 1464-1490.

Fehr, W. R., C. E. Caviness, D. T. Burmood and Penningt.JS. 1971. Stage of

Development Descriptions for Soybeans, *Glycine-Max* (L) Merrill. Crop Sci. 11: 929

Ferreyra M. L. F., Rius S. P., Casati P. 2012. Flavonoids: biosynthesis, biological functions, and biotechnological applications. Front. Plant Sci. 3:222

Githiri SM, Yang D, Khan NA, Xu D, Komatsuda T, Takahashi R. 2007. QTL analysis of low temperature induced browning in soybean seed coats. J Hered. 98(4):360-6.

de Groot H, Rauen U. 1998. Tissue injury by reactive oxygen species and the protective effects of flavonoids. Fundam Clin Pharmacol. 12:249–55.

Grotewold E. 2006. The genetics and biochemistry of floral pigments. Annu. Rev. Plant Biol. 57 761–780

Guo J., Han W., Wang M. H. 2008. Ultraviolet and environmental stresses involved in the induction and regulation of anthocyanin biosynthesis: a review. Afr. J. Biotechnol. 7 4966–4972.

Hancock K., Collette V., Chapman E., Hanson K., Temple S., Moraga R.. 2014. Progress towards developing bloat-safe legumes for the farming industry. Crop Pasture Sci. 65 1107–1113

Hernandez I. and Breusegem F.V. 2012. Opinion on the possible role of flavonoids as energy escape valves: novel tools for natures's Swiss army knife? Plant Sci. 179 297-

Hisano, H., S. Sato, S. Isobe, S. Sasamoto, T. Wada .2007. Characterization of the soybean genome using EST-derived microsatellite markers. DNA Res. 14: 271-281.

Heim KE, Tagliaferro AR, Bobilya DJ. 2002. Flavonoid antioxidants: chemistry, metabolism and structure-activity relationships. J Nutr Biochem. 13(10):572-584.

Iwashina, T., E. R. Benitez and R. Takahashi. 2006. Analysis of flavonoids in pubescence of soybean near-isogenic lines for pubescence color loci. J. Hered. 97: 438-443.

Iwata H. and Ninomiya S. 2006. AntMap: constructing genetic linkge maps using an ant colony optimization algorithm. Breed Sci 56:371-377

Kelly E. Heim, Anthony R. Tagliaferro, Dennis J. Bobilya. 2002. Flavonoid antioxidants: chemistry, metabolism and structure-activity Relationships. Journal of Nutritional Biochemistry. 13 572–584

Koes R.E., Quattrocchio, R. and Mol, J. N. M. 1994. The flavonoid biosynthetic pathway in plants: function and evolution. BioEssays. 16: 123-132

Lairson, L. L., B. Henrissat, G. J. Davies and S. G. Withers. 2008. Glycosyltransferases: Structures, functions, and mechanisms. Annu. Rev. Biochem. 77: 521-555.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly et al., 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181.

Liu-Gitz L., S. J. Britz & W. P. Wergin. 2000. Blue light inhibits stomatal development in soybean isolines containing kaempferol-3-O-2-glycosyl-gentiobioside (K9), a unique flavonoid glycoside. Plant, Cell and Environment 23, 883–891

Mackenzie, P. I., I. S. Owens, B. Burchell, K. W. Bock, A. Bairoch. 1997. The UDP glycosyltransferase gene superfamily: Recommended nomenclature update based on evolutionary divergence. Pharmacogenetics 7: 255-269.

Manach C., A. Scalbert, C. Morand, L. Jimene. 2004. Polyphenols: Food sources and bioavailability. American Journal of Clinical Nutrition. 79 (2004), pp. 727–747

Maggioni D, Biffi L, Nicolini G, Garavello W. 2014. Flavonoids in oral cancer prevention and therapy. Eur J Cancer Prev. 2014 Dec 17.

Masada S, Terasaka K, Mizukami H. 2007. A single amino acid in the PSPG-box plays an important role in the catalytic function of CaUGT2 (Curcumin glucosyltransferase), a Group D Family 1 glucosyltransferase from *Catharanthus roseus.* FEBS Lett. 581 (14) : 2605-10.

Middleton EJ. 1998. Effect of plant flavonoids on immune and inflammatory cell function. Adv Exp Med Biol. 439:175–82.

