Breeding Science 69: 564–572 (2019)
doi:10.1270/jsbbs.19027

Research Paper

Identification and characterization of a major QTL underlying soybean isoflavone malonylglycitin content

Satoshi Watanabe*[1)], Risa Yamada[1)], Hazuki Kanetake[1)], Akito Kaga[2)] and Toyoaki Anai[1)]

[1)] *Faculty of Agriculture, Saga University*, 1 Honjo-machi, Saga, Saga 840-8502, Japan
[2)] *Soybean and Field Crop Applied Genomics Research Unit, Institute of Crop Science, NARO (National Agriculture and Food Research Organization)*, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan

Isoflavones in soybean seeds are responsible for plant–microbe interactions and defend against pathogens, and are also beneficial to human health. We used two biparental populations and mini core collection of soybean germplasm to identify and validate QTLs underlying the content of isoflavone components. We identified a major QTL, *qMGly_11*, which regulates the content of malonylglycitin, on chromosome Gm11, in populations bred from parents with high, low, and null glycitein contents. *qMGly_11* explained 44.5% of phenotypic variance in a population derived from a cross between 'Aokimame' (high) and 'Fukuyutaka' (low) and 79.9% of that in a population between 'Kumaji-1' (null) and 'Fukuyutaka' (low). The effect was observed only in the hypocotyl. We further confirmed the effect of *qMGly_11* in a mini-core collection, where it explained 57.1% of the genetic diversity of glycitin production and 56.5% of malonylglycitin production. *qMGly_11* increased the contents of glycitin and malonylglycitin at the expense of daidzin and malonyldaidzin in all analyzed populations. We discuss the gene responsible for this QTL and the availability of the null allele for metabolic engineering of soybean seed isoflavones.

Key Words: soybean, isoflavone, QTL, malonylglycitin, glycitin.

Introduction

Soybean (*Glycine max* (L.) Merrill) is one of the most important crops for oil production and food for humans and livestock. Soybean seeds contain not only carbohydrates, proteins, and oils, but also secondary metabolites with functional benefits for human health. Among these metabolites, isoflavones function in plant–microbe interactions (Dakora and Phillips 1996, Graham 1991). Soybean accumulates three aglycone forms of isoflavones—genistein, daidzein, and glycitein—which are mostly glycosylated and then malonylated in mature seeds. These aglycones have health benefits, including reducing the risks of cardiovascular disease and osteoporosis, mitigating menopausal symptoms (van de Weijer and Barentsen 2002), and suppressing the proliferation of certain cancers (Yu and McGonigle 2005).

The biosynthesis of isoflavones in legumes is well characterized, and key enzymes and their physical interactions have been elucidated (Yu and McGonigle 2005). The first step in the flavonoid biosynthesis pathway is the synthesis of naringenin chalcone from 4-coumaroyl-CoA and three units of malonyl-CoA by chalcone synthase (CHS, Kreuzaler and Hahlbrock 1975). CHS and chalcone reductase also catalyze isoliquiritigenin synthesis (Welle and Grisebach 1988). Both chalcones (naringenin chalcone and isoliquiritigenin) are converted into isoflavone aglycones (genistein and daidzein) by chalcone isomerase (Kreuzaler and Hahlbrock 1975), isoflavone synthase (IFS, Jung *et al.* 2000, Steele *et al.* 1999), and 2-hydroxyisoflavanone dehydratase, sequentially (Akashi *et al.* 2005). Glycitein aglycone is synthesized from liquiritigenin by flavonoid 6-hydroxylase (F6H, Latunde-Dada *et al.* 2001) and an unidentified isoflavone *O*-methyl transferase. After the synthesis of these aglycones on the surface of the endoplasmic reticulum, glycosyltransferase and malonyltransferase respectively catalyze the synthesis of glycosides (daidzin, Da; glycitin, Gly; genistin, Ge) and malonyl glycosides (malonyldaidzin, M_Da; malonylglycitin, M_Gly; and malonylgenistin, M_Ge, Dhaubhadel *et al.* 2008, Funaki *et al.* 2015). These compounds are transported into the vacuole from the cytoplasm (Dhaubhadel *et al.* 2008). Among these enzymes, the expression profiles of CHS7 and CHS8 during seed development coincide well with the patterns of isoflavone accumulation in soybean seeds, and their gene expression levels differed significantly between high- and low-isoflavone-

Communicated by Donghe Xu
Received February 26, 2019. Accepted June 20, 2019.
First Published Online in J-STAGE on September 5, 2019.
*Corresponding author (e-mail: nabemame@cc.saga-u.ac.jp)

content lines (Dhaubhadel *et al.* 2006). In addition, transient expression of an R1-type MYB gene, GmMYB176 (accession DQ822924), induced endogenous expression of CHS8 (Yi *et al.* 2010). Further, overexpression of some R2R3-MYB transcription factors regulating isoflavone biosynthesis in soybean and *Lotus japonicus* (Chu *et al.* 2017, Kunihiro *et al.* 2017, Shelton *et al.* 2012) induced the up-regulation of genes encoding enzymes related to branched phenylpropanoid and isoflavone biosynthesis pathways. Nevertheless, knowledge of genes associated with the genetic diversity of isoflavone contents in soybean is still limited. Although the association between isoflavone content and SNPs found in *IFS* genes was reported, detailed effects of SNPs were not elucidated (Cheng *et al.* 2008).

