Molecular Genetic Studies on Flavonol Glycoside Biosynthesis in Soybean

January 2014

Felipe ROJAS RODAS

Molecular Genetic Studies on Flavonol Glycoside Biosynthesis in Soybean

A Dissertation Submitted to

the Graduate School of Life Environmental Sciences

the University of Tsukuba

in Partial Fulfillment of the Requirements

For the Degree of Doctor in Philosophy in Agricultural Sciences,

(Doctoral Program in Advanced Agricultural Technology and Sciences)

Felipe ROJAS RODAS

TABLE OF CONTENTS

CHAPTER I

1. Introduction

Flavonoids (from the Latin word *flavus* "Yellow") is the term to designate a large family of low molecular weight polyphenolic secondary metabolites that are widespread throughout the plant kingdom from mosses to angiosperms (Koes *et al*., 1994). They have been widely studied during the last decades because of their role in plants and benefits to human health (Harborne, 1986; Andersen and Markham, 2007). Flavonoids occur naturally, in fruit, vegetables, nuts, seeds, flowers and bark (Cook and Samman, 1996). They can be found in both the free form (aglycones) and as glycosides, differing in their substituents (type, number and position) and insaturation (Venketeshwer, 2012).

Flavonoids are an important part in human diet and have been intensively studied during more than two decades due to their potential health benefits. Mainly, they can be found in vegetables and fruits but the relevant concentrations are found in broccoli, soybean, green and black tea, wine and red fruits (Garcia-Tirado, *et al*., 2012). A high dietary intake can be beneficial to health (Chen and Chen, 2013). In animal models and *in vitro*, flavonoids are related to antioxidant or free radical scavenging activity (Kar, 2007; Hsue-Yin *et al*., 2008). Many of them display cancer-fighting potential, inhibition of vascular endothelial growth factor (VEGF) expression, cancer cell proliferation and

angiogenesis (Luo, *et al*., 2008; Chen and Chen, 2013), capacity to regulate cell enzymatic activity, induce apoptosis, promote cell differentiation, regulate the immune response and inhibit the cell proliferation in lung cancer (Garcia-Tirado, *et al*., 2012). Prospective epidemiologic studies demonstrate that consumption of foods and beverages rich in flavonoids are inversely correlated with coronary stroke mortality or prevalence of neurodegenerative diseases including Alzheimer's and Parkinson's diseases (Stoclet and Schini-Kerth, 2012).

Structure and classification of flavonoids

Flavonoids share the same basic skeleton, the flavan-nucleus consisting of two aromatic rings with six carbon atoms, ring A and B, interconnected by hetero cycle including three carbon atoms ring C, which can be a heterocyclic pyran or a pyron (Peterson and Dwyer, 1998; Aherne and O'Brien, 2002 Schijlen *et al*., 2004; Kar, 2007) (Fig. 1). Flavonoids are divided into classes according to their oxidation level on the C-ring, which include anthocyanidins, flavanols (catechins), flavones, flavonols, flavanones and isoflavones. Anthocyanins and flavanols (catechins) have a heterocyclic pyran conformation whereas flavones, flavonols, flavanones and isoflavones have a 4-oxo-pyrone conformation, which carry a carbonyl group on C4 of ring C (Hollman

and Katan, 1999; Manach *et al*., 2004; Venketeshwer, 2012) (Fig. 2). Flavones and flavonols have been identified in almost all the plants, the most frequently found being those with B-ring hydroxylation in the C-3 and C-4 positions (Rice-Evan *et al*., 1997).

The structures of flavonoids differ greatly within the major classifications and substitutions include glycosylation, hydrogenation, hydroxylation, malonylation, methylation and sulfation (Harborne, 1986). The pattern of conjugation, glycosylation, or methylation can be very complex, can modify the hydrophilicity of the molecule and its biological properties and markedly increase the molecular weight of the flavonoid. Flavonoid molecules not attached to sugar moieties are referred to as the aglycone form, whereas flavonoid molecules with sugar moieties are called flavonoid glycosides (Aherne and O`Brien, 2002, Manach *et al*., 2004).

Role of flavonoids

Flavonoids are in charge of providing pigmentation to flowers fruits and seeds, plant defense, protection against UV light, pollen fertility, (Bradley *et al*., 1998), attracting pollinators and seed dispersers, defending plant against pathogenic microorganisms, plant fertility and pollen germination, acting as signal molecules in plant microbe interactions (Koes *et al*., 1989; Dixon and Paiva, 1995; van der Meer *et*

al.,1992; Makoi *et al*., 2010). Flavonoids also act as superoxide scavengers and antioxidants (Chen *et al*., 1990; Middleton and Teramura 1993) and possible plant energy escape valves (Hernandez and Breusegem, 2010).

In soybean *Glycine max* (L.) Merr., flavonoids are in charge of providing black color to the seed (Kovinich *et al*., 2010), plant resistance against aphids (Sato *et al*., 2013), regulate plant salt tolerance (Yan *et al*., 2013), enhance seedling resistance to UV-B radiation (Bin *et al*., 2006), induce pubescence pigmentation (Iwashina *et al*., 2006-a; Nagamatsu *et al*., 2009), interconnecting signals in the symbiosis of arbuscular mycorrhizal fungi (Antunes *et al*., 2006), chilling tolerance and seed coat cracking (Takahashi and Asanuma, 1996; Takahashi, *et al*., 2005).

Distribution of flavonoids

Flavonoids are widely distributed in plants. This distribution depends on several factors including variation according to plant phyla/order/family and population variations within species (Harborne, 1986). The distribution pattern depends on the degree of accessibility to light and previous illumination because formation of the higher oxidized flavonoids is accelerated by light (Aherne and O'Brien, 2002). Usually, flavonoids are found in roots, leaves, flowers fruits, stem, buds, bark, resins, gums and

oils (Wollenweber and Dietz, 1981). Flavonoids deposit in the leaf epidermis and most of flavonoids are located in the central vacuoles of epidermal cells, some in epidermal cell walls and associated with cell nuclei (Hutzler *et al.*, 1998), others in trichomes, which contains large amounts of flavonoids including anthocyanins and others in chloroplasts, which appear to be capable of flavonoid biosynthesis (Agati *et al*., 2012).

In leafy vegetables and fruits, flavonols are almost exclusively present as glycosides (Andersen and Markhal, 2007). Flavonol glycosides are located mainly in the leaves, flowers and outer parts of plants such as skin and peel; and decrease in concentration toward the central core (Crozier *et al.*, 1997). Only trace amounts of flavonols are found in plant parts below the soil surface, with the notable exception of onions (Aherne and O'Brien, 2002). The pattern of flavonoid glycosides can differ within plant leaves, where epidermal cells have been shown to contain different glycosides from mesophyll cells. Quercetin glycosides have been shown to accumulate in the skin of red grapes (Figueiredo-Gonzalez *et al*., 2012), Spanish cherry, tomatoes (Stewart *et al*., 2000) and apples (Bhagwat *et al*., 2011)

In soybean, flavonoids deposit in various tissues with a tissue-specific manner. Flower petals usually contain four kinds of anthocyanins, eight flavonol glycosides and one dihydroflavonol (Iwashina *et al.*, 2007; 2008), pubescence exclusively deposits

flavones (apigenin or luteolin), flavone aglycons outside of the cell surface (so-called "surface flavonoids") and flavone glycosides inside of the cells (Iwashina *et al*., 2006-b). Leaves contain various kinds of flavonol glycosides with substantial varietal differences (Buttery and Buzzell, 1973).

The flavonoid biosynthetic pathway

A wealth of genes can be distinguished within the flavonoid pathway, structural, regulatory and vacuolar pH genes. Structural genes control single biosynthetic steps of the various flavonoid classes, or steps of flavonoid modification; and regulatory genes that switch on or off the whole pathway or parts of it, that influence flavonoid concentration, or lead to pattern formation; and genes that are responsible for vacuolar pH, co-pigmentation and interaction with metal ions (Forkman and Martens, 2001; Schijlen *et al*., 2004).

The precursors of the synthesis of most flavonoids are malonyl-CoA and p-coumaroyl-CoA, which are derived from carbohydrate metabolism and phenylpropanoid pathway, respectively (Forkmann and Heller, 1999; Schijlen *et al*., 2004). The biosynthesis of flavonoids is initiated by the enzymatic step catalyzed by chalcone synthase (CHS) and ended by flavonoid 3-*O*-glucosyltransferases, these last in

charge to providing stability to the flavonoid molecule, attaching glucose at the 3-position of ring C. In the majority of plants, chalcones are not the end-products and the pathway proceeds with several enzymatic steps to other classes of flavonoids like, flavanones → dihydroflavonols → anthocyanins, mediate by the enzymes chalcone isomerase (CHI), flavonoid hydroxylases (F3H/F3'H/F3'5'H), dihydroflavonol 4-reductase (DFR), anthocyanidin synthase (ANS) and flavonoid 3-*O*-glucosyltransferases (Bradley *et al*., 1998; Winkel-Shirley, 2001; Schijlen *et al*., 2004) (Fig. 3).

Glycosylation of flavonoids

Except for catechins, flavonoids do not occur in plants as aglycones; the most frequently occurring forms are the glycoside derivatives in plants (Harborne, 1986; Hertog *et al*., 1993). Glycosylation increases the polarity of the flavonoid molecule, which is necessary for storage in plant cell vacuoles. Flavonols and flavones occur in food usually as *O*-β-glycosides (Harborne, 1986). Of the major flavonoid classes, the flavonols predominate in fruits in which a variety of glycosides have been identified, whereas in vegetables quercetin glycosides predominate (Aherne and O'Brien, 2002).

When glycosides are formed, the preferred glycosylation site on the flavonol

molecule is the C-3 position, and less frequently, the C-7 position (Fossen, *et al*., 1998). D-glucose is the most usual sugar residue, but other carbohydrate substitutions include arabinose, galactose, glucorhamnose, lignin, L-rhamnose and xylose. For instance, quercetin can be linked to the 3-*O*-glycoside rhamnose to yield quercitrin, or glucorhamnose to yield rutin (Aherne and O'Brien, 2002).

Flavonol glycoside biosynthetic pathway in soybean

The dihydroflavonols can be converted into flavonols by the enzyme flavonol synthase (FLS); this enzyme catalyzes the introduction of a double bond between carbon 2 and 3 of the C-ring, producing the flavonols kaempferol, quercetin and myricetin respectively (Schijlen *et al*., 2004) (Fig. 3). Further, flavonol modifications include glycosylation of the glycoside attached at the 3-position of ring C, suggesting the existence of flavonol glycoside glycosyltransferase genes, called FGG genes, which encode UDP:flavonol 3-*O*-glucosyltransferases enzymes, in charge of transferring glycosides to the glycoside attached to 3-*O*-flavonol position, allowing more complex flavonol structures like flavonol di-glycosides and tri-glycosides. Flavonol glycosides in plant cells have been implicated in growth regulation, disease resistance and protection of the leaf mesophyll from UV radiation (Buttery and Buzzel, 1992). However, little is

known about the genes, synthesis and control; and also the characteristics they affect *in planta*.

In soybean leaves, eighteen flavonol glycosides have been found and used to make the steps of a hypothetic biosynthetic pathway. Buttery and Buzzel, (1975) found nine kaempferol 3-*O*-glycosides K1, K2, K3, K4, K5, K6, K7, K8, K9 and nine analogous quercetin 3-*O*-glycosides Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9 in leaves of soybean, and proposed a hypothetical biosynthetic pathway composed by four main FGG gene alleles in charge of flavonol biosynthesis, *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) (Buttery and Buzzel, 1975) (Fig. 4). These FGG genes are responsible for binding glucose or rhamnose, to a glucose that was bound to the 3-position of K5 at either 2" or 6"-positions and interacted each other to obtain K1, K3, K6 and K7 (diglycosides); and K2, K4, K8 and K9 (triglycosides) (Buttery and Buzzel, 1975, 1976). Also, other alleles for the *Fg2* locus, *Fg2-a*, *Fg2-b* and *fg2* have been identified (Buzzell and Buttery, 1992), *Fg3* and *Fg4* are linked with a recombination frequency of 12.0 % in the molecular linkage group C2 (chromosome 6) (Buzzell, 1974).

FGG genes of soybean are known to be associated with the control of important

traits like, photosynthesis rate, leaf chlorophyll content, specific leaf weight, bean yield and wavy leaf. The complementary action of *Fg1* and *Fg3,* results in a kaempferol triglucoside Kaempferol-3-*O*-2^{G}-glucosylgentobiose or K9 (Buttery and Buzzel, 1975, 1976). Crossing experiences and field evaluations showed that Kaempferol K9 is associated with lower chlorophyll content, lower photosynthetic rate, lower specific leaf weight, extremely low adaxial stomatal density, moderately reduced abaxial stomatal density (Buttery and Buzzel, 1976), inhibition of stomatal development (Liu-Gitz, *et al*., 2000) and waviness of leaf margins (Buzzel and Buttery, 1998).

Aim of this study

HPLC analysis of soybean leaves revealed that kinds and contents of flavonol glycosides were substantially different among cultivars (Table 1). The first 21 cultivars (group A) contain quercetins and kaempferols and the remaining 11 cultivars contain only kaempferols (group B). A crossing between cultivars between group A and group B may result in segregation of a F3'H gene, in addition to FGG genes, and make segregation more complicated. So, we made crossings only within cultivars of group B.

The aim of this research is to identify, clone and verify the function of FGG genes in soybean leaves and determine the physiological traits they affect *in planta*. For this

purpose, we chose two sets of cultivars with differences in their flavonol glycoside composition, Kitakomachi x Koganejiro (KK); and Nezumisaya x Harosoy (NH); and were crossed and successively grown until F_6 generation. The parents and RILs populations were subjected to HPLC analysis to determine their flavonol glycoside composition; genetic analysis, to localize candidate genes in SSR linkage maps; cloning of candidate genes to verify their function. Further, we developed near-isogenic lines (NILs), to determine the physiological traits they affect *in planta*. Understanding of the genetic biosynthesis pathway of flavonol glycosides in soybean leaves may lead to improvement of soybean cultivars and make them more profitable to farmers.

Table 1. Flavonol glycoside contents (%) in leaves of soybean cultivars. Cultivar we used are shown in red font.

Cultivar	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14+15	K16	K17	A1
Clark	33.3	12.1	6.2	4	5.8	5.2	2	1.4	-	-	-	-	-	-	11	9.4	9.5	-	-	-	-	-									
T31	10.6	-	2.3	2.7	1.9	1.6	5.8	0.9	16.3	10.3	13	1.5	1.6	16.3	2.7	2.6	1.4	-	-	-	-	-	6.9								
Shakojou	8.7	-	4.4	8.3	2.3	2.8	3.2	1.8	22.4	15.8	-	-	-	-	4.3	3.5	3.2	-	-	-	-	-	19.3								
Nakaide	30.9	13.6	4.7	4.1	4.1	3.6	0.6	0.2	-	-	-	-	-	-	11.3	18.7	8.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Clark (td)	44	13	4.6	3.1	4.6	4.1	0.7	0.4	-	-	-	-	-	-	8	9.5	7.9	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Clark (W4)	31.7	10.1	8.9	6.1	11.4	8.3	0.6	0.3	-	-	-	-	-	-	7.7	8.1	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Clark (W1)	36.2	13.4	6.4	4.2	5.8	4.9	0.4	0.2	-	-	-	-	-	-	9	8.7	7.8	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Clark (W3w4)	27.1	12.9	6.8	4.2	8.8	5.9	4.1	3.4	-	-	-	-	-	-	7.9	9.8	9.1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Harosoy (t)	37.2	10.3	7.2	4.7	9.3	7.7	3.8	1.9	-	-	-	-	-	-	5.5	7	5.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0518 BW8	33.4	12.6	8.5	4.6	5.6	4.8	0.8	0.1	-	-	-	-	-	-	11.5	10.2	7.9	-	-	-	-	-	-	-	-	-	-	-	-	-	-
0734 BW1	42.5	11.7	4.5	3.5	4.7	3.7	2.1	0.6	-	-	-	-	-	-	8.8	10.8	7.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Aomori	27.4	18	7.1	6	3.2	2.2	5.3	1.5	-	-	-	-	-	-	16.7	8.7	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Saga	37	11.8	2.7	-	1.8	1.5	4.6	-	-	-	-	-	-	-	19.8	17.3	3.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Miyako white	61.2	-	-	1.1	5.3	7.5	2.3	0.5	0.7	1.7	-	-	-	-	0.4	7.5	11.8	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Himeshirazu R3	7.3	-	7.6	3.1	0.5	3.5		1.1	17.2	11.1	-	-	-	-	16.8	0.5	6	-	-	-	-	-	25.1	-	-	-	-	-	-	-	-
Himeshirazu R5	8.4	-	9.4	2.8	0.9	3.6	0.8	1	18.6	12.2	-	-	-	-	14.9	0.6	4.9	-	-	-	-	-	21.7	-	-	-	-	-	-	-	-
Sodendaizu	8.2	-	9.3	3.9	2.7	3.1	2	0.8	17.7	17.8	-	-	-	-	15.4	1.8	2.3	-	-	-	-	-	14.6	-	-	-	-	-	-	-	-
Kosamame1	11.6	-	5.7	3.9	1.5	2.1	3.6	1.9	22.9	14.2	-	-	-	-	8.8	1.1	3.6	-	-	-	-	-	19	-	-	-	-	-	-	-	-
To7B	7.4	-	8.9	5.9	3.2	2.8	2.4	1.7	17.8	16.3	-	-	-	-	6.9	3.4	3.4	-	-	-	-	-	19.9	-	-	-	-	-	-	-	-
Shakojou	8.7		4.4	8.3	2.3	2.8	3.2	1.8	22.4	15.8					4.3	3.5	3.2						19.3								
T31	10.6		2.3	2.7	1.9	1.6	5.8	0.9	16.3	10.3	13	1.5	1.6	16.3	2.7	2.6	1.4						6.9								
Harosoy	-	-	-	-	-	-	-	-	-	-	-	-	-	-	19.6	20.7	16.1	14.4	10.8	9.8		3.8	-	-	-	-	-	-	-	-	3.3
Clark (t)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	33.9	31	25.8	-	-	-	2.3	-	-	-	0.6	0.1	3.1	-	-	-	3.2
Koganejiro	-	-	-	-	-	-	-	-	-	-	-	-	-	-	32.7	33.2	22.9	-	-	-	3.4	-	-	-	0.8	0.3	2.7	-	-	-	4
Toiku 237	-	-	-	-	-	-	-	-	-	-	-	-	-	-	32	35.8	22	-	-	-	3.8	-	-	-	1	0.2	2	-	-	-	3.2
Bay	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10.2	5.1	4.6	-	-	-	2.8	-	-	-	-	-	54.6	17.8	1.4	1.6	2
Nezumisaya	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8.5	5.7	5.3	-	-	-	2.1	-	-	-	-	-	49.6	21.6	1.6	2.1	3.6
ToG7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	9.7	7	6.1	-	-	-	-	-	-	-	-	-	52.2	18.3	-	-	6.7
Fukuibuki	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	5.3	5.9	-	-	-	-	-	-	-	-	-	59.5	23.1	-	-	4.2
Kitakomachi	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6.1	-	-	-	1.3	3.3	32.7	-	-	9.7	23	20.7	-	-	-	-	3.3
Toyosuzu	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5.1	-	-	-	2.4	6	31.1	-	-	8.7	27.4	17.8	-	-	-	-	1.5
Harosoy (wm)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	45.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	54.8

*Q = quercitin, K= kaempferol, A= apigenin

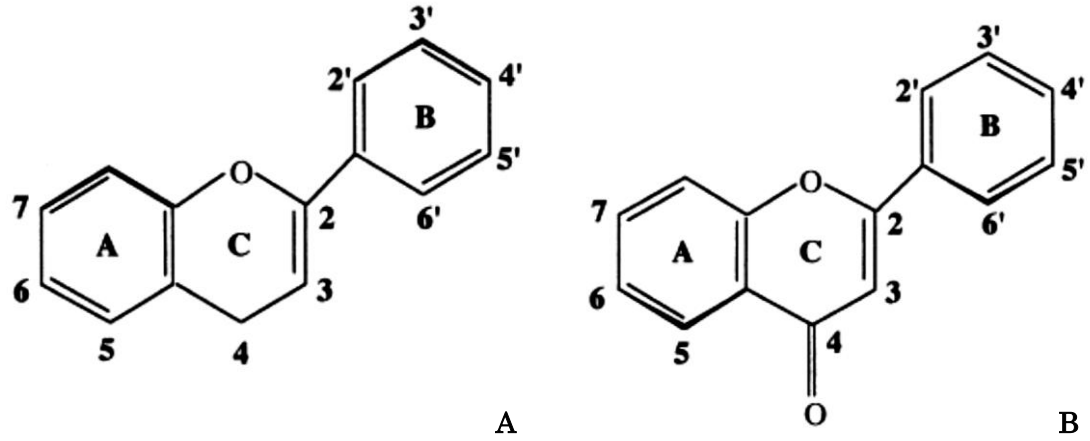


Fig. 1. Basic structures of flavonids (A) flavan nucleus (B) 4-oxo-flavonoid nucleus (Source: Arhene and O'Brien, 2002). The structure of the flavonoids is based on the flavonoid nucleus, which consists of three phenolic rings referred to as the A, B and C rings. The benzene ring A is condensed with a six-member ring C, which in the 2-position carries a phenyl benzene ring B as a substituent. Ring C may be a heterocyclic pyran (A) or a pyron, also called as 4-oxo-flavonoid nucleus (B).