Morita, Y., A. Hoshino, Y. Kikuchi, H. Okuhara, E. Ono .2005. Japanese morning glory dusky mutants displaying reddish-brown or purplish-gray flowers are deficient in a novel glycosylation enzyme for anthocyanin biosynthesis, UDP-glucose : anthocyanidin 3-O-glucoside-2 "-O-glucosyltransferase, due to 4-bp insertions in the gene. Plant J. 42: 353-363.

Murai, Y., R. Takahashi, F. R. Rodas, J. Kitajima and T. Iwashina. 2013. New Flavonol Triglycosides from the Leaves of Soybean Cultivars. Natural Product Communications 8: 453-456.

Murray MG, Thompson WF. 1980. Rapid Isolation of High Molecular-Weight Plant DNA. Nucleic Acids Research. 8:4321-4325.

Nijveldt RJ, van Nood E, van Hoorn DE, Boelens PG, van Norren K, van Leeuwen PA. 2001. Flavonoids: a review of probable mechanisms of action and potential applications. Am J Clin Nutr. 74(4):418-25.

Phang, T.H., Shao, G.H. and Lam, H.M. 2008 Salt tolerance mechanisms in soybean. J. Integr. Plant. Biol. 50: 1196–1212.

Paquette, S., B. L. Moller and S. Bak. 2003. On the origin of family 1 plant

glycosyltransferases. Phytochemistry 62: 399-413.

Pati S., Liberatore M. T., Gambacorta G., Antonacci D., La Notte E. 2009. Rapid screening for anthocyanins and anthocyanin dimers in crude grape extracts by high performance liquid chromatography coupled with diode array detection and tandem mass spectrometry. J. Chromatogr. A 1216 3864–3868

Rice-Evans CA, Miller NJ, Paganga G. 1996. Structure-antioxidant activityrelationships of flavonoids and phenolic acids. Free Radic Biol Med. 20:933–56.

Rojas Rodas, F., T. O. Rodrigues, Y. Murai, T. Iwashina, S. Sugawara .2014. Linkage mapping, molecular cloning and functional analysis of soybean gene *Fg2* encoding flavonol 3-*O*-glucoside (1-6) rhamnosyltransferase. Plant Mol Biol 84: 287-300.

Saito K., Yonekura-Sakakibara K., Nakabayashi R., Higashi Y., Yamazaki M., Tohge T., et al., 2013. The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. Plant Physiol. Biochem. 72 21–34

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

Song, Q. J., L. F. Marek, R. C. Shoemaker K. G. Lark, V. C. Concibido .2004. A new integrated genetic linkage map of the soybean. Theor. Appl. Genet. 109: 122-128.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei .2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. 28: 2731-2739.

Tanaka Y., Ohmiya A. 2008. Seeing is believing: engineering anthocyanin and carotenoid biosynthetic pathways. Curr. Opin. Biotechnol. 19 190–197

Toda K, Takahashi R, Iwashina T, Hajika M. 2011. Difference in chilling-induced flavonoid profiles, antioxidant activity and chilling tolerance between soybean near-isogenic lines for the pubescence color gene. J Plant Res. 124:173-182.

Weston L. A., Mathesius U. 2013. Flavonoids: their structure, biosynthesis and role in the rhizosphere, including allelopathy. J. Chem. Ecol. 39 283–297

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

Winkel-Shirley B. 2001b. It takes a garden. How work on diverse plant species has contributed to an understanding of flavonoid metabolism. Plant Physiol. 127 1399–1404

Wolfender JL, Rudaz S, Choi YH, Kim HK. 2013. Plant metabolomics: from holistic data to relevant biomarkers. Curr Med Chem. 20(8):1056-90

Yonekura-Sakakibara, K., and K. Hanada. 2011. An evolutionary view of functional diversity in family 1 glycosyltransferases. Plant J. 66: 182-193.

Yonekura-Sakakibara K, Nakabayashi R, Sugawara S, Tohge T, Ito T, Koyanagi M, Kitajima M, Takayama H, Saito K. 2014. A flavonoid 3-O-glucoside:2''-O-glucosyltransferase responsible for terminal modification of pollen-specific flavonols in Arabidopsis thaliana. Plant J, 79:769-782.