Many studies have identified loci controlling the accumulation of isoflavones through quantitative trait locus (QTL) analysis (Akond *et al.* 2014, Cai *et al.* 2018, Gutierrez-Gonzalez *et al.* 2010, 2011, Han *et al.* 2015, Kassem *et al.* 2004, Li *et al.* 2014, Meksem *et al.* 2001, Pei *et al.* 2018, Primomo *et al.* 2005, Wang *et al.* 2015, Yoshikawa *et al.* 2010, Zeng *et al.* 2009); at least 293 QTLs related to isoflavones (61 for daidzein, 69 for genistein, 72 for glycitein, 91 for total isoflavone contents) are listed in SoyBase (April 2019). Isoflavone contents in soybean seeds are also affected by environmental factors, such as temperature during seed maturation, soil, and cultivation conditions (Hoeck *et al.* 2000, Rasolohery *et al.* 2008, Tsukamoto *et al.* 1995). Multiple environmental evaluations with the same population indicated that the heritability of each isoflavone component is 0.6 to >0.9 (Yoshikawa *et al.* 2010). The comparison of F6H expression among several lines with high, low, and null glycitein contents revealed a putative association with accumulation of glycitein (Artigot *et al.* 2013). However, detailed knowledge of the effects of these QTLs is still limited.

We screened a soybean mini core collection (Kaga *et al.* 2012) and our own soybean genetic resources for their seed isoflavone contents and found three lines with similar M_Gly contents as reported previously (Artigot *et al.* 2013): 'Aokimame' (high content), 'Fukuyutaka' (low), and 'Kumaji-1' (null; **Table 1**). We used QTL analysis to find loci controlling seed M_Gly content in a population derived from a cross between 'Aokimame' and 'Fukuyutaka'. We also confirmed the reliability of one of the detected QTLs, *qMGly_11* on chromosome Gm11, which showed the highest effect on M_Gly accumulation in seeds of several different types of populations. We characterized this QTL by measuring the isoflavone contents of the cotyledons and the other seed tissues. *qMGly_11* controlled the accumulation of M_Gly and glycitin instead of daidzin and malonyl-daidzin in specific seed tissues, and can explain the majority of the genetic diversity underlying these isoflavones in soybean germplasms.

Materials and Methods

Plant materials

We used populations derived from a cross between the Chinese soybean (*Glycine max*) landrace 'Aokimame' (AO, JP28298 in Genebank of National Agriculture and Food Research Organization: NARO) and the Japanese breeding cultivar 'Fukuyutaka' (FUK, JP29668, bred at the Kyushu Okinawa Agricultural Research Center, NARO), and from a cross between the Japanese landrace 'Kumaji-1' (KUM, JP28377, collected from Kumamoto Prefecture, Kyusyu) and FUK. We also used a soybean mini core collection (developed by Kaga *et al.* 2012) of 158 lines. Parental lines, the derived populations, and the soybean mini core collection were grown in a field at Saga University (33°14′N 130°17E′) following standard cultivation practices under natural daylength conditions in 2016. The population derived from AO × FUK, with 96 individuals in the F_4 generation obtained by single-seed decent from the F_2 ("AF-F_4"), was used for construction of a genetic linkage map. The AF-F_4 and AF-F_5 populations were grown in 2013 (sown on July 17, harvested in late October to late November) and 2014 (sown on July 22, harvested in late October to mid November). We obtained phenotypic data (see *Measurement of isoflavone contents* below) from $F_{5:6}$ seeds (10 seeds per line) of AF-F_4 and selected one residual heterozygous line (RHL) heterozygous for *qMGly_11* from AF-F_4 (AF_RHL46), and evaluated F_5 seeds of AF_RHL46 (24 seeds for isoflavone extraction from whole seed, 48 seeds for tissue-specific accumulation of isoflavones; see *Measurement of isoflavone contents* below). The population derived from FUK × KUM, with 180 F_2 individuals ("KF-F_2"), was grown in 2017 (sown on July 21, harvested in mid November) and used for phenotypic evaluation with 5 bulked seeds

Table 1. Phenotypes of isoflavone contents of parental lines used in this study

Trait (abbreviation)[a]	Aokimame (AO)[a] Mean ± SD (μg/g)		Fukuyutaka (FUK) Mean ± SD		Kumaji-1 (KUM) Mean ± SD	
Daidzin (Da)	23.2	3.8	27.1	6.5	104.3	10.2
Genistin (Ge)	83.2	13.9	90.8	15.3	223.1	22.0
Glycitin (Gly)	78.4	4.8	35.2	1.2	ND[b]	ND
Malonyl daidzin (M_Da)	747.4	79.3	807.0	124.0	2002.1	180.6
Malonyl genitin (M_Ge)	1313.9	165.3	1524.7	179.6	2335.2	165.3
Malonyl glycitin (M_Gly)	194.9	13.7	73.3	2.4	ND	ND

[a] Abbreviations of line name and isoflavone molecules are shown in parentheses.
[b] ND, not detected: peak areas were below detection level.

of each individual. A single seed obtained from each F_2 individual of KF-F_2 was used for genotyping and measurement of isoflavone contents ("KF-F_3", 96 lines evaluated). The parental lines (FUK, AO, and KUM) and the mini core collection were grown again in 2017 (sown on July 21, harvested in October to November). Seeds were harvested and then held at 4°C until phenotypic evaluation.