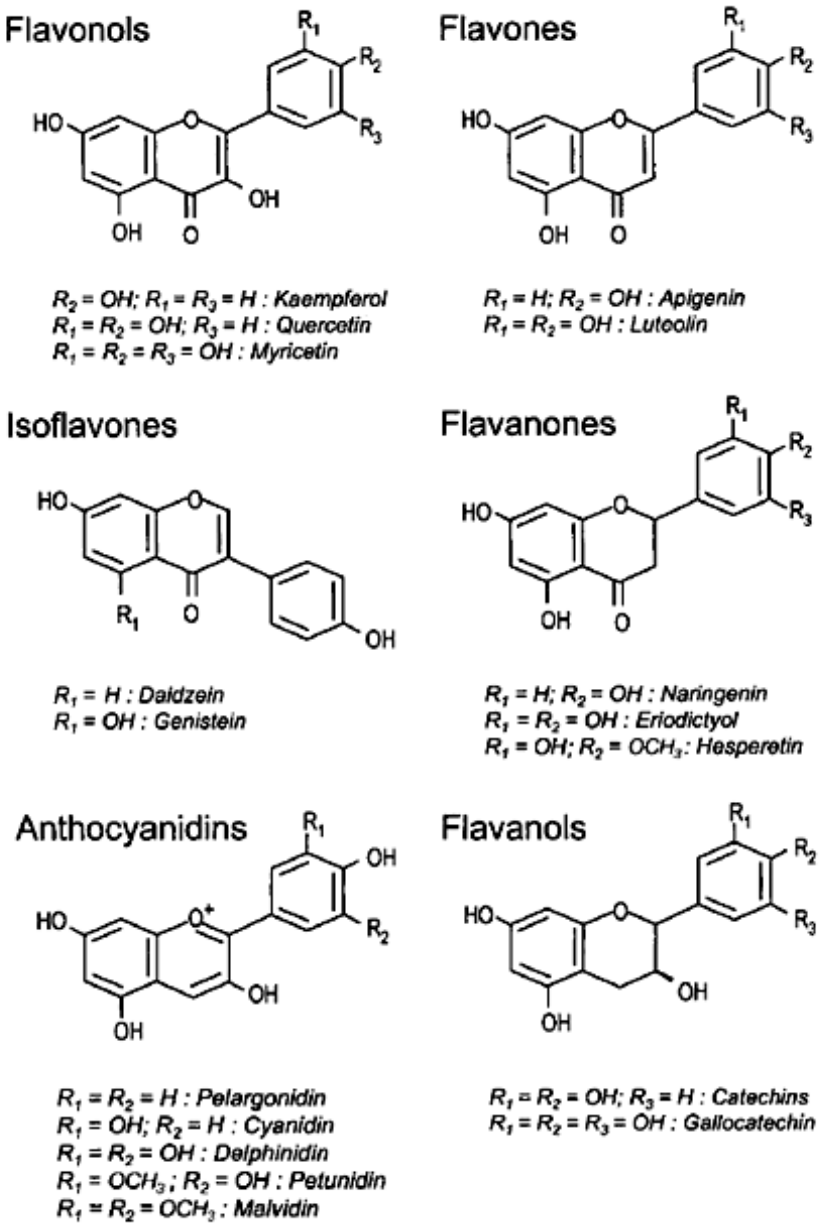


Fig. 2. Main flavonoid groups (Source: Manach *et al*., 2004). Flavonoids are divided into classes according to their oxidation level on the C-ring, which include anthocyanidins, flavanols (catechins), flavones, flavonols, flavanones and isoflavones on C-4 of ring C.

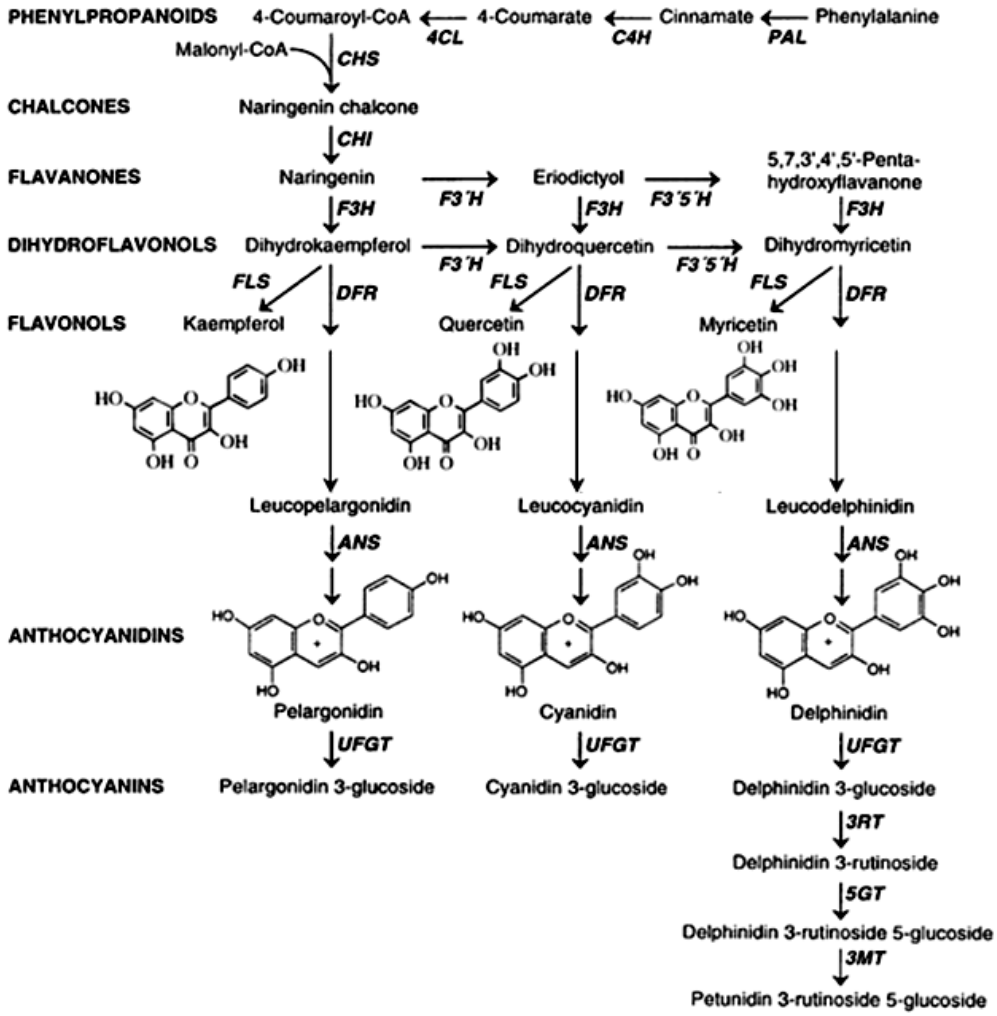


Fig. 3. Flavonoid biosynthetic pathway (Adapted from Bradley *et al*., 1998). Abbreviations are as follow: PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; DFR, dihydroflavonol 4-reductase; FLS, flavonol synthase; F3'H flavonoid 3'-hydroxylase; F3'5'H, flavonoid 3'5'-hydroxylase; ANS anthocyanidin synthase; UFGT, UDP-glucose:flavonoid 3-*O*-glucosyltransferase; 3RT, UDP-rhamnose:anthocyanidin-3-*O*-glucoside rhamnosyltransferase; 3MT, anthocyanin-3-*O*-methyltransferase.

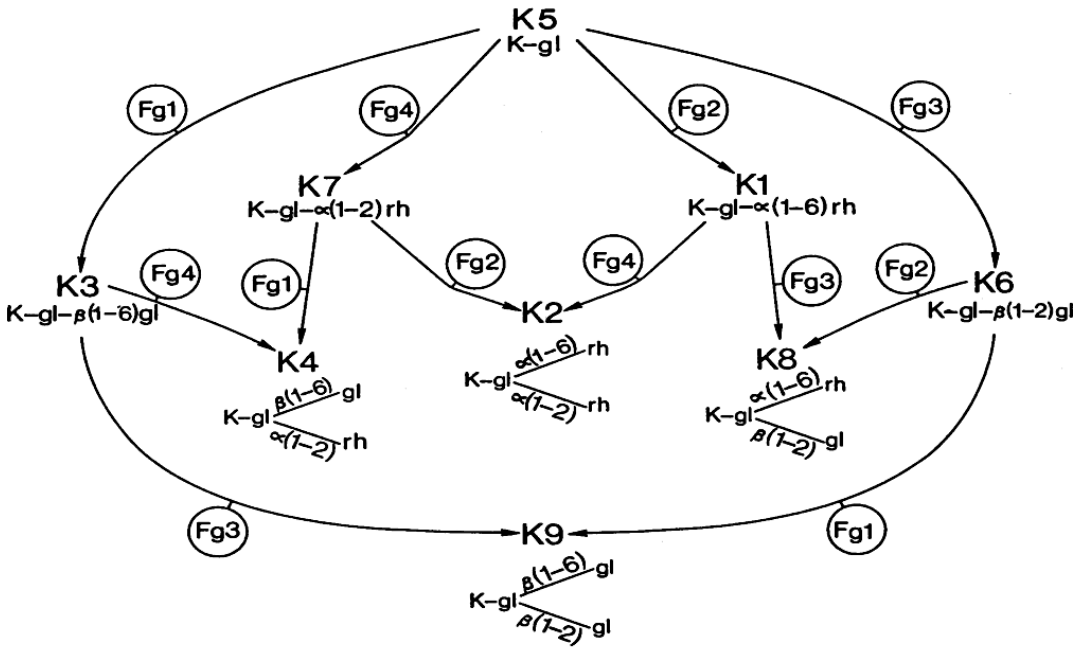


Fig. 4. Schematic presentation of flavonol glycoside genes (Source: Buttery and Buzzel, 1975). Scheme proposed for flavonol glycoside alleles in soybean, *Fg1* (β(1-6)-glucoside present), *Fg2* (α(1-6)-rhamnoside present), *Fg3* (β(1-2)-glucoside present) and *Fg4* (α(1-2)-rhamnoside present), defined by the ability to bind glucose or rhamnose at either 2″- or 6″-position of glucose that is bound to the 3-position of flavonols.

CHAPTER II

2. Analysis of flavonol glycoside genes segregated in Kitakomachi and Koganejiro (KK) population

2.1 Background

Preliminary experiments suggested that two Japanese cultivars, Kitakomachi and Koganejiro, had differences in the amount and composition of FGs extracted from leaves (Table 1). In Koganejiro, kaempferols K1, K2 and K3 are the main flavonols found and accounted for the 88.8 % of the total flavonols extracted from leaves; whereas in Kitakomachi K7, K11 and K12 were the main flavonols and accounted for the 76.4 %. These differences in flavonol composition suggested that one or a pool of genes can be in charge of FGs biosynthesis in both cultivars; however the number of genes, localization in the soybean genome, or function, is still unclear.

The objective of this study is to locate, identify, clone, sequence, verify the function of candidate genes in charge of FGs biosynthesis segregated in RILs developed from a cross between Kitakomachi and Koganejiro; and develop NILs to determinate the characters *in planta*, in order to understand how the flavonol glycoside pathway works and find ways to enhance the soybean crop production.

2.2 Materials and methods

Plant Materials

Kitakomachi and Koganejiro were crossed and selfing successively to develop

recombinant inbred lines (RILs) until F_6 generation. Koganejiro with yellow hilum, yellow seed coats, gray pubescence, purple flowers and narrow leaflets (*IIttW1W1W2W2w3w3W4W4WmWmWpWpLnLn*), was crossed with Kitakomachi with yellow hilum, yellow seed coats, gray pubescence, purple flowers and ovate leaflets (*IIttW1W1W2W2w3w3W4W4WmWmWpWplnln*) (Table 2). Flowers of Koganejiro were emasculated one day before opening and fertilized with pollen from Kitakomachi in year 2004. Hybridity of F_1 plants was ascertained based on ovate leaflets. A total of 120 RILs from the F_6 generation were developed without any selection using single-seed descent method. Seeds were planted at the National Institute of Crops Sciences, Tsukuba, Japan (36° 06'N, 140° 05'E) on 9th of June 2011 and fertilized with N, P and K at 3.0, 4.4 and 8.3 g/m^2 respectively. Plants were individually planted with spaces between rows and plants at 70 and 10 cm, respectively. Nine plants were grown for each parent and RILs.

Flavonoid extraction

A total of 94 RILs and parents were used. Trifoliated leaves were sampled from four plants of each RILs plus parents at stage R6 (Fehr *et al*., 1971) in three replications and subjected to MeOH extraction. The number of samples was determinate because of the 96 wells plates for PCR and gel comb for electrophoresis machine (94 samples and two parents). Leaf disks of 5 mm diameter were made from every sample using an office puncher in a total amount of 100 ± 1 mg, placed into 2 mL plastic tubes in three replications, soaked in 1 mL of MeOH, vortexed and kept at 4 ºC for 48 hours, performing vortex every 24 hours. Then, the samples were filtered through a disposable filtration unit Maishoridisk H-13-5 (Tosho) and subjected to HPLC analysis.

HPLC and genetic analysis

Quantitative HPLC analysis of the extract samples was performed with the Agilent 1100 HPLC System using an L-column 2 ODS [I.D. 60 x 150 mm (Chemicals Evaluation and Research Institute)] at a flow rate of 1.0 mL/min and detection wavelength range of 190-700 nm. The eluents were A: 0.2 % phosphoric acid and solvent B: (0.2:30:70 v/v/v) phosphoric acid/acetonitrile/water. The gradient was 0 min, B 20 %; 18 min, B 50 %; 25 min, B 70 %; 31 min, B 50 %; 34 min, B 20 %; total time 40 minutes. The amount of flavonol glycosides was estimated from the pertinent peak area in a HPLC chromatogram with a detection wavelength of 350 nm. HPLC results were compared with seven authentic specimens from soybean cultivars and other plant species (Iwashina *et al*., 2006a, 2010; Murai *et al*., 2013). In addition to HPLC comparisons, eight kaempferol glycosides isolated Kitakomachi and Koganejiro (parents) were identified by UV, LC-MS, acid hydrolysis and ^{1}H and ^{13}C NMR. The genetic model for FGs composition was hypothesized based on the HPLC chromatogram of the parents and the RILs; and the chemical structure of the pertinent peaks.

SSR analysis

Genomic DNAs from 96 plants selected for flavonoid extraction, were extracted from trifoliated leaves by CTAB method (Murray and Thompson, 1980). Polymorphisms between parents were screened using a total of 1025 SSR markers developed by USDA (Song *et al*., 2004) and by Kazusa DNA Research Institute (Hisano *et al*., 2007). PCR mixture contained 10 ng of genomic DNA, 2.25 pmol primers, 625 pmol nucleotides and 0.125 units of ExTaq in 1X ExTaq buffer (Takara Bio) in total

volume 5 μL. PCR program was performed as follows, 2 min denaturation at 94℃, followed by 35 cycles of 1 min denature at 94 ℃, 1 min annealing 49 ℃, 1 min extension at 72 ℃ and a final 5 min extension at 72 ℃. PCR products were separated in 8 % nondenaturing acrylamide gels. The fragments were visualized with ethidium bromide staining and UV light. PCR was performed in an Applied Biosystems 9700 thermal cycler.

Linkage mapping

A linkage map was constructed using AntMap software version 1.2 (Iwata and Ninomiya, 2006). The parameters 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 markers were tested by Chi-square test for segregation in 1:1 ratio.

QTL analysis

QTL analysis was performed to all the HPLC peaks F1, F2, F3, F4, F5, F6, F7, F8 and F9, in a F_6 population of 94 RILs derived from a cross between Kitakomachi and Koganejiro. The flavonol concentration of each peak was estimated by averaging the peak area of three replications. QTL analysis was performed by composite interval mapping (Zeng, 1993) using the QTL Cartographer version 2.5 (Basten *et al*., 2001). The threshold LOD score was determined by permutation test with 1000 repetitions corresponding to a genome-wide 5% level of significance. Forward and backward regression was applied.

RNA extraction and cDNA cloning

RNA was extracted from trifoliated leaves (200 mg) from Kitakomachi and Koganejiro using the TRIZOL Reagent (Invitrogen) according to the manufacturer's instructions. Reverse transcription of total RNA was obtained 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 from Kitakomachi and Koganejiro by end-to-end PCR using a pair of PCR primers (Table 3) 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 nmol of nucleotides and 1 unit of ExTaq in 1 X ExTaq Buffer in a total volume of 25 μl. PCR program was performed as follows, initial denature 30 sec at 94 ℃ followed by 30 cycles of 30 sec denature 94 ℃, 1 min annealing 59 ℃, 1 min extension 72 ℃ and 7 min final extension at 72 ℃. Finally the ~1.4 kbp PCR product was cloned into pCR vector 2.1 and sequenced.

Sequencing analysis

Nucleotides of both strands were determined with the BigDye terminator cycle method using an ABI3100 Genetic Analyzer (Applied Biosystems). Primer sequences are described in Table 3. Nucleotide and putative amino acid sequences were analyzed with the BLAST program (Altschul *et al.*, 1997). Intron/exon gene structure was estimated based on the comparison between the cDNA sequence and the corresponding genome sequences of Williams 82 deposited in the soybean genome database. The nucleotide sequences were aligned using ClustalW and the alignment was used to construct a phylogenetic tree using the neighbor-joining method

(http://clustalw.ddbj.nig.ac.jp/top-j.html).

dCAPS analysis

To detect the two-base deletion unique to Kitakomachi, the first primer contained a nucleotide (G) mismatched with its target DNA to artificially create a restriction site for *BcgI* (CGANNNNNNTGC) in Koganejiro (Fig. 10A). The deletion within the restriction site of Kitakomachi would abolish the restriction site in the amplified product to generate a polymorphism. dCAPS PCR mixture contained 30 ng of *g*enomic DNA, 10 pmol of each primer, 5 nmol of nucleotides and 1 unit of ExTaq in 1 x ExTaq Buffer in a total volume of 25 μl. PCR program consisted of initial 30 sec denaturation at 94 ℃, followed by 30 cycles of 30 sec denaturation at 94 ℃, 1 min annealing at 59 ℃ and 1 min extension at 72 ℃ and a final 7 min extension at 72 ℃. The amplified products were digested with *Bcg*I enzyme; and the digests were separated on 8 % nondenaturing polyacrylamide gels. After electrophoresis, the gel was stained with ethidium bromide and the DNA fragments were visualized under UV light.

Expression and purification of recombinant *GmF3G6"Rt* proteins (RIKEN)

The entire coding region of *GmF3G6"Rt* was amplified from Kitakomachi and Koganejiro by PCR, using the high fidelity KOD-Plus-DNA polymerase (Toyobo) and primers designed for the restriction enzyme sites *Sac*I and *Xho*I (Table 3). Since *GmF3G6"Rt* had no intron, genomic DNA was used as a template, the PCR mixture was performed with 30 ng of *g*enomic DNA, 10 pmol of each primer, 5 nmol of nucleotides, 50 nmol 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. KOD PCR program consisted on initial 2 min

denaturation at 94 ℃, followed by 40 cycles of 15 sec denaturation at 94 ℃, 30 sec annealing at 59 ℃ and 1 min extension at 68 ℃ and a final 7 min extension at 68 ℃. The PCR amplicon was digested with *Sac*I and *Xho*I restriction enzymes and then cloned into the pCold ProS2 vector (Takara Bio) previously digested with the same restriction enzymes.

Escherichia coli strain BL21star (DE3) was used as a host for expression of PCR-pCold ProS2 vector and the proteins were purified. Transformed cells were cultured at 37 ºC until the optical density A_{600} reached 0.5, then the cells were cultured at 15 ºC for 24 h after the addition of isopropyl-ß-D-thiogalactopyranoside to a final concentration of 1mM. Cells were collected and the protein was purified as a His fusion using TALON® Metal Affinity Resin (Clontech) according to the manufacturer's instructions. ProS2 tag was removed using HRV3C protease (Novagen) according to the manufacturer's instructions. After exchanging the buffer to 50 mM HEPES-KOH, pH 7.5, the proteins were concentrated using an Amicon Ultra filter (10,000 MWCO, Millipore) and were used for enzyme assays.

Enzyme assays (RIKEN)

The standard enzyme assay reaction mixture was described as before (Yonekura-Sakakibara *et al*, 2007). The mixture was pre-incubated at 30 ºC for 2 min and the reaction was started by the addition of the enzyme. Reactions were stopped after 0, 2, 4, 6, 60 or 80 min of incubation at 30 ºC by the addition of 50 µl ice-cold 0.5% (v/v) trifluoroacetic acid/methanol. The supernatant was recovered by centrifugation at 12,000 g for 3 minutes; flavonoids in the resultant solution were analyzed using a Shimadzu HPLC system with a Unison UK-C18 column (2.0 x 150 mm, 3µm, Imtakt

Corporation) at a flow rate of 0.2 ml/min at 35 °C. Compounds were separated with a linear eluting gradient with solvent A (0.1% trifluoroacetic acid in water) and solvent B (0.1% trifluoroacetic acid in acetonitrile) set according to the following gradient: 0 min, 10 % B; 3 min, 10 % B; 18 min, 75 % B; 18.01 min, 95 % B; 20 min, 95 % B; 30 min, 95 % B. PDA was used for the detection of UV-visible absorption in the range of 200 to 600 nm.

Gene expression assays

Real-time quantitative reverse transcription PCR was performed to estimate gene expression of *GmF3G6″Rt* in Kitakomachi and Koganejiro. Total RNA (5μL) extracted from three replicated leave samples of Kitakomachi and Koganejiro at R5 stage, was reverse-transcribed using the Superscript III First Strand Synthesis System and oligo d(T)primer. PCR mixture contained 2.0 μL of cDNA synthesis reaction mixture, 6 pmol of each designed primer (Table 3), 1X ROX reference dye, 1X SYBR Premix Dimer Eraser (Takara Bio) and water to a final volume of 20 μL. Analysis was carried out in a StepOnePlus Real-Time PCR System, with a program of initial 30 sec denaturation at 95 ºC, followed by 40 cycles of 3 sec denaturation at 95 ºC, 3 sec annealing at 58 ºC and 30 sec extension at 72 ºC. The expression level of soybean *actin gene* (GenBank accession number: J01298) was used to normalize target gene expression; and the resulting expression levels were compared by t-test using Minitab 16 statistical software. Tissue-specific gene expression data from a diverse set of fourteen tissues were extracted from the RNA-seq Atlas at SoyBase webpage (http://soybase.org/soyseq/) (Severin *et al*., 2010).

Photosynthesis rate and chlorophyll contents of NILs

To develop NILs, dCAPs analysis was performed for 360 seed samples from three heterozygous lines, F7-19, F7-21 and F7-75. A total of 90 plants with genotype of the father (Kitakomachi) or mother (Koganejiro) were planted at the National Institute of Crops Sciences, Tsukuba, on 7th of June 2013. Photosynthesis rate and chlorophyll contents measurements were carried out at 10 week after planting, between 10:00 am and 2:00 pm with a LICOR LI-6400 biosciences and a MINOLTA SPAD-502, respectively. Data were collected *in situ* to a total of 90 plants, divided into Kitakomachi and Koganejiro genotype; and subjected to t-test to compare means using the Minitab Software.