DNA extraction

Genomic DNA was extracted from young trifoliate leaves (~100 mg fresh weight) of growing plants or from seed powder (~10 mg) by the CTAB method (Murray and Thompson 1980). Linkage mapping with AF-F_4 and KF-F_2 was conducted with DNA extracted from leaves, and other genotyping experiments were performed with DNA extracted from seed powder.

DNA marker analysis

For construction of the linkage map, we used simple sequence repeat (SSR) markers and designated derived cleaved amplified polymorphic sequence (dCAPS) markers. Each amplicon was digested according to the manufacturer's (Wako, Osaka, Japan) instructions for dCAPS markers, and PCR products of SSR and digested DNAs were separated in a 12% polyacrylamide gel, stained with ethidium bromide, and visualized with a UV transilluminator. Automatic programs developed in a previous study (Watanabe *et al.* 2017) were used for designing primers for dCAPS analysis from SNPs obtained in a previous study from the parental lines (Yamagata *et al.* 2018). We also designed primers for additional DNA markers with the technique of nearest-neighboring-nucleotide substitution–high-resolution melting analysis (NNNs-HRM) to map or evaluate target QTLs as described previously (Yamagata *et al.* 2018). We used a Light Cycler 96 sequencer (Roche Diagnostics K.K., Tokyo, Japan) to detect polymorphism of PCR fragments containing SNPs for NNNs-HRM markers and determined genotype with the manufacturer's software. The primers used for NNNs-HRM markers are listed in **Supplemental Table 1**. The primer sequences and physical positions of the other SSR markers on the soybean genome were taken from SoyBase (https://soybase.org/).

Measurement of isoflavone contents

We extracted isoflavones from whole seeds and from embryo tissues. For the analysis of parents, we crushed freeze-dried whole seeds in a powder mill (5 of AF-F_4 and KF-F_2 and 1 of AF_RHL46). For the analysis of progeny, we removed the seed coat and separated seeds into cotyledons and "hypocotyl" (plumule, epicotyl, hypocotyl, and radicle) with a knife. We analyzed 1 hypocotyl of AF_RHL46 and KF-F_3 and 3 to 5 hypocotyls of the mini core collection for association analysis. We weighed freeze-dried hypocotyls for adjustment of isoflavone contents by their weight, and crushed them in a Multibeads shocker (Yasuikikai, Osaka, Japan) in a 3-mL tube with metal beads. We suspended 100 mg of powder obtained from whole seeds or cotyledons or all powder obtained from crushed hypocotyls in 1 mL of buffer containing 70% ethanol (v/v) and 0.1% acetic acid (v/v) and vortexed it. To enhance extraction efficiency, we sonicated the mixture for 30 min at room temperature with an ultrasonic cleaner (Bransonic 5800, Emerson Japan, Ltd., Kanagawa, Japan). After brief centrifugation, the supernatant was collected in a new tube. These extraction steps were performed three times, and the final extract (close to 3 mL) was filtered through a 0.45-µm membrane. Samples of 3 µL were analyzed by high-pressure liquid chromatography (HPLC; Jasco Corp., Tokyo, Japan) on a Hydrosphere C18 column (YMC Co. Ltd., Kyoto, Japan). Separation was performed with a water-based two-solvent system: solvent A contained 3% acetic acid; solvent B contained 3% acetic acid and 50% acetonitrile. The percentage of solvent B during one 20-min separation at 35°C was programmed as follows: 25% (+75% solvent A) for 1 min, a linear gradient from 25% to 60% for 10 min, 100% for 2 min, and 25% for 7 min. Isoflavones were detected at 254 nm with a UV detector.

External standard curves for daidzein, daidzin, malonyldaidzin, glycitein, glycitin, malonylglycitin, genistein, genistin, and malonylgenistin (all purchased from Wako, Osaka, Japan) were drawn, and peak areas of extracted isoflavones were calculated according to their retention times and areas of the standards.

Statistical analysis

We constructed the linkage map in AntMap software (Iwata and Ninomiya 2006) with default parameters. We identified QTLs by composite interval mapping in R/qtl software (Broman *et al.* 2003) with threshold LOD scores calculated by permutation testing (1000×) at the 5% significance level. We also used the "fitqtl" function to estimate the effect of each QTL. Genotype and phenotype data for AF-F_4 are provided as **Supplemental Text 1**. We used linear regression analysis and analysis of variance (ANOVA) to confirm the effects of QTLs and for association analysis of the genotype of the closest DNA marker and isoflavone contents. We calculated Pearson's correlation coefficient among pairs of traits. All analyses were performed in R software (R Development Core Team 2008).

Results

QTL analysis with AO × FUK (AF) population

The parental lines of AF-F_4 differed significantly in M_Gly (FUK, 73.3 µg/g; AO, 194.9 µg/g; $P < 0.01$) and Gly (FUK, 35.2 µg/g; AO, 78.4 µg/g; $P < 0.01$). Six isoflavones were extracted from whole seeds of the AF-F_4 population: daidzin (Da), genistin (Ge), glycitein (Gly), malonyldaidzin (M_Da), malonylgenistin (M_Ge), and malonylglycitin (M_Gly) (**Supplemental Text 1**). Contents of other isoflavones were below the detection limit (aglycone forms and acetylated forms). We also omitted Gly for QTL analysis of

AF-F_4 population because of the large number of missing data.

Accumulation patterns and ratios of each isoflavone coincided well with previous results (Tsukamoto *et al.* 1995); M_Ge (51.1% of total isoflavone content in AF-F_4, calculated from **Supplemental Text 1**), Ge (8.9%,), Da (3.6%), M_Da (29.4%) accounted for the majority of the content in mature seeds, while M_Gly accounted for only 4.4%. Among these molecules, correlations between malonyl glycosides (M_Da, M_Ge) and glycosides (Da, Ge) were very high (M_Da:Da, $r = 0.94$, $P < 0.01$; M_Ge:Ge, $r = 0.88$, $P < 0.01$, we also calculated these values from **Supplemental Text 1**). The correlation between M_Da and M_Ge ($r = 0.93$, $P < 0.01$) was higher than those between M_Da and M_Gly ($r = 0.23$) and between M_Ge and M_Gly ($r = 0.22$). We used the contents of these malonylated forms as phenotypic data for QTL analysis for their abundance and high correlations with glycosylated forms.

We constructed a linkage map covering the soybean genome with 335 DNA markers (**Supplemental Text 1**) for QTL analysis of malonyl glycosides (M_Da, M_Ge and M_Gly). The total map length was 1771.2 cM and the average marker interval was 5.6 cM. QTL analysis located 2 QTLs for M_Gly, and 1 QTL for M_Ge (**Table 2**). The QTL with the highest LOD score was associated with M_Gly on Gm11 at 37.0 cM (LOD, 7.6; proportion of phenotypic variance explained [PVE] by QTL, 27.9%). We named this QTL *qMGly_11* (QTL for malonylglycitin content located on Gm11); the AO allele increased M_Gly (additive effect, 21.8 μg/g that was calculated by a half of difference between homozygous phenotype of each parental line) in whole seed. Another QTL for M_Gly was detected on Gm15 (LOD, 4.4; PVE, 15.1%); the AO allele also increased M_Gly (additive effect, 16.0 μg/g). Hereafter, we focus on *qMGly_11*.

Evaluation of qMGly_11 by progeny tests with whole seeds and tissue-specific accumulation of glycitein

To confirm the effect of *qMGly_11*, we analyzed the progeny of line AF_RHL46, which was heterozygous for *qMGly_11*. The genotype of DNA marker M23_AF (SNP for Gm11, 8 182 360 bp, 35.0 cM), close to *qMGly_11* (37.0 cM), was significantly associated ($P \leq 0.001$, **Table 3**) with seed contents of M_Gly and Gly. The additive effect of *qMGly_11* in AF_RHL46 was 17.9 μg/g, almost the same as in AF-F_4 (21.8 μg/g). The genetic uniformity of the RHL increased the PVE by *qMGly_11* to 55.8% for M_Gly and 49.6% for Gly. Previous studies with several soybean cultivars (Rasolohery *et al.* 2008, Tsukamoto *et al.* 1995) showed that the pattern of M_Gly accumulation differed between the cotyledon and the "hypocotyl" (plumule, epicotyl, hypocotyl, and radicle). The cultivar 'Jack' showed a similar pattern, in which the embryo axes contained most of the glycitein (Lygin *et al.* 2013). To confirm whether this trend also applied to the phenotypes of the parental lines of AF-F_4 and KF-F_2, we measured the isoflavone contents of cotyledons and hypocotyls of mature seed of the soybean mini core collection: in AO and FUK, the hypocotyls contained all or most of the M_Gly (**Supplemental Table 2**); the average ratio of Gly + M_Gly to total isoflavone content in the hypocotyls was 34.5% (average of mini core collection). We further evaluated whether the effect of *qMGly_11* was associated with the accumulation of M_Gly in hypocotyls in AF_RHL46. Not only M_Gly and Gly contents but also M_Da and Da contents in hypocotyls were highly associated with the genotype of M23_AF (**Table 4**). The contents of Da and M_Da were oppositely associated with the genotype. PVEs by *qMGly_11* were 44.5% for M_Gly, 46.7% for Gly, 50.4% for M_Da, and 47.4% for Da.