2.3 Results

HPLC Analysis

HPLC chromatograms for Kitakomachi and Koganejiro showed different peaks, attributed to FGs (Fig. 5). Koganejiro chromatogram showed five primary peaks, 24.5 (F1), 25.1 (F2), 27.7 (F3), 28.8 (F4) and 30 min (F5); and Kitakomachi four primary peaks 26.4 min (F6), 26.9 (F7), 28.9 (F8) and 30 min (F5). All of them were identified as flavonol glycosides. Based on comparison with authentic specimens, these peaks correspond to the followings FGs:

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

F2, kaempferol 3-*O*-rhamnosyl-(1→2)-[rhamnosyl-(1→6)-glucoside]

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

F4, kaempferol 3-*O*-rhamnosyl-(1→6)-glucoside

F5, kaempferol 3-*O*-glucoside

F6, kaempferol 3-*O*-rhamnosyl-(1→2)-galactoside

F7, kaempferol 3-*O*-rhamnosyl-(1→2)-glucoside

F8, kaempferol 3-*O*-galactoside.

The results indicate that FGs specific of Koganejiro (F1 to F4) had rhamnose at the 6″-position of glucose or galactose that is bound to the 3-position of kaempferol, whereas FGs of Kitakomachi was devoid of rhamnose at the position.

Genetic analysis

HPLC chromatograms for 94 RILs, segregated into 53 RILs that had peaks pattern like Koganejiro (Koganejiro-type) and 41 RILs like Kitakomachi (Kitakomachi-type) (Table 4). The segregation fitted a 1:1 ratio (χ^2=1.53, P= 0.22) suggesting that a single gene, probably encoding flavonol 3-*O*-glucoside (1→6) rhamnosyltransferase, might control the composition of FGs.

SSR analysis and Linkage mapping

Among the 1025 SSR markers tested, 146 exhibited polymorphism between the parents and distinctly segregated in the RILs. These markers were used for linkage mapping. A total of 99 markers fell into 20 linkage groups spanning 748 cM (Fig. 6). The low map coverage of this population may be partly due to the shared pedigree of the parents as previously discussed (Githiri *et al*., 2007).

QTL analysis

QTLs responsible for each peak component (F1 to F8) were mapped in linkage

group O (chromosome 10), between the markers GMES3560 and Satt331, at the identical position of GMES0295 (Fig. 7). QTLs for F1 to F8 had a LOD score of 32.12 to 41.25; accounting for 80.60% to 92.23% of the phenotypic variation (Table 5).

cDNA cloning

A survey of the genome sequence of a US cultivar Williams 82 suggested that two candidate genes similar to *GT* genes, Glyma10g33790 and Glyma10g33800, aligned in tandem at a distance of about 6.2 kb in the vicinity of GMES0295. The entire coding region of Glyma10g33790 was amplified by PCR and subsequently cloned. Sequence analysis revealed that coding region of Glyma10g33790 was 1392 bp long encoding 464 amino acids in Koganejiro. We designated the gene as *GmF3G6"Rt*.

GmF3G6"Rt belongs to the family 1 glycosyltransferase and was designated as UGT79A6 by the UGT Nomenclature Committee (Mackenzie *et al*., 1997). *GmF3G6"Rt* had 56 % amino acid similarity with the *Rt* gene of petunia encoding anthocyanin 3-*O*-glucoside (1→6) rhamnosyltransferase (Kroon *et al*., 1994) and 58 % similarity with flavanone 7-*O*-glucoside (1→6) rhamnosyltransferase of *Citrus sinensis* (Frydman *et al*., 2013) (Fig. 8).

Compared with Koganejiro, the cDNA sequence of Kitakomachi had a two-base deletion at the nucleotide position 930 and two single-base substitutions in the downstream of the deletion (Fig. 9). The cDNAs of Koganejiro and Kitakomachi were designated as *GmF3G6"Rt-a* and *GmF3G6"Rt-b*, respectively. The deletion generated a truncated polypeptide that consisted of only 314 amino acids in Kitakomachi, *GmF3G6"Rt-b* lacked the PSPG (Plant Secondary Product Glycosyltransferase)-box around the C-terminal region (Masada *et al*. 2007) (Fig. 8). Comparison with the

genome sequence of Williams 82 suggested that *GmF3G6"Rt* had not intron.

The database survey suggested that Glyma10g33800 lacked an upstream region and may be was a pseudogene. Sequencing analysis of the coding region indicated that the transcript of Glyma10g33800 had premature stop codons in all reading frames. In addition, Koganejiro and Kitakomachi had identical sequence for this gene, suggesting that Glyma10g33800 may not be the responsible gene for *GmF3G6"Rt* that controls FGs in Kitakomachi and Koganejiro.

dCAPS analysis

Bands with molecular size of about 190 bp were amplified with the dCAPS primers in Koganejiro and Kitakomachi; and then digested with the restriction enzyme. *Bcg*I digestion generated a band of about 140 bp in Koganejiro, whereas Kitakomachi was unaffected (Fig. 10B), banding pattern of the RILs was classified into Koganejiro-type, Kitakomachi-type and heterozygous type (those having both bands). In total, 50 RILs had Koganejiro, 41 RILs had Kitakomachi and 3 RILs heterozygous type bands.

Banding patterns co-segregated with FGs patterns RILs, FGs of Kitakomachi type had bands classified as Kitakomachi-type, whereas RILs with FGs of Koganejiro-type had bands classified as Koganejiro-type or both types of bands. The 3 RILs with both bands had the FGs pattern classified as Koganejiro-type.

In vitro characterization of recombinant *GmF3G6"Rt-a*

The recombinant *GmF3G6"Rt* protein was expressed in *E. coli* as a ProS2 protein fusion and was semi-purified. The *GmF3G6"Rt-a* protein catalyzed the conversion of

kaempferol 3-*O*-glucoside to a single product kaempferol 3-*O*-rutinoside as confirmed by HPLC retention time (Fig. 11). Neither *GmF3G6"Rt-b* protein nor ProS2 protein alone as a negative control, catalyzed the conversion to 6″-*O*-rhamnoside. Thus, *GmF3G6"Rt-a* can be defined as a flavonol 3-*O*-glucoside (1→6) rhamnosyltransferase gene.

The specificity of *GmF3G6"Rt-a* as a sugar acceptor was also examined. *GmF3G6"Rt-a*, was similarly active on both 3-*O*-glucoside/galactoside of kaempferol and quercetin (Table 6). *GmF3G6"Rt-a* prefers kaempferol than quercetin as an aglycone; and 3-*O*-galactosides than 3-*O*-glucosides as a glycosylation pattern. The sugar donor specificity of *GmF3G6"Rt-a* was examined with UDP-rhamnose, UDP-glucose, UDP-galactose, UDP-glucuronic acid and UDP-arabinose as donors and kaempferol 3-*O*-glucoside as an acceptor. No UGT activity was detected for UDP-glucose, UDP-galactose and UDP-glucuronic acid, indicating that *GmF3G6"Rt-a* is highly specific to UDP-rhamnose. Only slight activity with UDP-arabinose was found, but the structure of the resultant product was not identified.

Gene expression

At the R5 stage, the transcript level of *GmF3G6"Rt* in leaves of Kitakomachi was 41.1 % of that in Koganejiro (t = 3.43*) (Fig. 12). According to the RNA-Seq Atlas of soybean, *GmF3G6"Rt* was expressed predominantly in young leaves and less so in flowers and young pod shells (Table 7). Almost no expression was observed in seeds, roots and root nodules.

Photosynthesis rate and chlorophyll contents of NILs

Photosynthesis rate for Kitakomachi, Koganejiro and the three pairs of NILs (F7-19, F7-21 and F7-75) are shown in Table 8-A. For F7-19, the mean of Koganejiro type (21.6) and Kitakomachi type (22.6) were not statistically different (t = 1.0, p>0.05). Similarly, for F7-21, the mean of Koganejiro type (15.9) and Kitakomachi type (17.6) had no statistical differences. For F7-75, the mean of Koganejiro type 18.7 and Kitakomachi type (18.0) with no statistical differences.

Interestingly, chlorophyll concentration showed different results as shown in Table 8B. For F7-19, the mean of Koganejiro type was 47.2 and Kitakomachi type 43.7. There was a significant difference between means, indicating that Koganejiro type had a higher content of chlorophyll than Kitakomachi. For F7-21, the mean of Koganejiro (45.0) and Kitakomachi (45.2) had not significant statistical difference. For F7-75, the mean of Koganejiro (42.0) and Kitakomachi (42.2) had no significant statistical difference. Thus, the first line, F7-19, showed statistical difference between Koganejiro and Kitakomachi for total 30 plants tested. In contrast, the other two lines did not showed statistical difference for the same number of plants tested.

2.4 Discussion

Soybean cultivars Koganejiro and Kitakomachi have gray pubescence and according to the HPLC analysis, their leaves contain predominantly kaempferol derivatives with substantial varietal differences. Koganejiro showed five primary HPLC peaks, F1 to F5, whereas Kitakomachi four peaks from F5 to F8. Both cultivars had the F5 peak, which amount was slight lower in Koganejiro, than in Kitakomachi where was abundant. Comparison with authentic specimens suggested that FGs of Koganejiro had

rhamnose at the 6″-position of *g*lucose or galactose that is bound to the 3-position of kaempferol, whereas FGs of Kitakomachi were devoid of rhamnose.

Genetic analysis suggested that a single gene controls FGs composition. SSR analysis, QTL analysis, linkage mapping and genome database survey revealed the existence of a gene responsible for FGs composition at the same position of GMES0295 in the molecular linkage group O (chromosome 10). The genome sequence of Williams 82 suggested two candidate genes Glyma10g33790 and Glyma10g33800, which were aligned in tandem in the vicinity of GMES0295. Both genes were cloned and sequenced, results of sequencing contrasted with genome sequence Williams 82 suggested that Glyma10g33800 might be a pseudogene, due to the presence of premature stop codons. On the contrary, Glyma10g33790 of Koganejiro was functional and was designated as *GmF3G6″Rt*. The coding region of *GmF3G6″Rt* was 1392 bp long, encoding 464 amino acids in Koganejiro (*GmF3G6″Rt-a*). It has an amino acid similarity of 56 % with the *Rt* gene of petunia encoding anthocyanin 3-*O*-glucoside (1→6) rhamnosyltransferase and 58 % similarity with flavanone 7-*O*-glucoside (1→6) rhamnosyltransferase of citrus (Frydman *et al*. 2013).

In contrast to Koganejiro, *GmF3G6″Rt* of Kitakomachi (*GmF3G6″Rt-b*) had a two-base deletion that generated a truncated polypeptide consisting of 314 amino acids. Detection of the two-base deletion by dCAPS analysis in Kitakomachi confirmed the nucleotide polymorphism. Banding patterns of the dCAPS marker co-segregated with FGs patterns. RNA-Seq Atlas of soybean suggested that *GmF3G6″Rt* was primarily expressed in leaves. The transcript level in leaves of Kitakomachi was lower than Koganejiro, probably because of nonsense-mediated mRNA decay, a surveillance mechanism to eliminate aberrant mRNA transcripts that contain premature stop codons

(Chang *et al*. 2007). Recombinant *GmF3G6"Rt-a* protein had a flavonol 3-*O*-glucoside (1→6) rhamnosyltransferase activity, indicating that *GmF3G6"Rt-a* gene encodes a functional flavonol 3-*O*-glucoside (1→6) rhamnosyltransferase; it presumably corresponds to the *Fg2* gene described by Buttery and Buzzel (1976). *GmF3G6"Rt-a* protein can utilize kaempferol 3-*O*-glucoside/3-*O*-galactoside as substrates. The *GmF3G6"Rt* had no intron, consistent with FGs genes of other plant species, most of which have either none or one intron (Paquette *et al*. 2003). Moreover, *GmF3G6"Rt-b* protein showed no activity probably because of the lack of a C-terminal region including the PSPG-box. These data are consistent with the flavonoid distribution pattern in Kitakomachi and Koganejiro.

Buzzell and Buttery (1992) reported a new allele for the *Fg2* locus, resulting in a series of alleles, *Fg2-a*, *Fg2-b* and *fg2*. Leaves of Kitakomachi and Koganejiro contained FGs with α(1-2)-rhamnoside but lacked FGs with β(1-6)-glucoside or β(1-2)-glucoside suggesting that they might have the allelic combination of *fg1 fg3 Fg4*. With this genetic background, *Fg2-a* generates three TLC spots corresponding to rutinoside, 2^{G}-rhamnosylrutinoside and glucoside, whereas *Fg2-b* generates six spots corresponding to gentiobioside, sophoroside and neohesperidoside in addition to the three spots produced by the *Fg2-a* allele (Buzzell and Buttery 1992). Leaves of Koganejiro contained rutinoside, 2^{G}-rhamnosylrutinoside and glucoside; and lacked gentiobioside, sophoroside and neohesperidoside, suggesting that the allele of Koganejiro is *Fg2-a*.

Currently, there is no evidence that *GmF3G6"Rt* affects visible phenotypes or agronomic characters. NILs developed for KK did not show differences for the photosynthetic rate and chlorophyll concentration. However, more detailed studies

should be carried out in order to obtain information on the effect of *GmF3G6"Rt* in relation to morphology, productivity and plant fitness in soybean.

Table 2. Plant materials used.

Cultivars (parents)	**Origin**	**Hilum Color**	**Pubescence color**	**Flower color**	**Leaflet shape**	**Genetic Characteristics**	**Cross abr.**
Kitakomachi	Japan	Yellow	Gray	Purple	Narrow	*IIttW1W1W2W2w3w3W4 W4WmWmWpWplnln*	KK
Koganejiro	Japan	Yellow	Gray	Purple	Ovate	*IIttW1W1W2W2w3w3W4 W4WmWmWpWpLnLn*	
Nezumisaya	Japan	Yellow	Gray	Purple-blue	Ovate	*IIttW1W1w2w2w3w3W4W 4WmWmWpWpLnLn*	NH
Harosoy	Canada	Yellow	Gray	Purple	Ovate	*IIttW1W1W2W2w3w3W4 W4WmWmWpWpLnLn*	

Table 3. Primers used for cloning, sequencing, dCAPS analysis, real-time PCR for *GmF3G6"Rt* gene.

Purpose	Target gene	Forward primer (5'-3')	Reverse primer (5'-3')
Cloning for sequencing	*GmF3G6"Rt*	GTATTGAGCAAAGTAGAGAC	TTTCCATTGGAAGAGACAAG
Sequencing	*GmF3G6"Rt*	1. GCATCCAACATTCCCAGAAT 2. ATCTTAAAAAGCCTCCACCT 3. AGTGAGACATTTCTGAGTGA	1. ATACATAGCACCCCACACTT 2. GCGTTTCTATGTAGTCTAAG 3. GTGTATGAGATTTCCAGCTA
dCAPS analysis	*GmF3G6"Rt*	GTTCTGAATTTCCCATCCGA	TCAATCACTGAACTAAAGCCAC
Cloning for functional analysis[a]	*GmF3G6"Rt*	GCCG<u>CATATG</u>CCTAGTGAATTAGCTATGAACAAT *Sac*I	GCCG<u>CTCGAG</u>CTAAGCCATAGACTTTAACTGGGCAA *Xho*I
Real-time PCR	*GmF3G6"Rt*	GAGAGAGCATTGCCAAAAGG	TAAAGCCACCATGGCATACA
	actin	GTCCTTTCAGGAGGTACAACC	CCACATCTGCTGGAAGGTGC

[a] *Sac*I and *Xho*I restriction sites are shown in forward and reverse primer respectively

Table 4. Genetic analysis of KK chromatograms.

A total of 94 RILs segregated into 53 RILs for Koganejiro-type pattern peaks; and 41 RILs for Kitakomachi-type. The segregation fitted a 1:1 ratio (χ^2=1.53, P= 0.22) suggesting that a single gene, probably encoding flavonol 3-*O*-glucoside (1→6) rhamnosyltransferase, might control the composition of FGs.

Koganejiro-type Peaks: 24.5, 27.7, 28.8 min			Kitakomachi-type Peaks: 26.4, 26.9, 28.9 min		
1	50	100	3	62	118
5	51	101	4	63	
8	54	102	7	64	
9	57	103	11	70	
10	59	104	12	73	
13	61	105	16	74	
19	68	106	18	76	
20	75	107	23	80	
21	77	108	29	82	
22	78	109	31	84	
24	79	112	33	87	
28	81	116	34	89	
30	83	117	35	92	
32	85		37	94	
36	88		41	96	
38	90		47	97	
39	93		48	110	
43	95		53	111	
45	98		55	114	
46	99		56	115	

Table 5. QTLs responsible for areas of HPLC peak F1, F2, F3, F4, F5, F6, F7 and F8, observed in a F_6 population developed from a cross between Kitakomachi and Koganejiro.

QTL name	Linkage group	Proximal marker	Marker position (cM)[1]	QTL position (cM)[1]	LOD score	Additive effect[2]	Variance (%)
F1	O	GMES0295	113.4	114.4	38.49	445.19	86.51
F2	O	GMES0295	113.4	114.4	42.77	59.77	82.89
F3	O	GMES0295	113.4	114.4	32.12	296.60	81.91
F4	O	GMES0295	113.4	114.4	39.76	554.82	80.60
F5	O	GMES0295	113.4	113.4	40.83	-292.35	83.42
F6	O	GMES0295	113.4	113.4	47.25	-341.34	85.92
F7	O	GMES0295	113.4	113.4	43.65	-253.94	92.23
F8	O	GMES0295	113.4	114.4	39.04	-99.75	81.94

[1] Distance from the top of each linkage group
[2] Additive effects of each QTL are those of Kitakomachi allele in contrast to Koganejiro allele

Table 6. Substrate specificity of *GmF3G6"Rt*.

Sugar acceptor[a]	Relative activity (%)
Kaempferol 3-*O* -glucoside[c]	100.0 ± 0.7
Kaempferol 3-*O* -galactoside[c]	152.4 ± 7.2
Quercetin 3-*O* -glucoside	75.3 ± 1.5
Quercetin 3-*O* -galactoside	129.0 ± 6.9
Sugar donor[b]	
UDP -rhamnose	100.0 ± 3.4
UDP -glucose	ND
UDP -galactose	ND
UDP -glucuronic acid	ND
UDP -arabinose	2.1 ± 0.1

(Experiment was carried out in RIKEN)

ND, not detected

[a] Reactions were performed with UDP -rhamnose as sugar donor

[b] Reaction were performed with kaempferol-3-*O*-glucoside as sugar acceptor

[c] Enzymatic products are identified based on comparisons with the standards.

Table 7. Tissue-specific gene expression of Glyma10g33790 based on the RNA-Seq Atlas of soybean. Raw data was normalized by following equation. $eq = reads/kilobase/millon$

Gene	Young leaf	Flower	One cm pod	Pod shell 10DAF[a]	Pod shell 14DAF	Seed 10DAF	Seed 14DAF	Seed 21DAF	Seed 25DAF	Seed 28DAF	Seed 35DAF	Seed 42DAF	Root	Nodule
Glyma10g33790	75	22	20	26	18	1	0	0	0	0	0	0	0	0

[a]DAF, Days after flowering.

Table 8. Photosynthesis rate (μmol CO_2 m^{-2} s^{-1}) (A); and chlorophyll concentration (nmol/cm^2) of NILs (B).

A

Sample No	Ko F7-19	Ki F7-19	Ko F7-21	Ki F7-21	Ko F7-75	Ki F7-75
1	25.4	27.8	9.8	16.2	19.6	13.0
2	20.4	10.9	14.3	23.6	16.5	20.1
3	24.7	23.2	15.0	13.1	18.9	13.6
4	20.9	15.0	15.6	20.4	12.2	12.0
5	25.9	21.5	10.7	13.0	19.8	21.0
6	24.1	25.9	9.1	19.9	22.8	24.5
7	19.8	27.5	19.0	16.9	12.5	24.9
8	24.8	24.0	27.1	16.9	15.3	13.3
9	21.0	24.3	17.4	18.6	19.7	14.3
10	21.8	23.3	12.5	14.7	22.4	14.4
11	24.9	24.2	17.0	10.7	18.6	23.9
12	17.1	17.1	24.3	21.2	23.4	15.8
13	13.2	22.7	13.2	21.0	18.5	21.1
14	17.2	24.2	19.9	18.6	22.1	20.3
15	23.0	19.7	13.1	19.9	18.6	17.8
Mean	21.6	22.6	15.9	17.6	18.7	18.0
Standard deviation	3.7	3.6	5.1	3.6	3.4	4.5
t – value		1.0		1.8		0.7
Probability		0.46		0.28		0.62

B

Sample No	Ko F7-19	Ki F7-19	Ko F7-21	Ki F7-21	Ko F7-75	Ki F7-75
1	43.3	45.6	45.0	46.2	40.0	42.9
2	45.1	42.2	47.0	41.0	42.8	42.4
3	48.7	43.8	45.1	43.6	40.5	42.2
4	52.1	42.1	45.2	45.3	39.3	45.2
5	44.6	45.5	44.2	42.9	41.3	43.8
6	44.9	44.5	45.7	40.4	42.8	44.8
7	44.3	46.4	44.9	45.7	42.2	41.2
8	50.9	42.7	44.5	47.1	39.0	41.6
9	46.6	42.7	41.8	45.2	41.9	42.8
10	48.7	42.8	46.2	49.8	44.0	42.0
11	50.4	43.4	45.3	44.1	43.3	39.6
12	46.3	41.9	44.8	46.1	45.8	39.5
13	47.7	43.6	45.1	46.5	39.5	41.9
14	43.9	44.3	44.1	46.5	43.1	40.7
15	49.8	43.9	45.6	47.3	45.0	42.4
Mean	47.2	43.7	45.0	45.2	42.0	42.2
Standar deviation	2.8	1.4	1.1	2.5	2.1	1.6
t-value		3.5*		0.2		0.2
Probability		0.00		0.76		0.81

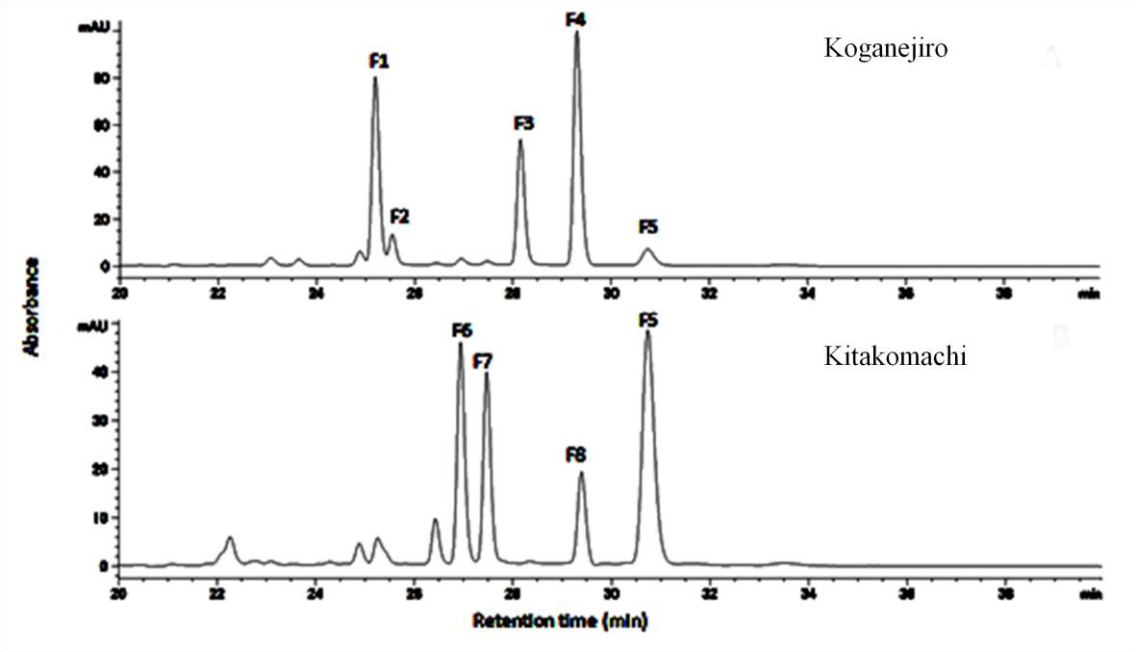


Fig. 5. HPLC Chromatogram of flavonol glycosides extracted from leaves of soybean cultivars Koganejiro and Kitakomachi.