Table 2. Summary of QTL analysis with the population derived from the cross between Fukuyutaka and Aokimame

Trait	Chromosome	LOD peak position (cM)	LOD	PVE[a]	Additive effect[b]	QTL name
M_Gly	Gm11	37.0	7.6	27.9	–21.8	*qMGly_11*
	Gm15	9.7	4.4	15.1	–16.0	*qMGly_15*
M_Ge	Gm12	78.1	5.3	19.7	–162.8	*qMGe_12*

[a] PVE, proportion of phenotypic variance explained.
[b] Additive effect (μg/g) of Fukuyutaka allele relative to Aokimame allele.

Table 3. Confirmation of qMGly_11 in progeny of AF_RHL46 (heterozygous for qMGly_11) by the genotype of SNP marker M23_AF

Trait	Homozygous for AO ($n = 4$)[a]		Heterozygous ($n = 13$)		Homozygous for FUK ($n = 7$)		PVE[b]	*F*-value	*P*-value
	Mean ± SD (μg/g)		Mean ± SD		Mean ± SD				
Da	Not significant						NC	0.8	0.445
Ge	Not significant						NC	0.2	0.823
Gly	101.0	34.4	88.1	18.2	46.8	22.2	49.6%	10.3	0.001
M_Da	Not significant						NC	0.4	0.653
M_Ge	Not significant						NC	0.3	0.748
M_Gly	60.4	18.7	51.3	11.8	24.7	10.9	55.8%	13.2	0.000

[a] Numbers of plants with indicated genotypes are shown in parentheses.
[b] PVE, proportion of phenotypic variance explained.

Evaluation of qMGly_11 in population derived from a line with null glycitin content

We screened the soybean mini core collection (Kaga *et al.* 2012) for M_Gly phenotypes and identified KUM as having null Gly or M_Gly (details in *Association analysis in soybean mini core collection* below). We analyzed the segregation of glycitin content in the F_2 population derived from KUM × FUK (KF-F_2). We conducted QTL analysis for M_Gly content with the linkage map of Gm11 constructed from KF-F_2 because *qMGly_11* had a large effect on M_Gly content in AF-F_4. QTL analysis of KF-F_2 using isoflavone contents of whole seeds detected a major QTL at almost the same location as *qMGly_11* in AF-F_4 (**Supplemental Fig. 1A**). The peak LOD score of *qMGly_11* detected in KF-F_2 was 44.3, and its PVE was 71.8% (23.5 cM on Gm11) at SNP marker M12_KF (8.2 Mbp) (**Supplemental Fig. 1B**). Relative to the allele of Fuk allele, the *qMGly_11* allele of KUM had an additive effect of –36.0 μg/g and a dominance effect of 1.8 μg/g for M_Gly (**Table 5**).

We confirmed the effect of *qMGly_11* on isoflavone contents of hypocotyls of 96 individuals in KF-F3. The genotype of M12_KF segregated completely with M_Gly and Gly phenotype (no content when homozygous for KUM allele) in hypocotyls (**Table 6**). *qMGly_11_KUM* explained 79.9% of phenotypic variance for M_Gly and 67.3% for Gly. It also explained 26.2% for M_Da and 20.2% for Da. The allele of KUM of *qMGly_11* reduced M_Gly and Gly contents but increase M_Da and Da contents (**Table 6**).

Association analysis in soybean mini core collection

Next, we performed association analysis of *qMGly_11* in the mini core collection. We measured isoflavone contents in cotyledons and hypocotyls of the 158 accessions of the mini core collection and analyzed the association between the contents and the *qMGly_11* genotype using DNA markers M1_AS to M6_AS (8.12 to 8.17 Mbp, **Supplemental**

Table 4. Association between genotypes of SNP marker M23_AF and contents of isoflavones extracted from hypocotyl in AF_RHL46

Traits	Homozygous for AO (n = 11)[a]		Heterozygous (n = 22)		Homozygous for FUK (n = 13)		PVE[b]	F-value	P-value
	Mean ± SD (μg/mg)		Mean ± SD		Mean ± SD				
Da	0.5	0.3	0.9	0.3	1.4	0.5	47.4%	19.4	0.000
Ge	Not significant						0.0%	0.0	0.995
Gly	3.5	1.1	3.1	0.8	1.6	0.6	46.7%	18.8	0.000
M_Da	2.6	1.0	4.0	1.1	6.0	1.7	50.4%	21.8	0.000
M_Ge	Not significant						1.5%	0.3	0.718
M_Gly	2.4	0.8	2.2	0.5	1.2	0.4	44.5%	17.2	0.000

[a] Numbers of plants with indicated genotypes of M23_AF are shown in parentheses.
[b] PVE, proportion of phenotypic variance explained.

Table 5. Contents of isoflavones in whole seed in KF-F2 by *qMGly_11* allele

Trait	Homozygous for KUM (n = 41)[a]		Heterozygous (n = 85)		Homozygous for FUK (n = 31)		PVE[b]	F-value	P-value
	Mean ± SD (μg/mg)		Mean ± SD		Mean ± SD				
Da	Not significant						NC	1.2	0.305
Ge	Not significant						NC	2.5	0.084
M_Da	Not significant						NC	0.4	0.662
M_Ge	Not significant						NC	2.0	0.133
M_Gly	13.2	7.8	48.8	16.0	81.8	16.9	72.1%	203.0	0.000

[a] Numbers of plants with indicated genotypes of M12_KF are shown in parentheses.
[b] PVE, proportion of phenotypic variance explained.