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

F2, kaempferol3-*O*-rhamnosyl-(1→2)-[rhamnosyl-(1→6)-glucoside]

F3, kaempferol3-*O*-rhamnosyl-(1→6)-galactoside

F4, kaempferol 3-*O*-rhamnosyl-(1→6)-glucoside

F5, kaempferol 3-*O*-glucoside

F6, kaempferol 3-*O*-rhamnosyl-(1→2)-galactoside

F7, kaempferol 3-*O*-rhamnosyl-(1→2)-glucoside

F8, kaempferol 3-*O*-galactoside

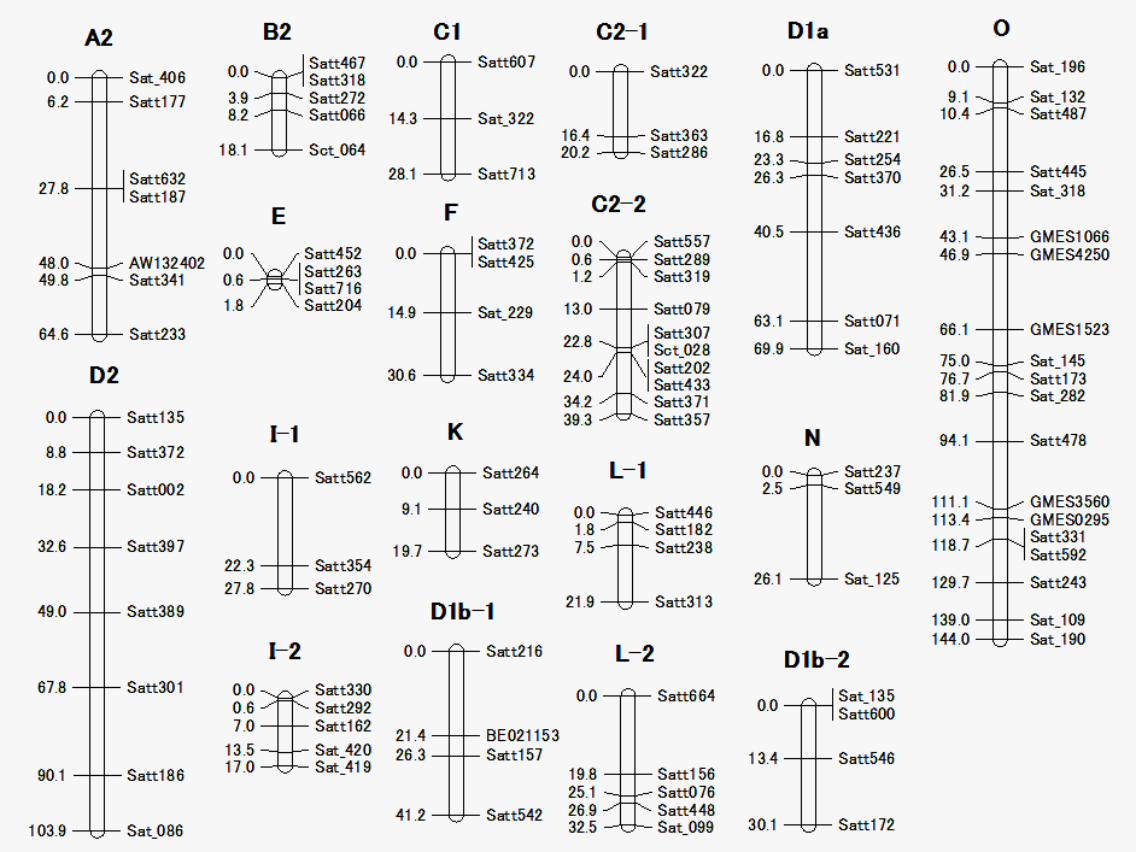


Fig. 6. Linkage map of soybean F_6 population derived from cross between Kitakomachi and Koganejiro. Numbers on the left side of each linkage group indicate the genetic distance (cM) from the top of each linkage group. The name of each linkage group is indicated at the top. The linkage groups were named using the consensus map (Cregan *et al*., 1999).

Group O (Chromosome 10)

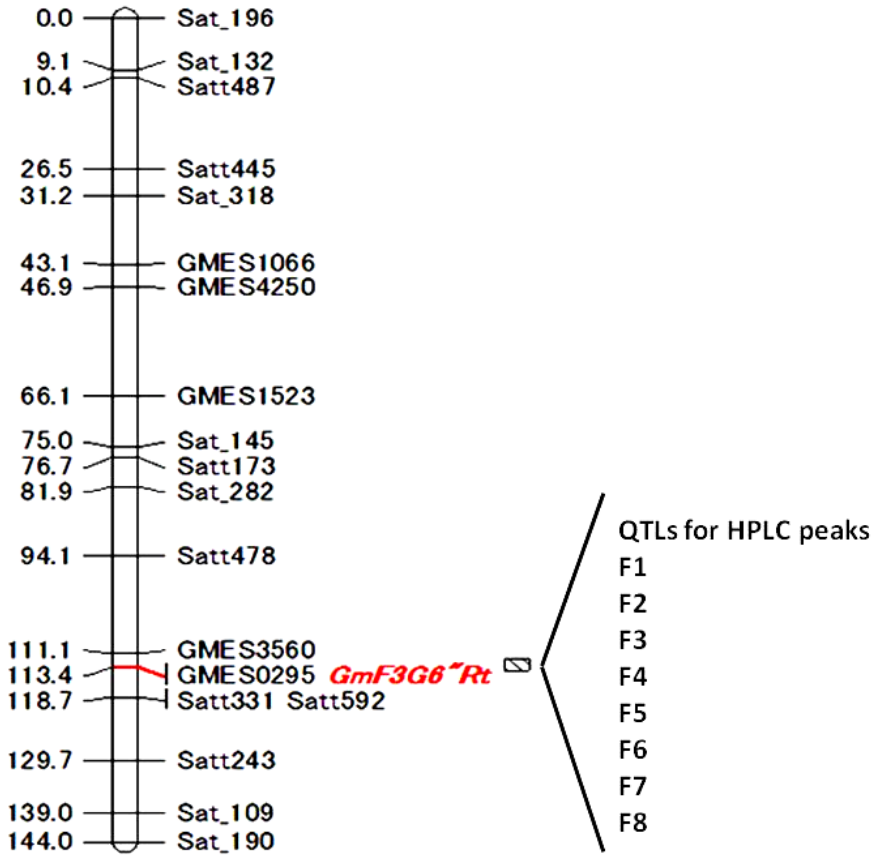


Fig. 7. Linkage mapping of gene *GmF3G6″Rt* responsible for attachment of rhamnose to the 6″-position of glucose or galactose that is bound to the 3-position of kaempferol, observed in a F_6 RILs population derived from a cross of soybean cultivars, Koganejiro and Kitakomachi. Position of QTLs for areas of HPLC peaks F1 to F8 is shown on the right in diagonal hatch square. The name of the linkage group 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.

Fig. 8. Amino acid sequence alignment of soybean *GmF3G6"Rt* from cultivar Koganejiro, *Petunia Rt* encoding anthocyanin 3–*O*–glucoside (1→6) rhamnosyltransferase (accession number: X71059) and citrus Cs1,6RhaT encoding flavanone 7–*O*–glucoside (1→6) rhamnosyltransferase (accession number: ABA18631). Identical amino acids are indicated in black, similar amino acids in gray. Dashes represent gaps introduced to improve the alignment. Position of two–nucleotide deletion found in a cultivar Kitakomachi is shown as an asterisk. The plant secondary product glycosyltransferase (PSPG) box is underlined.

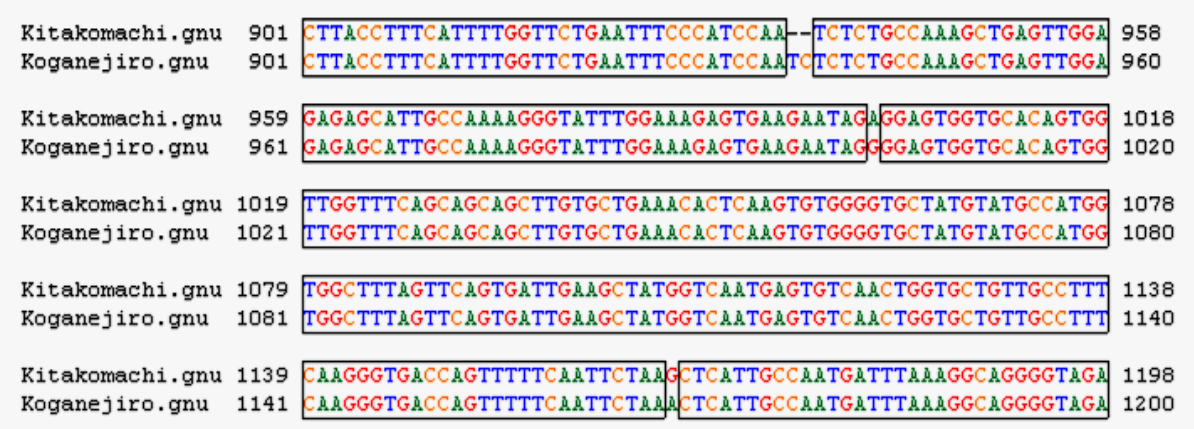


Fig. 9. Sequence alignment of *GmF3G6"Rt* from Kitakomachi and Koganejiro. cDNA sequence of Kitakomachi had a two-base deletion at the nucleotide position 930 and two single-base substitutions, A and G, in the downstream of the deletion when compared with Koganejiro. Identical nucleotides are boxed.

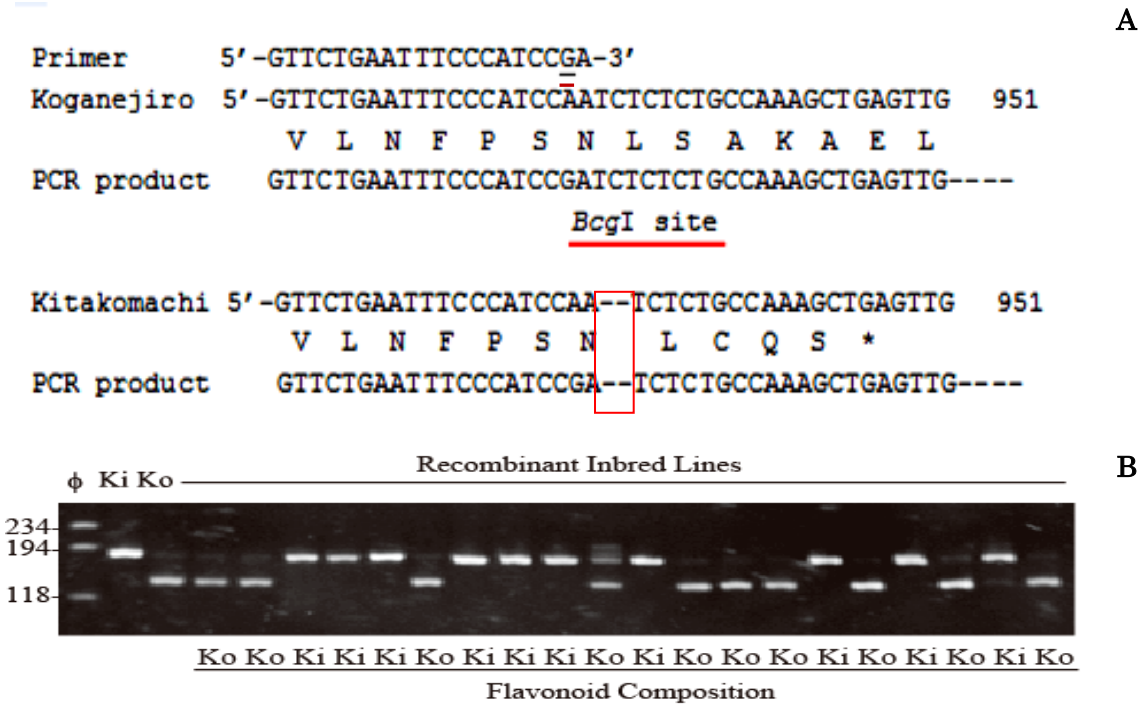
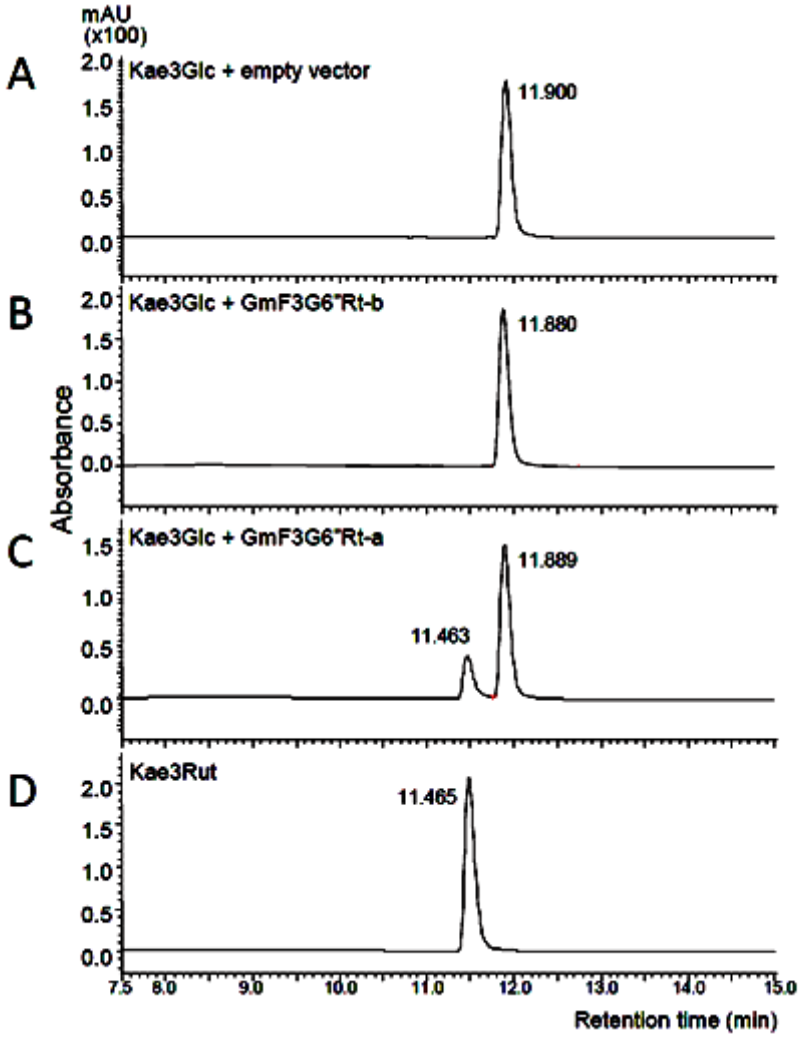


Fig. 10. Outline and results of dCAPS analysis of *GmF3G6''Rt* gene in soybean. (A) Schematic presentation of dCAPS analysis. Partial nucleotide and amino acid sequences around the region polymorphic between Kitakomachi and Koganejiro are exhibited. The PCR primer has a mismatched base of G that is double-underlined. PCR generated an amplification product having *Bcg*I site (CGANNNNNNTGC) that is single underlined in Koganejiro. The two-base deletion in Kitakomachi abolished the restriction site. (B) Results of dCAPS analysis of the parents and the recombinant inbred lines derived from a cross between Koganejiro and Kitakomachi. PCR products amplified with dCAPS primers were digested by *Bcg*I and the digests were separated on an 8 % polyacrylamide gel. ϕ, molecular marker ϕx174/*HaeIII*; Ki, Kitakomachi; Ko, Koganejiro. FGs pattern of the RILs is exhibited below the gel, RILs with heterozygous bands had FGs of Koganejiro-type. The migration of size markers (bp) is shown to the left of the gel.



(Experiment carried out in RIKEN)

Fig. 11. HPLC analyses of the reaction products of *GmF3G6"Rt-a* recombinant protein. Elution profiles of reaction products of ProS2 protein (empty vector) (A), *GmF3G6"Rt–b* protein (B), *GmF3G6"Rt–a* protein (C) with kaempferol 3-*O*-glucoside and the standards (kaempferol 3-*O*-rutinoside, EXTRASYNTHESE, France) (D) are shown. Kae3Glc, kaempferol 3-*O*-glucoside; Kae3Rut, kaempferol 3-*O*-rutinoside.

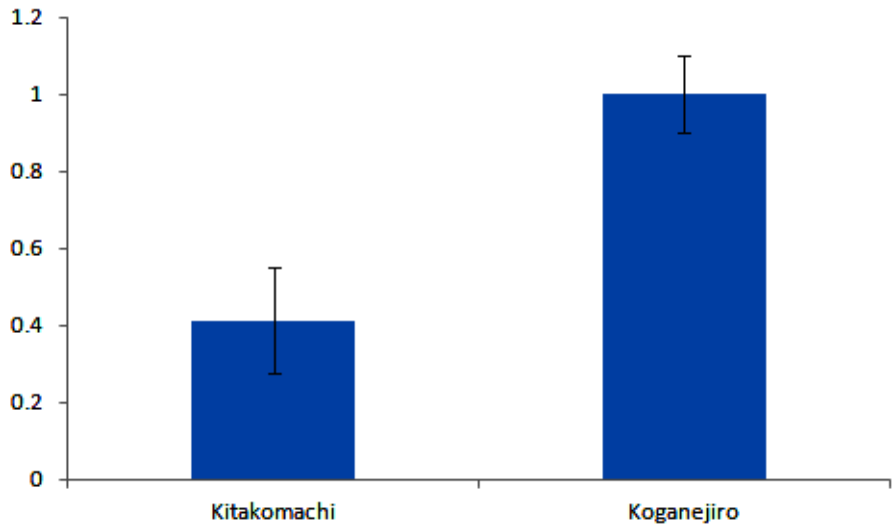


Fig. 12. *GmF3G6"Rt* gene expression of Kitakomachi and Koganejiro. Real-time quantitative reverse transcription PCR of Kitakomachi leaves, at R5 stage, showed a transcript level of *GmF3G6"Rt*, 41.1 % of that in Koganejiro (t=3.43*).

CHAPTER III

3. Analysis of flavonoid glycoside genes segregated in Nezumisaya and Harosoy (NH) population

3.1 Background

Preliminary HPLC experiments suggested that two Japanese cultivars, Nezumisaya and Harosoy, had differences in the amount and composition of FGs extracted from their leaves (Table 1). HPLC chromatogram from Nezumisaya leaves revealed three main peaks corresponding to kaempferols K13 and K14+15, accounted for the 71.2 % of the total flavonols extracted from this cultivar. In contrast, HPLC chromatogram of Harosoy showed three different main kaempferols K1, K2, K3, K4 and K5 that accounted for the 81.6 %. These FGs differences suggested the existence of genes in charge of FGs biosynthesis. However the number of genes, localization in the soybean genome, or function, was unclear.

The objective of this study is to locate, identify, cloning, sequence, verify the function, of candidate genes in charge of FGs biosynthesis in Nezumisaya and Harosoy, and develop NILs to determinate the characters they affected *in planta*. For this purpose, samples of parents and RILs derived from a NH crossing will subjected to HPLC, genetic analysis, SSR analysis, linkage mapping, molecular cloning and sequencing analysis.

3.2 Materials and methods

Plant Material

Plant material was obtained from crossing two soybean cultivars and selfing successively to develop RILs until a F_6 generation. A Japanese soybean Nezumisaya cultivar with yellow hilum, yellow seed coats, gray pubescence and purple-blue flowers (*IIttW1W1w2w2w3w3W4W4WmWmWpWpLnLn*) was crossed with the Canadian cultivar Harosoy with yellow hilum, yellow see coat, gray pubescence and purple flowers (*IIttW1W1W2W2w3w3W4W4WmWmWpWpLnLn*) (Table 2). Flowers of Harosoy were emasculated one day before opening and inoculated with pollen of Nezumisaya, in the year 2004; hybridity of F_1 plants was ascertained based on ovate leaftlets. A total of 120 RILs from the F_6 generation were developed without any selection using single-seed descent method, seeds were planted at the National Institute of Crops Sciences, Tsukuba, Japan on 9th June 2011. Fertilizer application and planting conditions were the same with KK population

Flavonoid extraction

Leaf samples were collected from a total of 91 RILs and the parents; and were used for methanol (MeOH) flavonoid extraction. Extraction method was same as KK population.

HPLC and genetic analysis

Quantitative HPLC analysis of the extracted samples was performed with the Agilent 1100 HPLC System using an L-column 2 ODS [I.D. 60 x 150 mm (Chemicals

Evaluation and Research Institute)] at a flow rate of 1.0 mL/min and detection wavelength range of 190-700 nm. An isocratic method was performed to the 96 samples and the eluent solvents were phosphoric acid/acetonitrile/H_2O (0.2:18:82). The amount of flavonol glycosides was estimated from the pertinent peak area in a HPLC chromatogram with a detection wavelength of 350 nm. HPLC results were compared with seven authentic specimens from soybean cultivars and other plant species (Iwashina *et al*., 2006-a, 2010; Murai *et al*., 2013). In addition to HPLC comparisons eight kaempferol glycosides isolated Nezumisaya and Harosoy (parents) were identified by UV, LC-MS, acid hydrolysis and ^{1}H and ^{13}C NMR. The genetic model for FGs composition was hypothesized based on the HPLC chromatogram of the parents and the RILs; and the chemical structure of the pertinent peaks.

SSR analysis

Genomic DNA samples were extracted from the parents and the 91 RILs; and were subjected to SSR analysis. Methods for DNA extraction and SSR analysis were the same as KK population.

Linkage mapping

All the markers were tested by Chi-square test for segregation in 1:1 ratio. A linkage map was constructed using MAPMAKER/EXP. ver.3.0 (Lander *et al*., 1987) with a LOD score of 3.0; and the linkage groups designation was followed by Cregan *et al*., (1999).

Molecular cloning

RNA was extracted from trifoliated leaves (200 mg) from Nezumisaya and Harosoy using the TRIZOL Reagent (Invitrogen) according to the manufacturer's instructions. Reverse transcription of total RNA was performed using the Superscript III First-Strand Synthesis System (Invitrogen) and an oligo(dT) primer according to the manufacturer's instructions, then the full-length cDNA was cloned from Nezumisaya and Harosoy by end-to-end PCR using a pair of PCR primers (Table 8) designed based on the genome sequence of US cultivar Williams 82 deposited in the soybean genome database. 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 in a total volume of 25 μl. PCR was performed according to the program, initial denaturation 30 sec at 94 °C followed by 35 cycles of 30 sec denaturation at 94 °C, 1 min annealing at 59 °C, 1 min extension at 72 °C and 7 min final extension at 72 °C. Finally, the ~1.4 kbp PCR product was cloned into pCR 2.1 vector and sequenced. 5' upstream region was also cloned in the same conditions with the CDS using primers shown in Table 8.

Sequencing analysis

Sequence analysis procedure was the same as KK population. Primer sequences are exhibit in Table 8. Transcription factor binding motifs in the 5' upstream region, were searched with TFSEARCH ver. 1.3 (http://mbs.cbrc.jp/research/db/TSEARCH.html)

dCAPS analysis

The first primer contained a nucleotide (C) that mismatched with its target DNA

to artificially create a restriction site for *EcoN*I (CCTNNNNNAGG) in Harosoy (Fig. 17A). The Substitution within the restriction site of Nezumisaya would abolish the restriction site in the amplified product to generate a polymorphism. PCR mixture contained 30 ng of genomic DNA, 10 pmol of each primer, 5 pmol of nucleotides and 1 unit of ExTaq in 1 x ExTaq Buffer in a total volume of 25 µl. dCAPS PCR program consisted on initial 30 sec denaturation at 94°C, followed by 30 cycles of 30 sec denaturation at 94°C, 1 min annealing at 59°C and 1 min extension at 72°C; and a final 7 min extension at 72°C completed the program. The amplified products were digested with *Eco*NI enzyme; and DNA fragments were separated in 8 % nondenaturing polyacrylamide gel, stained with ethidium bromide, and visualized under UV light.

Gene expression assays

Real-time quantitative reverse transcription PCR was performed to estimate gene expression of *GmF3G6"Gt* in Nezumisaya and Harosoy. Total RNA was extracted from three replicated leaves samples of Nezumisaya and Harosoy at R5 stage. cDNA synthesis, PCR conditions and data normalization were carried out as described in KK population.

3.3 Results

HPLC Analysis

HPLC analysis revealed that Harosoy had eight primary peaks corresponding to FGs, 4.5 min (F1), 6.1 min (F2), 6.7 min (F3), 7.4 min (F4), 9.6 min (F5), 11.2 min (F6), 13.7 min (F7) and 16.3 min (F8). Nezumisaya had seven primary peaks, 4.9 min (F9),

5.0 min (F10), 6.1 min (F2), 6.4 min (F11), 6.6 min (F12), 9.6 min (F5) and 11.2 min (F6) (Fig. 13). Based on comparison with authentic specimens, the peaks correspond to the followings FGs:

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

Peaks F1, F3 and F4 that were specific to Harosoy and corresponded to FGs having glucose at the 6"-position of glucose or galactose that is bound to the 3-position of kaempferol, whereas FGs with 6" glucose were absent in Nezumisaya. Peaks F9, F10 and F12 that were specific to Nezumisaya corresponded to FGs having glucose at the 2"-position of glucose or galactose that is bound to the 3-position of kaempferol, whereas these components were absent in Harosoy.

Genetic analysis

Genetic analysis revealed that among 91 RILs, 21 RILs had peaks of Harosoy-

type, 23 RILs had peaks of Nezumisaya-type, 17 RILs had a peak distribution designated as ‘type 3’ that lacked three peaks, F1, F3 and F4 from the Harosoy-type; and 28 RILs had a peak distribution designated as "type 4", having peaks of a mixture of peaks from both cultivars in addition to a unique peak F13 (Table 9). The remaining two RILs had unique peak distribution and might have heterozygous allele(s). Excluding the two RILs, the segregation fitted to 1:1:1:1 ratio suggesting that two genes control the FGs pattern (χ^2 = 2.82, P = 0.42), one of them may be responsible for attachment of glucose to the 6"-position probably encodes flavonol 3-*O*-glucoside-(1→6)-glucosyltransferase. And the other gene could attach glucose to the 2"-position; and probably encodes 3-*O*-glucoside-(1→2)-glucosyltransferase

SSR analysis and linkage mapping

Among the 465 SSR markers tested, 185 markers exhibited polymorphism between the parents and distinctly segregated in the RILs and were used for linkage mapping. A total of 99 markers fell into 28 linkage groups spanning 2172 cM (Fig. 14). For mapping of the gene responsible for attachment of 6" glucose, RILs having FGs composition of the Harosoy type and "type 3" were considered to have the genotype of Harosoy, whereas RILs having FGs composition of the Nezumisaya type and "type 4" was considered to have the genotype of Nezumisaya. Linkage mapping revealed that the gene responsible for attachment of glucose to the 6"-position was located in the molecular linkage group I (chromosome 20) between the SSR markers Satt623 and Sat_419; and the gene responsible for attachment of glucose to the 2"-position was located in the molecular linkage group C2 (chromosome 06) between the SSR markers Satt307 and Sat_202 (Fig. 14).

Molecular Cloning

Survey of the genome sequence of a US cultivar Williams 82 suggested three genes similar to GT gene, Glyma20g33810, Glyma20g33820 and Glyma20g33831. These genes were aligned spanning about 15 kb, between Satt623 and Sat_419, BLAST analysis suggested that Glyma20g3320 and Glyma20g33831 might be truncated and might be pseudogenes. In contrast, Glyma20g33810 seemed to be functional. Therefore entire CDS of Glyma20g33810 was amplified by RT-PCR from Harosoy and cloned. Sequencing analysis revealed that the CDS was 1386 bp long encoding 462 amino acids; the gene was designated as *GmF3G6"Gt*. BLAST analysis suggested that *GmF3G6"Gt* had 88 % amino acid similarity with *GmF3G6"Rt* of soybean encoding flavonol 3-*O*-glucoside-(1→6)-rhamnosyltransferase (Fig. 15). In contrast to Harosoy, entire CDS of *GmF3G6"Gt* could not be amplified by RT-PCR from Nezumisaya. So, the genomic fragment containing the entire CDS was PCR amplified using the genomic DNA as a template. The region corresponding to the CDS is 1386 bp long; capable of encoding 462 amino acids similar to Harosoy. A single nucleotide substitution was found at nucleotide position 1041, but amino acids were identical between the cultivars. Comparison with the genome sequence of Williams 82 and genome sequence of Nezumisaya suggested that *GmF3G6"Gt* had not intron similar to the *GmF3G6"Rt*.

In contrast to the CDS, 5'-upstream region of *GmF3G6"Gt* for Harosoy and Nezumisaya was quite polymorphic. Alignment of Harosoy and Nezumisaya 5'-upstream region of *GmF3G6"Gt* showed six single-nucleotide substitutions, an indel of 61 nucleotides and a substitution of 26 consecutive nucleotides (Fig 16A). The last substitution may have occurred by a duplication of a fragment of 26 nucleotides in Harosoy. The nucleotide motif (AACTACCCG) similar to a binding site of myb-

homologous P gene responsible for phlobaphene pigmentation in maize (Groteworld and Peterson, 1994) existed in the duplicated region.

dCAPS analysis

Bands with molecular size of about 170 bp were amplified with the dCAPS primers in Harosoy and Nezumisaya. *Eco*NI enzyme digestion generated a band of about 150 bp in Harosoy, whereas the PCR amplicon of Nezumisaya was unaffected. Banding pattern of the RILs was classified into Harosoy-type, Nezumisaya-type and heterozygous type (having both bands) (Fig. 16B). In total, 37 RILs had bands of Harosoy type, 51 RILs had bands of Nezumisaya type and 3 RILs had both bands. Banding patterns were co-segregated with FGs patterns; RILs with FGs of Nezumisaya type had bands of only Nezumisaya-type, whereas RILs with FGs of Harosoy type had bands of only Harosoy-type or both bands.

3.3.1 Gene expression assays

At the R6 stage, transcript level of *GmF3G6''Gt* in leaves of Nezumisaya was 0.2 % of that in Harosoy ($t=3.43$, $P=0.03$*) (Fig. 17)

3.4 Discussion

Soybean cultivars, Harosoy and Nezumisaya have gray pubescence and deposit predominantly kaempferol derivatives in leaves. However, HPLC analysis showed FGs contents were different. HPLC analysis and comparison with authentic specimens suggested that FGs contents from Harosoy and Nezumisaya were different, Nezumisaya predominately had FGs with glucose at the 2"-position of glucose or galactose that is

bound to the 3-position of kaempferol , whereas FGs of Harosoy were devoid of the 2" glucose (Shaokang *et al*. 2013 not published results). Conversely, Harosoy predominately had FGs with glucose at the 6"-position of glucose or galactose that is bound to the 3-position of kaempferol, but Nezumisaya was absent of these FGs.

Genetic analysis for Harosoy and Nezumisaya showed a distribution (1:1:1:1) suggesting that two genes were in control of FGs pattern. One of the genes may be responsible for attachment of glucose to the 6"-position and probably encodes flavonol 3-*O*-glucoside-(1→6)-glucosyltransferase. Harosoy may have a dominant and Nezumisaya may have a recessive allele of the gene. The other gene was involved in attachment of glucose to the 2"-position and encoded flavonol 3-*O*-glucoside-(1→2)-glucosyltransferase. Nezumisaya may have a dominant and Harosoy a recessive allele of the gene. RILs of the 'type 3', in which FGs with 2" glucose and FGs with 6" glucose were absent, may have double-recessive alleles. RILs of the "type 4", in which FGs with 2" glucose and FGs with 6" glucose were abundant, may have double-dominant alleles.

The gene responsible for attachment of glucose to the 6"-position was mapped in the molecular linkage group I (chromosome 20). Survey of the genome sequence of a US cultivar Williams 82 suggested existence of three genes similar to GT genes, Glyma20g33810, Glyma20g33820 and Glyma20g33831, were aligned at the region. BLAST analysis suggested that the last two genes might be truncated and probably pseudogenes.

cDNA of Glyma20g33810 was cloned and sequenced from Harosoy and designated as *GmF3G6"Gt*. The gene was 1389 bp long encoding 462 amino acids. The *GmF3G6"Gt* had no intron, in consistent with FGs genes of other plant species, in which either existence of one or none intron is predominant (Paquette *et al*. 2003). In

contrast to Harosoy, CDS of *GmF3G6"Gt* could not be amplified by RT-PCR from Nezumisaya. Genome sequence of the region corresponding to the CDS in Nezumisaya was similar to Harosoy. Further, the deduced amino acid sequence was identical between the cultivars. However, 5'-upstream region was quite polymorphic between the cultivars; six single-nucleotide substitutions, an indel of 61 nucleotides and a substitution of 26 consecutive nucleotides existed. The last substitution may have occurred by a duplication of a fragment of 26 nucleotides in Harosoy. The nucleotide motif similar to binding site of myb-homologous P gene responsible for phlobaphene pigmentation in maize was duplicated in Harosoy because of the fragment duplication. Transgenic experiments may be necessary to determine which polymorphism(s) is critical for gene expression.

A CAPS marker was designed to detect the fragment substitution in *GmF3G6"Gt*. Banding patterns of the CAPS marker was co-segregated with FGs patterns; RILs with FGs of Nezumisaya type had only bands of Nezumisaya type, whereas RILs with FGs of Harosoy type had bands of only Harosoy type or both bands. Three RILs with both bands may be heterozygotes for *GmF3G6"Gt*. CAPS marker confirmed the nucleotide polymorphism, association between genotype and FGs composition; and could be used to develop NILs.

Based on the above results, the gene we cloned likely corresponds to flavonol 3-*O*-glucoside-(1→6)-glucosyltransferase gene (*Fg1*), but the function should be ascertained by functional analysis, as performed in Kitakomachi and Koganejiro. In addition, NILs should be developed to determine the characteristics this gene affects *in planta*.

Table 8. Primers used for cloning, sequencing, dCAPS, real-time PCR and analysis of *GmF3G6"Gt* gene.

Purpose	Target gene	Forward primer (5'-3')	Reverse primer (5'-3')
Cloning of CDS for sequencing	*GmF3G6"Gt*	CCATAAGGAACTTGCAGAAA	CAAGAAATCCGAATCCTGAA
Cloning of 5' upstream region and dCAPS analysis	*GmF3G6"Gt*	GTGTCTGATGGATATTTTGG	CAGCTGACAAGAAAGTGATA
Sequencing of CDS	*GmF3G6"Gt*	1. ACCTTACTCAACCTCATGTA 2. GATCAAATCAAAGAAGTGGC	1. AAAGGTAAGCCACTGAGTTC 2. TGAGATTTGCAGCCAAGTTT
Sequencing of 5' upstream region	*GmF3G6"Gt*	TGGGAATGTAAATGACTCTC	TTCCTTATGGATGTTGGCTT
Cloning for functional analysis[a]	*GmF3G6"Gt*	GCCG<u>GAGCTC</u>ATGTCTTGTGAAGTTGTGAAC *Sac*I	GCCG<u>AAGCTT</u>CTAAGCCATAGACTTCAACT *Hind*III
Real time-PCR	*GmF3G6"Gt*	ATGTCTTGTGAAGTTGTGAAC	GAGGCTTTTTAAGATCCTCAA
	Actin	CTGGGGATGGTGTCAGCCACAC	CACCGAACTTTCTCTCGGAAGGTG

[a] Restriction sites incorporated in primers, *SacI* site in forward primer and *HindIII* site in reverse primer are underlined.

Table 9. Genetic analysis of NH chromatograms.

A total of 91 RILs were segregated into 21 RILs of Harosoy-type; 23 RILs of Nezumisaya-type; 17 RILs of type 3; and 28 RILs of type 4. Two samples 55.5 and 73 had a unique distribution and might have heterozygous allele(s). Chromatogram segregation fitted a 1:1:1:1 ratio (χ^2=2.82, P= 0.42) suggesting that two genes were in charge of controlling FGs composition, one of them probably encodes flavonol 3-*O*-glucoside-(1→6)-glucosyltransferase.

Harosoy-type Peaks: 4.5, 6.1, 6.7, 7.4, 9.6, 11.2 min		Nezumisaya-type Peaks: 4.9, 5.0, 6.1, 6.4, 6.6 min		Type 3 Peaks: 6.1, 9.6, 11.2, 13.7, 16.8 min		Type 4 Peaks: 3.8, 4.5, 4.9, 5.0, 6.1 min	
1	59	5	62	2	83	7	69
4	65	6	71	3	98	8	72
14	81	16	77	9	106	13	88
15	84	28	80	10		19	89
17	85	34	86	11		21	92
18	103	38	93	12		24	94
20	110	40	96	26		25	95
27		42	104	29		37	99
33		43	111	30		44	100
39		47		31		45	101
41		49		35		52	105
50		53		55		57	107
54		58		75		67	108
56		60		79		68	109

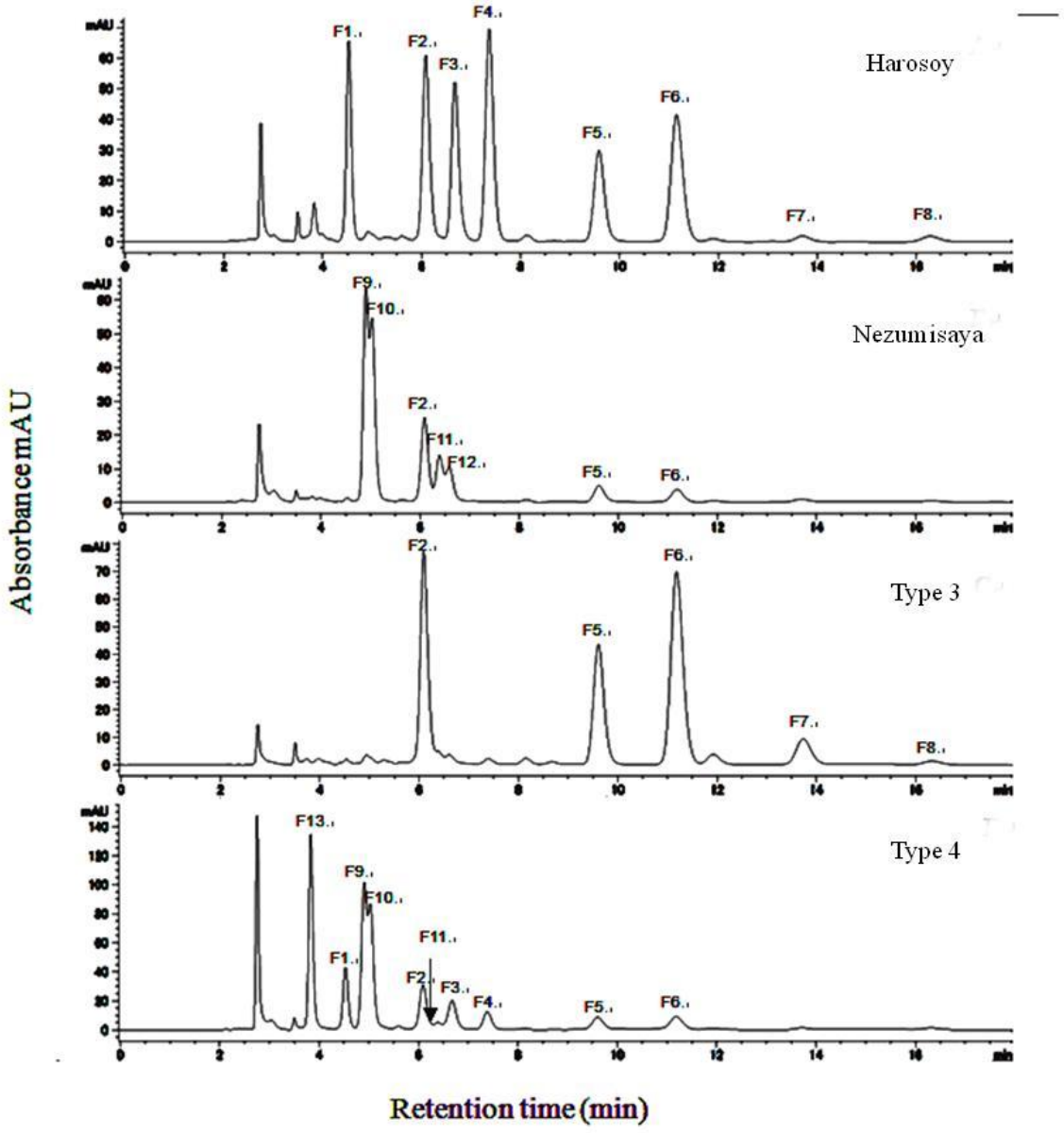


Fig. 13. HPLC chromatogram of flavonol glycosides extracted from leaves of soybean cultivars Harosoy and Nezumisaya; and two typical peak distributions of Type 3 and Type 4 found in RILs derives from cross of the cultivars. 100 mg of trifoliated leaves was extracted with 1 mL of MeOH. HPLC conditions, eluent phosphoric acid/acetonitrile/H_2O (0.2:18:82). Flow rate 1.0 mL/min, sample injection 10 μL, detection wavelength 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 Kaempferol glycoside

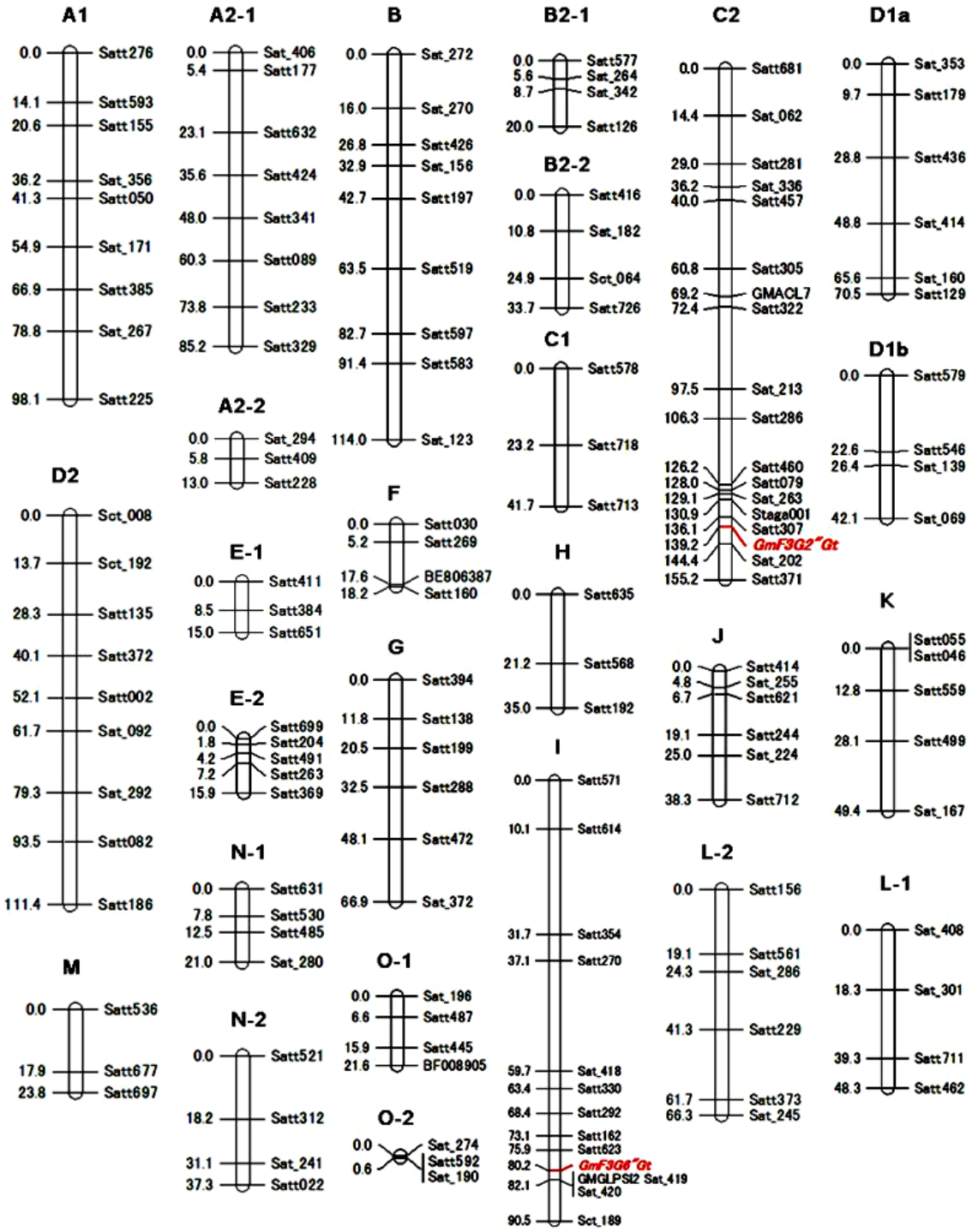


Fig. 14. Linkage map of soybean F_6 RIL population derived from a cross between Nezumisaya and Harosoy. Genes responsible of glucose attachment to 6"-position *GmF3G6"Gt* and 2"-position *GmF3G2"Gt* are shown in red font. The name of each linkage group is indicated at the top, numbers on the left side indicate the genetic distance (cM) from the top of each linkage group. The linkage groups were named using the consensus map (Cregan *et al.*, 1999).