Table 6. Association between genotypes of SNP marker M12_KF and contents of isoflavones extracted from hypocotyl in KF-F_3

Trait	Homozygous for KUM (n = 34)[a]		Heterozygous (n = 30)		Homozygous for FUK (n = 32)		PVE[b]	F-value	P-value
	Mean ± SD (μg/mg)		Mean ± SD		Mean ± SD				
Da	0.9	0.4	0.6	0.3	0.6	0.2	20.2%	11.8	0.000
Ge	Not significant						NC	1.6	0.204
Gly	0.0	0.0	0.4	0.2	1.0	0.4	67.3%	95.9	0.000
M_Da	13.2	4.3	9.1	2.4	9.3	2.6	26.2%	16.5	0.000
M_Ge	Not significant						NC	1.7	0.182
M_Gly	0.0	0.0	0.9	0.3	1.8	0.6	79.9%	184.6	0.000

[a] Numbers of plants with indicated genotypes of M12_KF are shown in parentheses.
[b] PVE, proportion of phenotypic variance explained.

Table 1), close to M23_AF (8.18 Mbp). M_Gly contents in cotyledons of most lines (88.0%) were below the detection level. KUM and three other lines had no M_Gly or Gly in hypocotyls (Supplemental Table 2). We analyzed correlations among isoflavone components and detected high correlations (0.84 to 0.96) between glycosides and malonyl glycosides (Supplemental Table 3). However, the correlation between M_Da of hypocotyl and that of cotyledon was very low (0.05), as were those of M_Ge and M_Gly between hypocotyl and cotyledon (0.20 and 0.22). We found significant associations of the contents of Da, Gly, M_Da, and M_Gly in hypocotyls with the genotype of marker M4_AS (8.16 Mbp bp): *qMGly_11* explained 19.1% of the phenotypic variance of Da, 57.1% of Gly, 18.2% of M_Da, and 56.5% of M_Gly (Table 7). There was no association with M_Gly and Gly contents of cotyledons, but there was a weak association between contents of Da, M_Da, Ge, and M_Ge in cotyledons (Table 7). To confirm the chromosomal region conferring the high association between M_Gly and Gly contents in hypocotyl, we analyzed associations with markers M58_AF (7.76 Mbp) and M66_AF (9.42 Mbp) also. Neither had a significant association with M_Gly like M4_AS (data not shown). Allelic association (linkage disequilibrium) was absent at these loci, and the gene underlying *qMGly_11* is located between these two markers.

Discussion

We identified a QTL associated with the contents of Gly and M_Gly on chromosome Gm11. Other major QTLs at this locus explaining 40% to 50% of variance were also reported in crosses between 'Essex' and 'Forrest' (Kassem *et al.* 2004, Meksem *et al.* 2001) and between 'Peking' and 'Tamahomare' (Yoshikawa *et al.* 2010). Genome-wide association analysis detected significant association with glycitein content on Gm11 (Chu *et al.* 2017). These QTLs are probably *qMGly_11*, because *qMGly_11* explained most of the genetic diversity of M_Gly and Gly contents in soybean hypocotyl. *qMGly_11* had specific effects relating to the accumulation of glycitein instead of daidzein in hypocotyl. The results of progeny testing (Table 4) and of association study in the mini core collection (Table 7) show that the FUK allele of *qMGly_11* (or G/G genotype of M4_AS) reduces M_Gly and Gly, but increases Da or M_Da. The amounts of both glycosylated and malonylated forms of glycitein were inversely related to those forms of daidzein. The *qMGly_11* FUK allele, however, did not cause any accumulation of glycitein aglycone (data not shown). Hence, *qMGly_11* probably regulates the divergence between daidzein and glycitein. F6H, which regulates the biosynthesis step from liquiritigenin (4′,7-dihydroxyflavanone) to 6,7,4′-trihydroxyflavanone, would play an important role in this divergence. A CYP71D9 (cytochrome P450) with F6H activity was previously identified in soybean (Latunde-Dada *et al.* 2001), and three homologous soybean genes (*F6H1*, *Glyma18g08950*; *F6H2*, *Glyma18g08930*; and *F6H3*, *Glyma08g43890*) were identified by their sequence similarity to the gene encoding CYP71D9 (Artigot *et al.* 2013). Comparison of the expression profiles of these *F6H* genes among tissues and time points during seed development of cultivars with null, low, medium, and high glycitein contents revealed that PI567580A (Chinese landrace Qi Si Mi), with the null phenotype for glycitein, did not express *F6H3* (*Glyma08g43890*), but the other cultivars expressed it similarly. *F6H1* and *F6H2* had no association with glycitein content (Artigot *et al.* 2013).