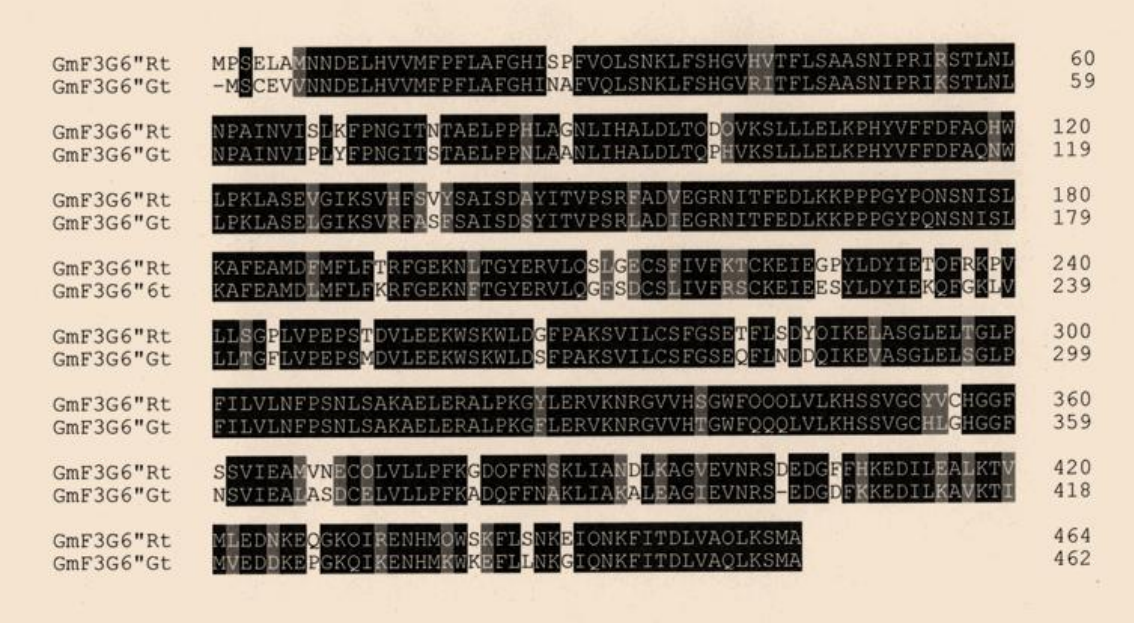


Fig. 15. Amino acid alignment of soybean *GmF3G6"Gt* from cultivar Harosoy and soybean *GmF3G6"Rt* from cultivar Koganejiro. Identical amino acids are indicated in black, similar amino acids in gray.

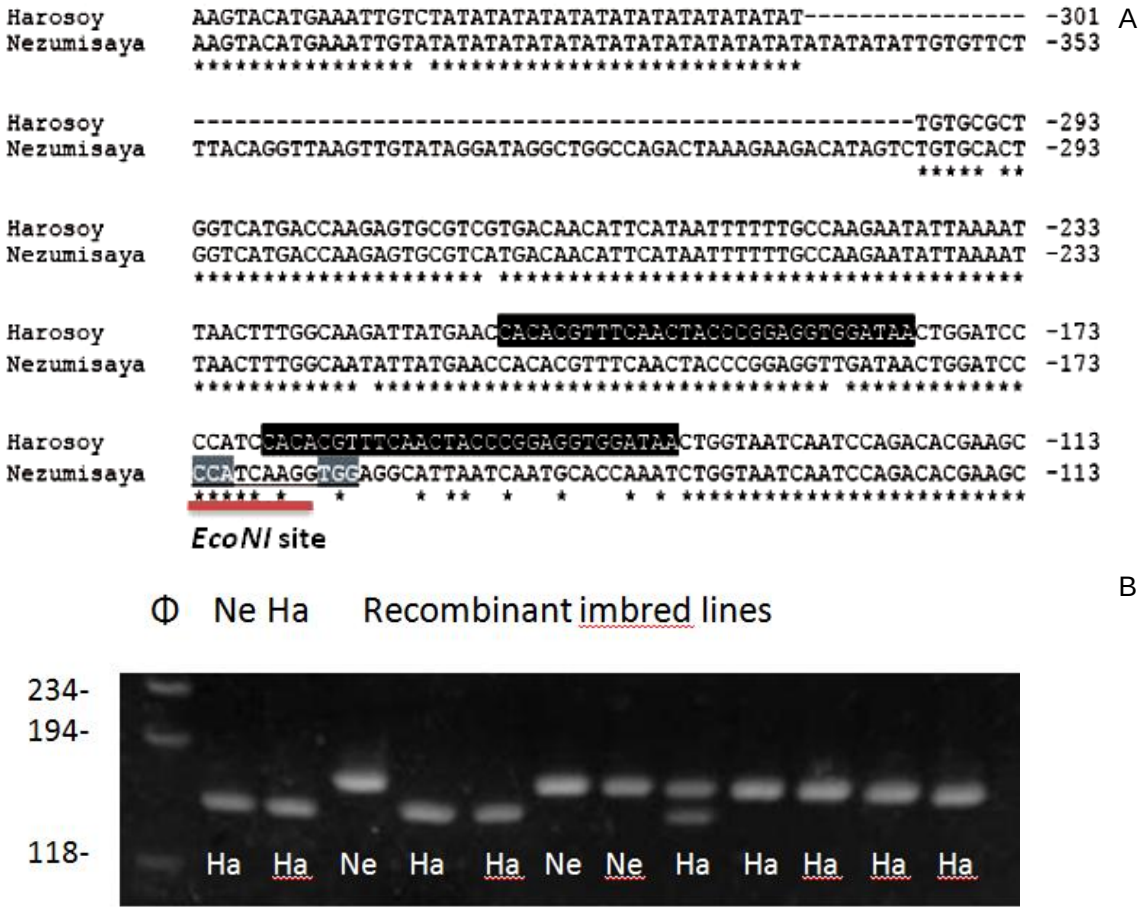


Fig. 16. Outline and results of dCAPS analysis of *GmF3G6"Gt* gene in soybean. (A) schematic presentation of dCAPS analysis. Partial nucleotide and amino acid sequences around the region polymorphic between Nezumisaya and Harosoy are exhibited. The PCR primer has a mismatched base of G that is underlined. PCR generated an amplification product having *EcoN*I site (CCTNNNNNAGG). The first primer contained a nucleotide that mismatched with its target DNA to artificially create a restriction site *EcoN*I in Harosoy that is underlined. The two-base deletion in Kitakomachi abolished the restriction site. (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 *EcoN*I and the digests were separated in 8 % polyacrylamide gel. ϕ, molecular marker ϕx174/*HaeIII*; Ne, Nezumisaya; Ha, Harosoy. FGs pattern of the recombinant inbred lines is exhibited below the gel. Heterozygous plants had FGs of Harosoy-type. The migration of size markers (bp) is shown to the left of the gel.

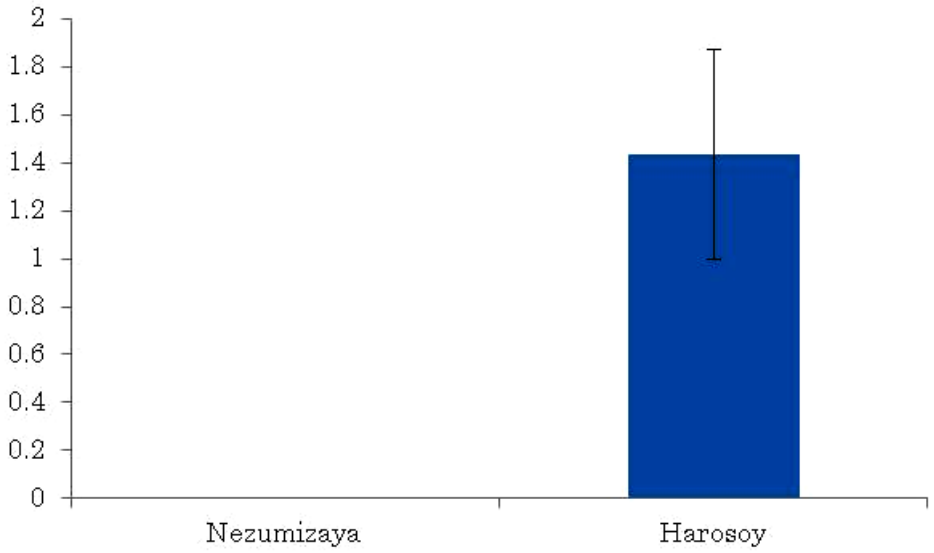


Fig. 17. *GmF3G6"Gt* gene expression of Nezumisaya and Harosoy. Real-time quantitative reverse transcription PCR of Nezumisaya leaves, at R6 stage. Transcript level for *GmF3G6"Gt* was 0.2 % of that in Harosoy (t = 3.43*).

CHAPTER IV

4. General discussion

Chemical structure

HPLC analysis revealed differences in FGs from cultivars Kitakomachi, Koganejiro, Nezumisaya and Harosoy. From Kitakomachi and Koganejiro, a total of eight FGs were identified, four of them in Koganejiro with rhamnose at the 6''-position of the glycoside attached to the 3-*O*-kaempferol; and from Nezumisaya and Harosoy, a total of twelve FGs, three of them in Harosoy with glucose at the 6''-position of the glycoside attached to the 3-*O*-kaempferol and other three in Nezumisaya with glucose at the 2''-position of the glycoside attached to the 3-*O*-kaempferol. These glycosides distribution among FGs, indicated that at least three genes were in charge of their synthesis, *Fg1* flavonol-3-*O*-glycoside-(1→6)-glucosyltransferase, *Fg2* flavonol-3-*O*-glycoside-(1→6)-rhamnosyltransferase, found in this study; and probably a third gene *Fg3* flavonol-3-*O*-glycoside-(1→2)-glucosyltransferase. Thus, either glucose or galactose was attached to the 3-position of kaempferol. This contrasts with an earlier report that only glucose was attached to the 3-position (Buttery and Buzzell, 1975). Probably, the analytical method used could not discriminate between glucose and galactose. The same authors, Buzzell and Buttery (1974), reported another gene *Fg4* encoding flavonol 3-*O*-glucoside (1→2) rhamnosyltransferase. In addition, FGs having rhamnose at the 4″-position of 3-*O*-galactose have been identified (Murai *et al*., 2013), kaempferol-3-*O*-rhamnosyl-(1→4)-[glucosyl-(1→6)-galactoside] from Harosoy and kaempferol-3-*O*-rhamnosyl-(1→4)-[rhamnosyl-(1→6)-galactoside] from Koganejiro,

exhibiting a new glycosidic bond [glycoside-(1→4) rhamnose], suggesting the existence of a new FGG gene flavonol 3-*O*-glucoside (1→4) rhamnosyltransferase. This new gene could transfer rhamnose to the 4''-position of glucose or galactose, attached to the 3''-position of kaempferol or quercetin. We designated this new gene as *Fg5*.

Notwithstanding, the schematic diagram on the genetic control of FGs biosynthesis proposed by Buzzell and Buttery (1975) should be revised. We propose a new scheme for flavonol glycosides biosynthesis in soybean leaves shown in Fig. 19. The scheme conserves four basic FGG genes described by Buttery and Buzzel (1975), *Fg1*(K-3-*O*-Glc-(1→6)-Glc) and *Fg2*(K-3-*O*-Glc-(1→6)-Rha), described in this study; *Fg3*(K-3-*O*-Glc-(1→2)-Glc) and *Fg4*(K-3-*O*-Glc-(1→2)-Rha); and includes the new gene *Fg5* (K-3-*O*-Glc-(1→4)-Rha) and the interactions with the other four FGG genes, to assembly new flavonol di-glycosides and tri-glycosides.

Gene structure

In this study, two genes in charge of transferring glycosides to the glycoside attached to the 3-*O*-kaempferol position were cloned. The first gene had the function of attaching rhamnose to the 6″-position of glucose or galactose bound to the 3'-position of kaempferol and was cloned from KK. The second gene attached glucose to the 6''-position of galactose or glucose bound to the 3'-position of kaempferol and was cloned from NH. Both genes are FGG genes, designated as *GmF3R6''Rt* and *GmF3R6''Gt* and corresponded to the characteristics of *Fg2* and *Fg1* flavonol glycosyltransferase genes, respectively, reported by Buttery and Buzzel (1975).

Phylogenic analysis of *GmF3R6''Rt* and *GmF3G6''Gt* showed high similarity between genes. *GmF3R6''Rt* and *GmF3G6''Gt* belonged to the family 1

glycosyltransferases, *GmF3R6"Rt* was designated as UGT79A6 by UGT nomenclature committee (Mackenzie *et al*., 1997), the flavonoid glycosyltransferase phylogenic tree (Fig. 20) suggests that both genes belongs to the 1,2/1,6 branch forming the glycosyltransferase gene cluster 4. *GmF3G6"Rt* and *GmF3G6"Gt* comprised a separate branch in the tree, suggesting that both genes come from duplication of the same antecessor and this was a recent event (Yonekura-Sakakibara and Hanada, 2011; Ono *et al*., 2010).

Application to agriculture

Flavonol glycosides in leaves were determined in soybean cultivars Kitakomachi, Koganejiro, Harosoy and Nezumisaya. All the cultivars showed differences in their flavonol glycoside type and concentration. Some of them are known to have biological activity against some health disorders. Leaves of Koganejiro contained Kaempferol-3-*O*-rhamnosyl-(1→6)-glucoside and Kaempferol-3-*O*-rutinoside, these flavonol glycosides are related to the decrease of the blood pressure in rats (Mansoor *et al*., 2006). Leaves of Harosoy contained Kaempferol-3-*O*-rhamnosyl-(1→6)-galactoside and Kaempferol-3-*O*-robinobioside, which are related to inhibition of lymphocyte proliferation *in vitro* (Brochado *et al*., 2003). Leaves of Kitakomachi contained Kaempferol-3-*O*-rhamnosyl-(1→2)-glucoside and Kaempferol-3-*O*-neohesperidose, these kaempferol glucosides are related to an insulinomimetic effect in rats, stimulating the glucose uptake in the rat soleus muscle (Zanatta *et al*., 2008). And, leaves of Nezumisaya contained Kaempferol-3-*O*-glucosyl-(1→2)-glucose and Kaempferol-3-*O*-sophorose, which have anti-inflammatory properties in human

endothelial cells (Kim *et al*., 2012). Certainly, these FGs identified give an added value to these cultivars, which can be used in the pharmaceutical industry or traditional natural medicine.

On the other hand, similar FGs contents could lay on soybean cotyledons from soybean sprouts, which can be eaten as functional foods. In recent years, Japanese food has been coming back under the global spotlight—as the formula for Japanese longevity and for its health benefits. The concept of functional foods has evolved as the role of food in the maintenance of health, well-being and prevention of diseases; and has received increased scientific and commercial interest (Jooyandeh, 2011). Seeds and sprouts from legume crop plants have received attention as functional foods, because of their nutritive values including amino acid, fiber, trace elements, vitamins, flavonoids and phenolic acids. Consumption of seeds and sprouts has become increasingly popular among people interested in improving and maintaining their health status by changing dietary habits, as they lowering the risk of various diseases and/or exerting health promoting effects in addition to its nutritive value (Chon, 20013).

In soybean, the nutritional value of sprouts change during the germination; the free amino acid content increases and the vitamin C content increases approx. to its 200-fold value compared to the non-germinated seed (Marton *et al*., 2010), at the same time, sprouting increases the total flavonoid and aglycone contents in soybean, but interestingly not in black soybean (Lin and Lai, 2006); and some flavonoids like rutin, apigenin and quercetin are found in a considerable concentrations (Konar, 2013). However, genetic characteristics of cultivars and sprouting conditions like, light, UV, IR radiation, temperature and days of spouting (Marton *et al*., 2010), have effects on flavonoid contents from soybean sprouts. Our futures studies should focus on evaluate

the FGs contents in sprouts, their variation among cultivars and environmental requirements.

Furthermore, GMO soybean overexpressing FGs could be an alternative to increase the amount and kinds of FGs, in leaves and cotyledons. Different studies suggest that this is possible. Muir *et al*., (2001) increased the total flavonol content of tomato, to similar level of onions, by using petunia *chi-a* gene encoding chalcone isomerase, and proposed a rich flavonol tomato processed paste as functional foods. Kostyn *et al.* (2013) increased the content of flavonoids in potato (*Solanum tuberosum*) through the over expression of dihydroflavonol reductase gene (DFR). Transgenic plants showed elevated levels of flavonoids and exhibit nutritional benefits in rats, also made plants more resistant to pathogens. Hypothetically, overexpressing *Fg1* and *Fg2* could result in an increment of the total FGs contents in soybean plant leaves, many of these FGs can be used as medicines, and could be well sold to the pharmaceutical industry; also the sprouts could also have a high concentration of FGs and can be used direct as functional foods, as a part of animal diet, or in concentrated animal feed. Probably this increase in FGs sprouts could help breeding more healthy animals.

In conclusion, two FGG genes *GmF3G6"Rt* and *GmF3G6"Gt* were segregated from two populations of the soybean crossings KK and NH; and cloned for first time in plants, those genes correspond to *Fg2* and *Fg1* reported by Buttery and Buzzel (1975), and transfer rhamnose and glucose to the glucose/galactose attached to the 3-*O*-Kaempferol, respectively. However, visible phenotypic or physiological characteristics these genes confer to soybean plants are still unknown. *In situ* evaluation of KK NILs did not show differences between photosynthesis rate and chlorophyll contents, but more evaluations under controlled conditions that also include leaf density,

yield and stoma distribution, should be conducted.