The gene responsible for the null glycitein in PI567580A is likely to be the same as that in KUM. So further studies, such as expression analysis of *F6H3* with near isogenic lines for *qMGly_11*, will be necessary to uncover the gene underlying this QTL. If we can find differences in the

Table 7. Association of *qMGly_11* with contents of isoflavones extracted from hypocotyl and cotyledon of soybean mini core collection

Trait[a]	SNP genotype at 8161366 (M4_AS) on Gm11						
	G/G (n = 97)[b]	±SD[c]	A/A (n = 57)	±SD	PVE[d]	F-value	P-value
Da_hy	1.7	0.8	1.0	0.7	19.1%	18.3	0.000
Ge_hy	Not significant				NC	1.8	0.169
Gly_hy	2.7	1.1	5.7	1.3	57.1%	103.2	0.000
M_Da_hy	11.4	3.6	8.1	2.9	18.2%	17.3	0.000
M_Ge_hy	Not significant				NC	3.9	0.022
M_Gly_hy	2.5	1.0	5.2	1.3	56.5%	100.6	0.000
Da_co	134.2	81.3	220.3	142.6	13.9%	12.5	0.000
Ge_co	262.5	130.4	364.6	189.6	10.5%	9.0	0.000
Gly_co	Not significant				NC	0.7	0.521
M_Da_co	524.8	289.1	808.7	448.5	13.7%	12.3	0.000
M_Ge_co	805.9	367.6	1077.7	477.8	10.4%	8.9	0.000
M_Gly_co	Not significant				NC	4.6	0.011

[a] "_hy", hypocotyl (μg/mg); "_co", cotyledon (μg/g).
[b] Genotypes of marker M4_AS are indicated in parentheses. Four individuals with undetermined genotypes were omitted from calculations.
[c] SD, standard deviation.
[d] PVE, proportion of phenotypic variance explained by qMGly_11.
[e] NC, not calculated.

expression of *F6H3* between such near isogenic lines, a transcription factor regulating the expression of *F6H3* would become a strong candidate for the product of *qMGly_11*. When we considered the physical position of *qMGly_11*, DNA markers located at 8.12 to 8.17 Mbp (M1_AS to M6_AS) showed a strong association with M_Gly in the mini core collection, while markers at 7.76 and 9.42 Mbp showed a very weak association. In addition, a marker at 8.21 Mbp (M12_KF) co-segregated with the null M_Gly phenotype in the mapping experiment with KF-F_3 (**Table 6**), and some recombinants (within 1 cM) were observed between M12_KF and M13_KF (8.5 Mbp, **Supplemental Table 1**). From these results, the region from 8.0 to 8.5 Mbp is likely to encompass *qMGly_11*. Further studies for positional cloning with a large number of individuals and complementation testing with soybean mutant lines with little or no M_Gly will be necessary to uncover the identity of the multiallelic *qMGly_11* locus.

There was, however, weak associations of M_Da, M_Ge, Da, and Ge in cotyledons with the marker of M4_AS were detected in the mini core collection (**Table 7**). As we used only a simple regression model to estimate the effect of *qMGly_11* in the mini core collection without consideration of population structure or kinship, further study using our data (**Supplemental Table 2**) and information on SNPs across the whole genome of the mini core collection would provide better knowledge of the effects of loci controlling isoflavone accumulation in cotyledons and hypocotyls.

The negative relationship between the accumulation of Gly + M_Gly and of Da + M_Da under the control of *qMGly_11* was restricted to the hypocotyl. Moreover, we found a low correlation of isoflavones between cotyledon and hypocotyl (**Supplemental Table 3**). During soybean seed development, the axis at the early maturity stage (about 40 days after flowering: DAF) contains root meristem, vascular tissue, shoot meristem, and plumule tissue (Le *et al.* 2007). This stage is followed by cell expansion in the cotyledons and the accumulation there of seed proteins and lipids. The accumulation of M_Gly peaks by 40 DAF, but other malonyl glycosides, M_Ge and M_Da, continue to increase after (Kudou *et al.* 1991). These results and our data indicate that the biosynthesis of M_Gly might stop earlier in the hypocotyl than in the cotyledons, and other malonyl glycosides continue to accumulate during late seed development mainly in the cotyledons, and therefore the mechanisms of isoflavone biosynthesis might differ between cotyledons and hypocotyls. Enzymes related to the biosynthesis of isoflavones are considered to be involved in metabolon that perform sequential reactions (Laursen *et al.* 2015). Isoflavone synthase (IFS, a membrane-bound P450 protein) is anchored to the endoplasmic reticulum membrane, and some related cytoplasmic enzymes, such as chalcone isomerase, CHS, and IFS, are physically associated with each other (Waki *et al.* 2016). The affinity of protein–protein interactions and the efficient transfer of substrates between enzymes, however, are affected by the combination of homologous genes (Mameda *et al.* 2018). In addition, *qMGly_11* might affect the components of the metabolon (controlled by dose and type of enzymes) with tissue-dependent manner.