Finally, soybean leaves have an added value to soybean plant due to the FGs contents. We propose soybean leaves can be used in the pharmaceutical industry or traditional medicine; and sprouted grains can be eaten as functional food with high benefits to nutrition and health. This way the farmers can find more profitable soybean cultivation, enrich their diet and help to preserve the soybean germoplasm to future generations

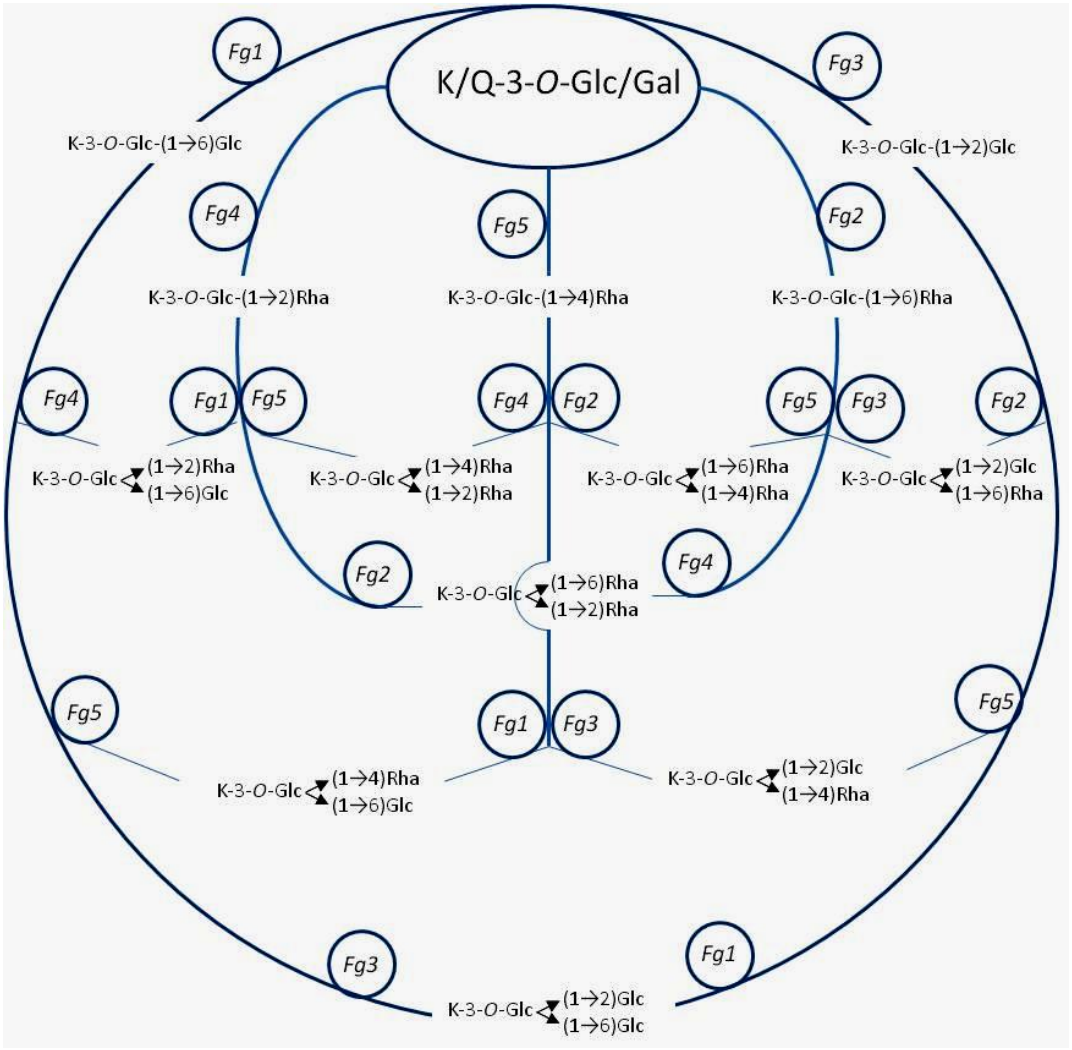


Fig. 19. Newly proposed flavonol glycoside biosynthesis scheme in soybean leaves. Starting from a kaempferol/quercitin-3-*O*-glucose/galactose (K/Q-3-*O*-Glc/Gal), five FGG genes *Fg1* (K-3-*O*-Glc-(1→6)-Glc), *Fg2* (K-3-*O*-Glc-(1→6)-Rha), *Fg3* (K-3-*O*-Glc-(1→2)-Glc), *Fg4* (K-3-*O*-Glc-(1→2)-Rha) and *Fg5* (K-3-*O*-Glc-(1→4)-Rha). These genes attach glucose (Glc) to the 2" or 6"-position, or rhamnose (Rha) to the 2", 4" or 6"-position of the glucose/galacatose attached to the 3-position of kaempferol/quercetin, to generate different flavonol di-glycosides and tri-glycosides.

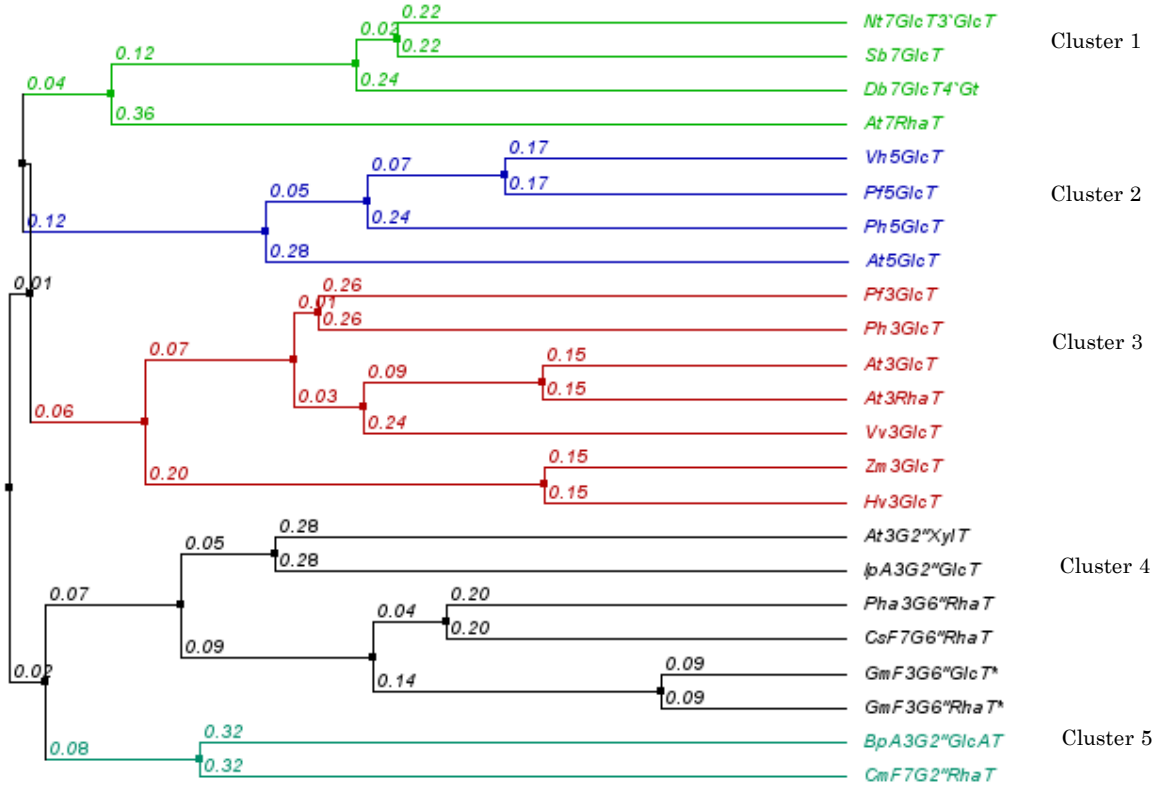


Fig. 20. Unrooted molecular phylogenetic tree for *GmF3G6''Gt* and *GmF3G6''Rt* and some flavonoid glycosyltransferases. The bootstrap values are given to each linked branch and the GenBank accession numbers for the sequences are shown in parentheses: At3RhaT (NM_102790); At3GlcT (NM_121711); Vv3GlcT (AF000371); Ph3GlcT (AB027454); Pf3GlcT (AB002818); Hv3GlcT (X15694); Zm3GlcT (X13501); At5GlcT (NM_117485); Pf5GlcT (AB013596); Vh5GlcT (BAA36423); Ph5GlcT (AB027455); Db7GlcT 4′GlcT (CAB56231); Nt7GlcT 3GlcT (AAB36653); Sb7GlcT (BAA83484); At7RhaT (AY093133); CmF7G2″RhaT (AAL06646); CsF7G6"RhaT (ABA18631); IpA3G2"GlcT (AB192315); PhA3G6"RhaT (X71059); BpA3G2″GlcAT (AB190262); At3G2″XylT (Q9FN26). GlcT, glucosyltransferase; RhaT, rhamnosyltransferase; XylT, xylosyltransferase; GlcAT, glucuronosyltransferase. Ac, *Actinidia chinensis*; At, *Arabidopsis thaliana*; Bp, *Bellis perennis*; Cm, *Citrus maxima*; Cs, *Citrus sinensis*; Db, *Dorotheanthus bellidiformis*; Gm, *Glycine max*; Hv, *Hordeum vulgare*; Ip, *Ipomoea purpurea*; Nt, *Nicotiana tabacum*; Pf, *Perilla frutescens*; Ph, *Petunia hybrida*; Sb, *Scutellaria baicalensis*; Vh, *Verbena hybrida*; Vv, *Vitis vinifera*; Zm, *Zea mays*.

SUMMARY

In this study, two flavonol glycoside glycosyltransferase genes, designated as *GmF3G6"Rt* and *GmF3G6"Gt* in charge of transferring rhamnose or glucose to the 6"-position of the glycoside attached to the 3-*O*-kaempferol respectively, were for first time cloned from plants. The genes may correspond to *Fg2* and *Fg1* reported by Buttery and Buzzel, (1975) and were segregated from a population of 94 recombinant F_6 inbred lines from the soybean cultivars crossings between Kitakomachi and Koganejiro; and Harosoy and Nezumisaya, which showed differences in their flavonol glycosides types and HPLC chromatogram patterns. A genetic analysis and SSR linkage mapping allowed the identification of candidate genes Glyma10g33790 and Glyma20g33810 and were cloned and sequenced. *GmF3G6"Rt* was 1392 bp long enconding 464 aminoacids with no introns, the recombinant protein of *GmF3G6"Rt-a* from Koganejiro was able to catalyze UDP-rhamnose+Kaempferol-3-*O*-glucoside to kaempferol-3-*O*-rutinose. *GmF3G6"Gt* was 1389 bp long encoding 462 amino acids and was 88 % identical to *GmF3G6"Rt,* sharing the same branch in the phylogenic tree. Near-isogenic lines derived from *GmF3G6"Rt* did not show distinct differences in their photosynthesis rate and chlorophyll content, but further analysis are required. New flavonol glycosides were indentified suggesting the existence of a new FGG gene *Fg5* in charge of transferring rhamnose to the 4"-position of glycoside attached to the 3-*O*-Kaempferol position, a new scheme for flavonol glycoside pathway control was proposed.

ACKNOWLEDGEMENTS

First, I want to give thanks to my God for giving me life and the chance to come true this wonderful dream; my gratitude to my academic advisor and sensei Dr. Ryoji Takahashi for believe in me, wait for me for one year while I was in Colombia, and show me day by day that the most important in life are the familiar values and the constant work; I'm thankful with Dr. Eduardo Rodriguez Benitez, for guiding me during the most difficult moments; special thanks to my lab mate Tito O. Rodriguez, for helping me in my personal and spiritual grow, and the experiment as well; and the laboratory members Mr. Di Shaokan, Ms. Yan Fan, Ms. Auchithya Dissanayaka, Mrs. Eiko Izumi and Mrs. Miyako Ogawa, for their help in greenhouse, field, laboratory, and for the nice coffee breaks and great humor; I want also to thanks to Dr. Setsuko Komatsu and her group for the sequencing machine, her advices, guide, and kindness; to Dr. Yoshinori Murai and Dr. Tsukasa Iwashina at the Tsukuba Botanical Garden for their friendship, help and wonderful job identifying the flavonoid samples; to Dr. Yokenura-Sakakibara and her group at RIKEN for their good job at gene expression; to Dr. Tomiya Maekawa at National Agricultural Research Center, for providing and teaching us how to use the photosynthesis machine; to the referees of this thesis Professor Ryo Ohsawa, Professor Chikako Kiribuchi-Utobe and Professor Junichi Tanaka, for their advice and corrections

that help to enrich this writing; and the people of the first floor of NIAS, Dr. Kyoko Toda, for her help with the HPLC machine and advices, Dr. Susumu Hiraga, for his technical advices and friendship; Dr. Makoto Togo, for the printer setup and friendship; And the five most beautiful ladies of the building Kawai-san, Hamano-san, Nakaizumi-san, Norimatsu-san and Kiba-san for their great humor and friendship. Also I want to thanks to Dr. Takahashi's family for their kindness and friendship during all the doctorate time, and their help and special care during the Fukushima nuclear emergency; to my family Tulia, Anasol and Santiago for supporting me every day and full me with love and energy, and my parents in Colombia my country, for their prays and unconditional love. Finally I want to thanks the MEXT scholarship without this work could not be possible, and the country of Japan and its people for be a wonderful host for my family and me.

"¡No solo no hubiéramos sido nada sin ustedes, sino con toda la gente que estuvo a nuestro alrededor desde el comienzo; algunos siguen hasta hoy! ¡Gracias totales!" (Gustavo Cerati)

REFERENCES

Agati G., Azzarello E., Pollastri S. and Tattini M. (2012) Flavonoids as antioxidants in plants: localization and functional significance. Plant Sci 196:67-76

Aherne A.S. and O`Brien N.M., (2002) Dietary flavonols: chemistry, food content and metabolism. Nutr 18:75-81

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

Andersen O.M. and Markham K. R. (2007) Flavonoids, chemistry, biochemistry, and applications. p. 219-263; 319-397, Taylor and Francis Group, USA.

Antunes P., Rajcan I. and Goss M. (2006) Specific flavonoids as interconnecting signals in the tripartite symbiosis formed by arbuscular mycorrhizal fungi, *Bradyrhizobium japonicum* (Kirchner) Jordan and soybean (*Glycine max* (L) Merr.). Soil Biol and Bioch 38(3):533-543

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

Bhagwat S.B., Haytowitz D.B. and Holden J.M. (2011) USDA database for the flavonoid content of selected foods. p.16-20, U.S Department of Agriculture, Baltimore, USA.

Bin L., Xiaohua H., Guangsheng Z., Feng Z. and Qing Z. (2006) effect of Lanthanum on plants under supplementary ultraviolet-B radiation: effect of lanthanum on flavonoid contents in soybean seedlings exposed to supplementary ultraviolet-B radiation. J of rare earths 24(5):613-616

Bradley M.J., Davies K. M., Delores C.S., Bloor J. S. and Lewis D. (1998) The maize Lc regulatory gene up-regulates the flavonoid biosynthetic pathway in *Petunia*. The Plant J 13(3):381-392

Brochado C.O., Almeida A. P., Barreto B. P., Costa L.P., Ribeiro L.S., Pereire R. L. C., Goncalves Koatz V.L. and Costa S. S. (2003) Flavonol robinobiosides and rutinosides from *Alternanthera brasiliana* (Amaranthaceae) and their effect on lymphocytes proliferation *in vitro*. J Braz Chem Soc 14(3):449-451

Buttery B.R. and 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. Can J Bot 53:219-224

Buttery B.R. and Buzzell R.I. (1976) Flavonol glycoside genes and photosynthesis in soybean. Crop Sci 16: 547-550

Buzzell R.I. (1974) Soybean linkage tests. Soybean Genet Newsl 1:11-14

Buzzell R.I. and Buttery B.R. (1974) Flavonol glycoside genes in soybeans. Can J Genet Cytol 16:897-899

Buzzell R.I and Buttery B.R. (1992) Inheritance of an anomalous flavonol glycoside gene in soybean. Genome 35:636-638

Buzzell R.I and Buttery B.R. (1998) Genetics of leaf waviness in soybean. Soybean Genet Newsl 25:23-24

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

Chen Y., Zheng R., Jia Z. and Ju Y. (1990) Flavonoids as superoxide scavengers and antioxidants. Free Rad Biol and Medic 9:19-21

Chen A. Y. and Chen C.Y. (2013) A review of dietary flavonoid, kaempferol on human health and cancer chemoprevention. Food chem 138:2099-2107

Chon S.U. (2013) Total polyphenols and bioactivity of seeds and sprouts in several legumes. Curr Pharm Des 19 (34):6112-6124

Cook, N.C., and Samman, S. (1996) Flavonoids-Chemistry, metabolism, cadioprotective effects and dietary sources. Nutri Bioch 7:66-76

Cosio E.G. and McClure J.W. (1984) Kaempferol glycosides and enzymes of flavonol biosynthesis in leaves of a soybean strain with low photosynthetic rates. Plant Physiol 74:877-881

Cregan P.B., Jarvic T., Bush A.L., Shoemaker R.C., Lark K.G., Kahler A.L., Kaya N., VanToai T.T., Lohnes D.G., Chung J. and Specht J.E. (1999) An integrated genetic map of the soybean genome. Crop Sci 39:1464-1490

Crozier A., Lean M.E.J., McDonald M.S. and Black C. (1997) Quantitative analysis of the flavonoid content of commercial tomatoes, onions, lettuce and celery. J Agric Food Chem 45:590

Dixon R.A. and Pavia N.L (1995) Stress-induced phenylpropanoid metabolism. Plant Cell 7:1085-1097

Fehr W.R., Caviness C.E., Burmood D.T. and Pennington J.S. (1971) Stage of development descriptions for soybeans, GLYCINE MAX (L.) MERRILL. Crop Sci 11:929-931

Figueiredo-Gonzalez M., Martinez-Carballo E., Cancho-Grande B., Santiago J.L., Martinez M.C. and Simal-Gandara J. (2012) Patter recognition of three *Vitis vinifera* L. red grapes varieties based on anthocyanin and flavonol profiles, with correlations between their biosynthesis pathways. Food Chem 130:9-19

Forkman G. and Martens S. (2001) Metabolic engineering and applications of flavonoids. Curr opinion in biotech 12:155-160

Forkmann G. and Heller W. (1999) Biosynthetic pathway of flavonoids. In "Comprehensive Natural Products Chemistry". Eds Barton and Nakanishi, p.713-748. Elsevier Science Ltd., Amsterdam-Netherlands.

Fossen T., Pederson A.T. and Anderson O.M. (1998) Flavonoids from red onion (*Allium cepa*). Phytochem 47:281-285

Frydman A., Liberman R., Huhman D.V., Carnmeli-Weissberg M., Sapir-Mir M., Ophir R., Sumner L.W. and Eyal Y. (2013) The molecular and enzymatic basis of bitter/non-biter flavor of citrus fruit: evolution of branch-forming rhamnosyltransferases under domestication. Plant J 73:166-178

Gachon C.M., Langlois-Meurinne M. and Saindrenan P. (2005) Plant secondary metabolism glycosyltransferases: the emerging functional analysis. Trends Plant Sci. 10: 542–549.

Garcia-Tirado J., Rieger-Reyes C. and Saz-Peiro P. (2012) Effect of flavonoids in the prevention of lung cancer: systematic review (in Spanish). Med Clin (Barc) 139(8):358-363

Githiri S.M., Yang D., Khan N.A., Xu D., Komatsuda T. and Takahashi R. (2007) QTL analysis of low temperature-induced browning in soybean seed coats. J Hered 98:360-366

Harborne, J.B. (1986) Nature, distribution and function of plant flavonoids. In "Plant Flavonoids in Biology and Medicine: Biochemical Pharmacological, and Structure-Activity Relationships" p. 15-24, Alan R. Liss, New York, NY USA.

Hernandez I. and Breusegem F.V. (2010) Opinion on the possible role of flavonoids as energy escape valves: novel tools for natures's Swiss army knife? Plant Sci 179:297-301

Hertog M.G.L., Hollman P.C.H., Katan M.B. and Kromhout D., (1993) Intake of potentially anticarcinogenic flavonoids and their determinants in adults in the Netherlands. Nutrit and Canc 20(1):21-29

Groteworld E. and Peterson T. (1994) Isolation and characterization of maize gene encoding chalcone flavones isomerase. Mol Gen Genet 242(1):1-8

Hisano H., Sato S., Isobe S., Sasamoto S., Wada T., Matsuno A., Fujishiro T., Yamada M., Nakayama S., Nakamura Y., Watanabe S., Harada K. and Tabata S. (2007) Characterization of the soybean genome using EST-derived microsatellite markers. DNA Res 14:271-281

Hollman P.C.H. and Katan M.B. (1999) Dietary flavonoids: intake, health effects and bioavailability. Food and Chem Toxicol. 37:937-942

Hutzler P., Fischbach R., Heller W., Jungblut T.P., Reuber S., Schmitz R., Veit M., Weissenböck G. and Schnitzler J. P. (1998) Tissue localization of phenolic compounds in plants by confocal laser scanning microscopy. J of Exp Bot 49(323):953-965

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

Iwashina T., Kitajima J. and Matsumoto S. (2006-b) Flavonoids in the species of Cyrtomium (Dryopteridaceae) and related genera. Biochem Syst Ecol 34:14-24

Iwashina T., Githiri S.M., Benitez R.E., Takemura T., Kitajima J. and 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

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

Iwashina T., Yamaguchi M., Nakayama M., Onozaki T., Yoshida H., Kawanobu S., Ono H. and Okamura M. (2010) Kaempferol glycosides in the flowers of carnation and their contribution to the creamy white flower color. Nat Prod Commun 5:1903-1906

Iwata H. and Ninomiya S. (2006) AntMap: constructing genetic linkage maps using an ant colony optimization algorithm. Breed Sci 56:371–377

Jooyandeh H. (2011) Soy products as healthy and functional foods. Middle-East J. of Scient Res 7(1):71-80

Kim T.H., Ku S.K. and Bae J. S. (2012) Inhibitory effects of kaempferol-3-O-sophoroside on HMGB1-mediated proinflammatory responses. Food Chem Toxicol. 50(3-4):1118-1123

Koes R. E., Quattrocchio, R. and Mol, J. N. M. (1994) The flavonoid biosynthetic

pathway in plants: function and evolution. BioEssays 16:123-132

Koes R.E., Spelt C.E. and Mol J.N.M. (1989) The chalcone synthase multigene family of *Petunia hibrida* (V30): differential light-regulated expression during flower development and UV light induction. Plant Mol. Biol. 12:213-225

Konar N. (2013) Determination of phytostrogenic compounds of soybean sprouts grown in Antalya, Turkey. Int J. of Chem Env and Biol Sci (online)

Kostyn K., Szatkowski M., Kulma A., Kosieradzka I. and Szopa J. (2013) Transgenic potato plants with overexpression of dihydroflavonol reductase can serve as efficient nutrition source. Agri. Food. Chem. 61:6743-6753

Kovinich N., Saleem A., Arnason J.T. and Miki B. (2010) Functional characterization of UDP-glucose:flavonoid 3-*O*-glucosyltransferase from seed coat of black soybean (*Glycine max* (L.) Merr.). Phytochemistry 71:1253-1263

Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E. and Newburg L. (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage map of experimental and natural populations. Genomics 1:174-181.

Lin P.Y. and Lai H.M (2006) Bioactive compounds in legumes and their germinated products. J. Agric Food Chem 54(11):3807-3814

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

Luo H., Jianq B.H., King S.M. and Chen Y.C. (2008) Inhibition of cell growth and VEGF expression in ovarian cancer cell by flavonoids. Nutric Cancer 60(6)800-809

Mackenzie P.I., Owens I.S., Burchell B., Bock K.W., Bairoch A., Belanger A., Fournel-Gigleux S., Green M., Hum D.W., Iyanagi T., Lancet D., Louisot P., Magdalou J., Chowdhury J.R., Ritter J.K., Schachter H., Tephly T.R., Tipton K.F. and Nebert D.W. (1997) The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. Pharmacogen 7:255–269

Manach C., Scalbert A., Morand C., Remesy C. and Jimenez L. (2004) Polyphenols: food sources and bioavailability. Am J Clin Nutr 79:727-747

Mansoor A., Anwar-UI-Hassan G., Khalid A. and Viqar U A. (2006) Effect of Kaempferol-3-O-rutinoside on rat blood pressure. Phytothe Reser 7(4):314-316

Marton M., Csapo-Kiss Zs., Mandoki Zs. and Csapo J. (2010) The role of sprouts in human nutrition. Acta Univ. Sapientatiae Alim 3:81-117

Masada S., Terasawa K. and 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 581:2605-2610

van der Meer I.M, Stam M.E, van Tunen A.J, Mol J.N.M. and Stuitje A.R. (1992) Antisense inhibition of flavonoid biosynthesis in petunia anthers results in male sterility. Plant Cell 4: 253-262

Middleton E. M. and Teramura A.H. (1993) The role of flavonol glycosides and carotenoids in protecting soybean from Ultraviolet-B damage. Plant Physiol 103:741-752

Muir S.R., Collins G.J., Hughes S., Bovy A., Ric De Vos C. H., van Tunen A. J. and Verhoeyen M.E. (2001) Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased level of flavonols. Nat Biotechnol 19(5):470-474

Murai Y., Takahashi R., Rojas Rodas F., Kitajima J. and Iwashina T. (2013) New flavonol triglycosides from the leaves of soybean cultivars. Nat Prod Commun (in press)

Murray M.G. and Thompson W.F. (1980) Rapid isolation of high molecular weight plant DNA. Nucl Ac Res 8:4321-4325

Nagamatsu A., Masuta C., Matsuura H., Kitamura K., Abe J. and Kanazawa A. (2009) Down-regulation of flavonoid 3'-hydroxylase gene expression by virus-induced gene silencing in soybean reveals the presence of a threshold mRNA level associated with

pigmentation in pubescence. J of Plant Physio 16:32-39

Ono E., Homma Y., Horikawa M., Kunikae-Doi S., Imai H., Takahashi S., Kawai Y., Ishiguro M., Fukui Y. and Nakayama T. (2010) Functional differentiation of the glycosyltransferases that contribute to the chemical diversity of bioactive flavonol glycosides in grapevines (*Vitis vinifera*). Plant Cell 22:2856-2871

Paquette S., Møller B.L. and Bak S. (2003) On the origin of family 1 plant glycosyltransferases. Phytochem 62:399-413

Peterson, J. and Dwyer J. (1998) Flavonoids: Dietary occurrence and biochemical activity. Nutri Reser 18(12):1995-2018

Rice-Evans C., Miller J. N. and Paganga G. (1997) Antioxidant properties of phenolic compounds. Tren in Plant Sci 2(4):151-159

Rode M.W. and Bernard R.L. (1975) Inheritance of wavy leaf. Soybean Genet Newsl 2:42-44

Sato D., Akashi H., Sugimoto M., Tomita M. and Soga T (2013) Metabolomic profiling of the response of susceptible and resistant soybean strains to foxglove aphid, *Aulacorthum solani* Kaltenbach. J of Chromat 925(15)95-113

Schijlen G.W.M.E., Ric de Vos C.H., Tunen A.J.V. and Bovy G.A. (2004) Modification of flavonoid biosynthesis in crop plants. Phytochemistry 65:2631-2648

Song Q.J., Marek L.F., Shoemaker R.C., Lark K.G., Concibido V.C., Delannay X., Specht J.E. and Cregan P.B. (2004) A new integrated genetic linkage map of the soybean. Theor Appl Genet 109:122-128

Stewart A. J., Bozonnet S., Mullen W., Jerkins G. I., Lean M. E. J. and Crozier A. (2000) Occurrence of flavonols in tomatoes and tomato-based products. J. Agric. Food. Chem. 48:2663-2669

Stoclet J.-C. and Schini-Kerth V. (2011) Dietary flavonoids and human health (in French). Ann Pharmac Franç 69:78-90

Takahashi R. and Asuma S. (1996) Association of T gene with chilling tolerance in soybean Crop Sci 37:1755-1759

Takahashi R., Benitez R. E., Funatsuki H. and Ohnishi S. (2005) Soybean maturity and pubescence color genes improve chilling tolerance. Crop Sci 45:1387-1393

Van der Meer I.M., Stam M., van Tunen A.J., Mol J.M. and Stuitje A.R. (1992) Antisense inhibition of flavonoid biosynthesis in *Petunia* anthers results in male sterility. Plant Cell 4:253-262

Venketeshwer, R. (2012) Phytochemicals a global perspective of their role in nutrition and Health, p. 33-35 Intech, Croatia.

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

Wollenweber E. and Dietz V. (1981) Occurrence and distribution of free flavonoid aglycones in plants. Phytochem 20(5): 869-932

Yan J., Wang b., Jianq Y., Cheng L. and Wu T. (2013) GmFNSII-controlled soybean flavones metabolism responds to abiotic stresses and regulates plant salt tolerance. Plant cell physiology (in press, November 4, 2013)

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

Yonekura-Sakakibara K., Tohge T., Niida R. and Saito K. (2007) Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in Arabidopsis by transcriptome coexpression analysis and reverse genetics. J Biol Chem 282:14932-14941

Zanatta L., Rosso A., Folador P., Figueiredo M.S., Pizzolatti M.G.,, Leitte L.D. and Silva F.R. (2008) Insulinomimetic effect of kaempferol 3-neohesperidoside in the rat soleus muscle. J Nat Prod 71(4):532-535

Zeng Z.B. (1993) Theoretical basis for separation of multiple linked gene effects in

mapping quantitative trait loci. Proc. Natl. Acad. Sci. 90:10972-10976

Appendix

Protocols and reagents

1. Soybean DNA extraction by CTAB method:

- Freeze leaf samples with liquid nitrogen (store at -80ºC if necesary)
- Add 10ml 1.5 × CTAB into screw-capped tubes
- Pre-chill mortars and pestles, and pre-heat tubes in a water bath at 55ºC
- Grind leaf samples into a fine powder with liquid nitrogen in a pre-chilled mortar. (Never thaw the samples DNA will be broken!)
- Transfer leaf powder with pre-chilled spatula into pre-warmed 1.5×CTAB and briefly vortex
- Keep 55ºC for 30 min
- Add 10ml Chloroform /Isoamyl alcohol (24:1)
- Shake for 20 min, cfg 3000 rpm for 20 min at room temperature
- Transfer the upper phase to new tubes
- Add 1 ml 10% CTAB, mix
- Add 10 ml Choroform/Isoamyl alcohol (24:1)
- Shake for 20 min, cfg 3000 rpm for 20 min at room temperature
- Transfer the upper phase to new tubes
- Gently add 15 ml Precipitation Buffer
- Very gently invert to precipitate DNA
- Stand at room temperature for 1.5 hr or over night
- Cfg 3000 rpm for 20 min at room temperature
- Decant supernatant and leave tubes on tissue papers
- Add 5 ml 1M NaCl and 5µl of Rnase solution
- Mix well and keep at 55ºC until pellet is dissolved (3 hr to overnight)
- Add 10 ml of cold (-20ºC) EtOH and mix well
- Keep at 4ºC for 1 hr to overnight
- Centrifuge at 3000 rpm for 20 min at room temperature
- Discard supernatant and add 10 ml of 70% EtOH
- Centrifuge at 3000 rpm for 10 min at room temperature
- Discard supernatant and dry pellets
- Dissolved with 500µl TE buffer and transfer to 1.5 ml tube
- Check the amount and integrity of DNA by agarose gel electrophoresis

2. Reagents for CTAB method of DNA extraction:

a. (1.5) CTAB Solution (500ml)

Reagent	Quantity
CTAB	7.5 g
1M Tris-HCL (pH8)	37.5 ml
0.5M EDTA (pH8)	15.0 ml
NaCl	30.7 g
Polyvinylpyrrolidone (PVP)	3.8 g

Dissolve with about 400 ml of destilled water; adjust volume to 500 ml and autoclave
Store at room temperatures

b. Chloroform/Isoamyl alcohol (24:1)

Reagent	Quantity
Chloroform	500 ml
Isoamyl alcohol	20.8 ml

Mix well

c. 10% CTBA (100 ml)

Reagent	Quantity
CTAB	10 g
5M NaCl	14 ml

Completely dissolve with about 80 ml of water distilled water, adjust volume to 100 ml and autoclave

d. Precipitation buffer (500 ml)

Reagent	Quantity
CTAB	5.0 g
1M Tris- HCl (pH8)	25.0 ml
0.5M EDTA (pH8)	10.0 ml

Fill up to 500 ml and autoclave
Store at room temperatures

3. PCR by Extaq

Reaction solution was prepared as below:

Contents	Volume (µl)	Final concentration
Stillized Milli-Q water	17.5	
10× ExTaq Buffer	2.5	1×
dNTP Mix (2.5mM Each)	2	200µM
Forward Primer (10µM)	1	400nM
Reverce Primer (10µM)	1	400nM
Template DNA*	1	
ExTaq Enzyme (5U/µl)	0.125	7.25U/100µl
Total volume	25µl	

*10 to 100 diluted genomic DNA by CATB method or not diluted single strand cDNA samples synthesized by Superscript III

Sequencing thermal program:

Step	Temperature (ºC)	Time	
Denaturation	94	30 sec	
Denaturation	94	30 sec	30 Cycles
Annealing	56*	30 sec	
Extention	72	1min	
Final Extention	72	5 min	
Store	4	—	

*Adjust for primers

4. Ligation (TA- Cloning)

a. Mix these contents in 1.5 ml microtube:

Reagent	Quantity (µl)
PCR Product	2
PCR2.1 linearized vector (2.5ng/µl)*	1
DNA Ligation kit <Mighty Mix>**	3
Total	6

*invitrogen TA Cloning kit (K2020-20) 25ng/µl, 10 times dilution
**Takara (TKR6023)

b. Incubate at 16ºC, overnight
c. Mix 25µl of competent cell (*E.coli*), 20 min on ice
d. Heat shock at 42ºC in 20 sec, using water bath
e. Add 100µl of SOC medium and incubate at 37ºC for 1hr
f. Split to LAG plate
g. Incubate at 37ºC for 12 to 18 hr
h. Select and culture single white colonies in 3ml TB media

5. Transformation: mini scale plamid preparation (Mini-prep by Alkali-SDS):

- Culture transformed single colonies in 3ml of TB medium at 37ºC for over night
- Centrifuge at 3000 rpm for 5 min
- Remove supernatant (this supernatant is a bio-hazard waste!!)
- Re-suspend the cell pellet with 100µl of SOLUTION-I and transfer into a 1.5ml micro-tube
- Add 200µl of SOLUTION-II and mix gently by up-side-down (about 5 times)
- Place on ice for min
- Add 150µl of SOLUTION-III
- Mix by upside down turning, leave standing for up to 5 minutes maximum
- Add 10µl of chloroform
- Mix by vortex
- Centrifuge at 15,000 rpm for 10 min
- Transfer supernatant (about 450µl) into a new 1.5ml micro-tube
- Add 450µl of isopropanol and mix
- Stand on room temperature for 10 min
- Centrifuge at 15,000 rpm for 5 min
- Remove supernatant and add 500µl of 70% ethanol

- Centrifuge at 15,000 rpm for 1 min
- Remove supernatant and dry the pellet for 20 to 60 minutes
- Dissolve the pellet with 50µl of TE buffer in which 0.05µl of RNase solution was added

Reagents

SOLUTION-I [50mM glucose, 25mM Tris-HCL, 10mM EDTA]
Mix0.9of glucose,2.5ml of 1M Tris-HCL (pH8.0) and 2ml of 500mM EDTA (pH8.0) Adjust volume to 100 ml by Milli-Q water

SOLUTION-II [0.2N NaOH, 10%SDS] (Always freshly constituted)
Mix 860µl of Milli-Q water and add 5N NaOH and 10%SDS, mix

SOLUTION -III [3M K^+, 5M CH_2COO^- solution] Mix 29.4g of Potassium Acetate and 11.5 ml of acetic acid, adjust volume to 100 ml by Milli-Q water

Chloroform, Isopropanol, 70% Ethanol, TE buffer, RNase solution (10 mg/ml, nippon-gene)

6. Determination of vector quality (for presence of an insert):

Mix the following:

Solution	Volume
10×H buffer	1.0µl
Milli-Q water	8.0µl
EcoR1 enzyme	0.1µl
Pasmid DNA sample	1.0µl
Total volume	10.1µl

Run electrophoresis by agarose gel using labda hind III digest as the marker

7. DNA Sequencing by Big-Dye Terminator (ver 1.1 or 3.1)

- Mix the contents below

Contents	Volume (µl)
Distilled water	3
5Xsequencing buffe*	2
Primer (0.8µM)	4
BigDye Reaction Ready*	1
Total	10

*****In Kit (ABI 4337450 (vl.l) or 4337355 (v3.l))

- Add 0.2µl of extracted plasmid vector by Alkali-SDS method
- Thermal program for sequencing reaction:

Step	Temperature (ºC)	Time	
Denaturation	96	5 min	
Denaturation	96	30 sec	25 Cycles
Annealing	50	5 sec	
Extention	60	1min	
Store	4	—	

- Add 1µl each of 3M Sodium acetate and 125mM (pH8.0) solutions and mix
- Add 25µl of ethanol mix stand on ice for 10 min
- Centrifuge 3000rpm 10 min
- Remove supernatant
- Add 30µl of 70% ethanol
- Centrifuge 3000rpm 5 min
- Remove supernatant and dry the DNA pellet in the air
- Add 20µl of Hi-Di formamide solution
- Heat at 95ºC for 5 min and immediately cool on ice
- Run capillary electrophoresis and analyzed by AB13100 Avant Genetic Analyzer

8. Media preparations

a. LB media

Reagent	Quantity
Tryptone	10 g
Yeast Extract	5 g
NaCl	10 g

Adjust to pH 7.0 with 1N NaOH
Dilute to 1000 ml with distilled water

b. LAG plate

100 ml of un-autoclave liquid LB (Luria Bertani) in flask
Add Bacto-Agar 1.5g
Autoclave and cool about 60ºC
Add 1/1000 (v/v) of 50mg/ml ampicillim sodium and 1/1000 (v/v) 2% X-gal
Aliquot to 90mm plates and seal with parafilm

c. Terrific Broth (TB)

A		B	
KH2PO4	2.31 g	Tryptone	12 g
K2HPO4	12.54 g	Yeast Extract	24 g
		Glycerol	4 ml
Adjust 100 ml		Adjust 900 ml	

Autoclave A and B and cool to under 60ºC and mix
Add 1/1000 (v/v) of 50mg/ml ampicillin sodium (optional)
A and B should be constituted separately and mixed later

d. 6x Loading Dye (BPB 1/5 Ver.):

Reagent	Quantity
Bromophenol Blue (BPB)	5 mg
Xylene Cyanol (XC)	25 mg
Ficoll 400	2.5 g
MilliQ-water	7 ml

Adjust volume to 10 millilitter and autoclave

e. 10x TBE

Reagent	Quantity
Tris	108 g
Boric acid	55 g
0.5M EDTA (pH8.0)	40 ml
Distilled water	900 ml

Adjust volume to 1 liter and autoclave

f. Making Gels (two 8% gels)

Reagent	Small	Large
MilliQ- Water	35 ml	42 ml
10 x TBE	5 ml	6 ml
40% acrylamide (39:1)	10 ml	12 ml
TEMED	25-30 µl	40µl
30% APS	150-170 µl	250 µl
Total	50 ml	60 ml

9. Real-time PCR by Applied Biosystems StepOnePlus

a. Extract total RNA using Trizol reagent
b. Synthesize cDNA using Superscript III First-Strand Synthesis System
c. Desing matrix for 96 - well plate
d. Make serial dilution of sample for standar curve (1/5, 1/10, 1/20, 1/40, 1/80,1/160)
4µl cDNA + 16µl autoclaved MQ water, mix
10 µl diluted cDNA + 10µl MQ water, mix
Repeat up to 1/160
e. Diluted Cdna samples for 20 times (4µl cDNA + 76µl autoclaved MQ water, mix)

f. Make master mix of PCR reaction mix (for 24 saples)

1 Sample		x 25
SYBR Premix Dimer Eraser	10µl	250µl with yellow tip
Forward primer (10 pmol/µl)	0.6µl	15µl for actin or target gene
Reverse primer (10 pmol/µl)	0.6µl	15µl for actin or target gene
Rox	0.4µl	10µl
Autoclaved MQ water	6.4µl	160µl
Total	18µl	450µl

g. Put 2µl of cDNA samples to 96-well plate
h. Aliquot 18µl of PCR reaction mix to 96 well-plate
i. Carefully cover plate with film using plastic apparatus (avoid wrinkle)
j. Briefly centrifuge
k. Apply to real-time PCR machine

10. Design of primers

a. Design primers using Primer3 website at http://frodo.wi.mit.edu/
   Amplification size: 80-200 bp
   Primer size: 17-25 bp
   GC content: 40-60%
   Approximately adjust annealing temperature of primers
   If possible, avoid consecutive G or C
b. Dilute primers to 10 pmol/µl with MQ water

11. RNA Extraction (small scale)

a. Leaf (flower) sample
   grind to fine powder with mortar and pestle under liquid nitrogen
b. Transfer to 2ml tubes containing 1 ml TRIZOL Reagent
c. Vortex
d. Leave for 5 min at room temperature
e. Add 200µl chloroform
f. Vortex for 15 seconds
g. Leave for 3 min at room temperature
h. cfg 15 krpm × 15 min at 4ºC
i. Transfer aqueous phase to new tubes

j. Add 500µl IPOH
k. Vortex
l. Leave for 10 min at room temperature
m. cfg 15 krpm × 10 min at 4ºC
n. Pipette out supernatant
o. Wash pellet with 1ml 75% EtOH
p. Vortex
q. cfg 15 krpm × 5 min at 4ºC
r. Air dry for min
s. Dissolve with 50µl DEPC water by pipetting 5 times
t. 55ºC for 10 min
u. Check integrity by agarose gel electrophoresis
store at -80ºC

12. **RT-PCR**

2µl total RNA
add 1 µl 10 mM dNTP
6 µl DEPC water

Total 10µl

65^{O}C for 5 min
chill on ice
prepare cDNA synthesis mixture (for 2 reactions)

10×RT buffer 4µl
25mM MgCl2 8 µl
0.1M DTT 4µl
RNaseOUT 2µl
Superscript III 2µl

Total 20µl

Add 10µl of cDNA Synthesis mix to RNA/primer mixture
Mix gently
Spin
50ºC for 50 min
85ºC for 5 min

Chill on ice
Spin
Add 1µl RNase H
37ºC for 20 min
Store at -20ºC

13. cDNA cloning

Mix the following components in PCR tube:

17.3µl	Distilled water
1µl	RT-PCR product
2.5µl	10× ExTaq Buffer
2µl	dNTP solution
1µl	Forward primer (10pmol)
1µl	Reverse primer (10pmol)
0.2µl	Extaq

PCR reaction

94ºC for 30 sec
30 cycles of (94ºC for 30 sec, 59ºC for 1 min, 72ºC for 1 min)
72ºC for 7 min
4ºC for ∞

Take 3µl PCR product and check by 0.8% agarose gel electrophoresis
Transfer remaining PCR product to -20ºC immediately
If distinct unique band with expected size appeared, proceed to vector ligation
If additional band(s) appeared in addition to expected band, optimized PCR condition (annealing temperature) or cut out expected band from agarose gel and extract DNA

14. pCOLD cloning

a. Phenol/chloroform extraction of PCR product

PCR product with single band (25µl)
Add sterile distilled water up to 50 µl
Add 50 µl phenol/chloroform
Vortex
Cfg at 5000 rpm × 5 min

Transfer aqueous phase to new tube
Add 5µl 3M NaOAC, mix
Add 125µl EtOH, mix well
Keep at -20 degree for 30 min
Cfg at 15000 rpm × 10 min at 4 degree
Pipette out liquid
Add 1mL 70% EtOH
Cfg at 15000 rpm × 5 min at 4 degree
Pipette out liquid
Air dry for 10 to 15 min
Dissolve with 30µl sterile distilled water

b. Restriction digestion

Take 10µl of above purified DNA
Add 3µl 10 × M Buffer
Add 25µl sterile distilled water
Add 1µl enzyme
Add 1µl Xhol (total of 30µl)
Mix well
Incubate at 37 degree overnight
Add sterile distilled water up to 50µl
Add 5µl 3M NaOAC, mix
Add 125µl EtOH, mix well
Keep at -20 degree for 30 min
Cfg at 15000 rpm × 10 min at 4 degree
Pipette out liquid
Add 1mL 70% EtOH
Cfg at 15000 rpm × 5 min at 4 degree
Pipette out liquid
Air dry for 10 to 15 min
Dissolve with 10µl TE buffer

c. Ligation with pCOLD vector

pCOLD vector digested with enzyme and Xhol 0.5µl (100ng/µl)
Add insert DNA 2µl
Add Might Mix 2.5µl
Incubate at 16 degree for 1 hr to overnight
Transformation

15. Bacterial expression assay

PCR amplification with KOD-Plus enzyme with high fidelity

Genome DNA	1µl
10 × KOD buffer	2.5µl
dNTP mix	2.5µl
$MgSO_4$	2µl
Primer-1	1µl
Primer-2	1µl
KOD polymerase	0.5µl
Distilled water	14.5µl

96ºC 2 min
40 cycles of (94ºC 15 sec, 59ºC 30 sec, 68ºC 1 min)
68ºC 7 min, 4ºC ∞

16. CAPS Analysis

Make master mix for 96 samples and add following component in 96-well PCR plate

Genomic DNA	1µl (92 F2 plants and two parents)
Distilled water	5.1µl
10× ExTaq buffer	1µl
dNTP mix	0.8µl
CAPS primer -fwd	1µl
CAPS primer -rv	1µl
ExTaq	0.1µl
Total	**10µl**

PCR reaction

94ºC 30 sec
30 cycles of (94ºC 30 sec, 59ºC 1 min, 72ºC 1 min)
72ºC 7 min
4ºC ∞

Restriction digestion

Transfer 5 µl of PCR product to new 96 well plate
Make master mix and add:

2.9µl	distilled water
1µl	10×M buffer
1µl	BSA

0.1µl RsaT
Total 10 µl
37ºC overnight
PAGE