Finding genes responsible for the diversity of isoflavone contents in soybean germplasms would provide not only knowledge about components of the isoflavone metabolon and its regulatory mechanisms, but also alleles for metabolic engineering of isoflavones. One derivative of daidzein, glyceollin, a phytoalexin, is induced by multiple external stimuli in legumes (reviewed by Dakora and Phillips 1996), and transgenic soybean plants in which the accumulation of isoflavones was suppressed by an IFS transgene were damaged severely by pathogens owing to a deficiency of glyceollin synthesis (Lygin *et al.* 2013). Glyceollin also promotes human health in multiple ways (Nwachukwu *et al.* 2013), notably through antioxidant and anti-inflammatory activities (Kim *et al.* 2012), and inhibits the growth of prostate and breast cancer cells in mouse models (Salvo *et al.* 2006). Increasing the flow of substrate molecules into daidzein production by reducing the biosynthesis of glycitein might increase daidzein and glyceollin in soybean seeds. Further study to identify the gene responsible for *qMGly_11* will be necessary. Genetic material with alleles that increase or decrease glycitein will provide new possibilities for improving the nutritional effects soybean in the human diet.

Author Contribution Statement

SW and TA designed the study. SW, RY and HK collect the data. SW and AK contributed to analysis of data.

Acknowledgments

We are grateful to Tokiko Kitajima for technical support. This study was supported by a MEXT KAKENHI grant (JP 17K07604).

Literature Cited

Akashi, T., T. Aoki and S. Ayabe (2005) Molecular and biochemical characterization of 2-hydroxyisoflavanone dehydratase. Involvement of carboxylesterase-like proteins in leguminous isoflavone biosynthesis. Plant Physiol. 137: 882–891.

Akond, M., S. Liu, S.K. Kantartzi, K. Meksem, N. Bellaloui, D.A. Lightfoot, J. Yuan, D. Wang and M.A. Kassem (2014) Quantitative trait loci for seed isoflavone contents in 'MD96-5722' by 'Spencer' recombinant inbred lines of soybean. J. Agric. Food Chem. 62: 1464–1468.

Artigot, M.P., M. Baes, J. Daydé and M. Berger (2013) Expression of flavonoid 6-hydroxylase candidate genes in normal and mutant soybean genotypes for glycitein content. Mol. Biol. Rep. 40: 4361–4369.

Broman, K.W., H. Wu, S. Sen and G.A. Churchill (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889–890.

Cai, Z., Y. Cheng, Z. Ma, X. Liu, Q. Ma, Q. Xia, G. Zhang, Y. Mu and

H. Nian (2018) Fine-mapping of QTLs for individual and total isoflavone content in soybean (*Glycine max* L.) using a high-density genetic map. Theor. Appl. Genet. 131: 555–568.

Cheng, H., O. Yu and D. Yu (2008) Polymorphisms of *IFS1* and *IFS2* gene are associated with isoflavone concentrations in soybean seeds. Plant Sci. 175: 505–512.

Chu, S., J. Wang, Y. Zhu, S. Liu, X. Zhou, H. Zhang, C. Wang, W. Yang, Z. Tian, H. Cheng *et al.* (2017) An R2R3-type MYB transcription factor, GmMYB29, regulates isoflavone biosynthesis in soybean. PLoS Genet. 13: e1006770.

Dakora, F.D. and D.A. Phillips (1996) Diverse functions of isoflavonoids in legumes transcend anti-microbial definitions of phytoalexins. Physiol. Mol. Plant Pathol. 49: 1–20.

Dhaubhadel, S., M. Gijzen, P. Moy and M. Farhangkhoee (2006) Transcriptome analysis reveals a critical role of *CHS7* and *CHS8* genes for isoflavonoid synthesis in soybean seeds. Plant Physiol. 143: 326–338.

Dhaubhadel, S., M. Farhangkhoee and R. Chapman (2008) Identification and characterization of isoflavonoid specific glycosyltransferase and malonyltransferase from soybean seeds. J. Exp. Bot. 59: 981–994.

Funaki, A., T. Waki, A. Noguchi, Y. Kawai, S. Yamashita, S. Takahashi and T. Nakayama (2015) Identification of a Highly Specific Isoflavone 7-*O*-glucosyltransferase in the soybean (*Glycine max* (L.) Merr.). Plant Cell Physiol. 56: 1512–1520.

Graham, T.L. (1991) Flavonoid and isoflavonoid distribution in developing soybean seedling tissues and in seed and root exudates. Plant Physiol. 95: 594–603.

Gutierrez-Gonzalez, J.J., X. Wu, J.D. Gillman, J.-D. Lee, R. Zhong, O. Yu, G. Shannon, M. Ellersieck, H.T. Nguyen and D.A. Sleper (2010) Intricate environment-modulated genetic networks control isoflavone accumulation in soybean seeds. BMC Plant Biol. 10: 105.

Gutierrez-Gonzalez, J.J., T.D. Vuong, R. Zhong, O. Yu, J.-D. Lee, G. Shannon, M. Ellersieck, H.T. Nguyen and D.A. Sleper (2011) Major locus and other novel additive and epistatic loci involved in modulation of isoflavone concentration in soybean seeds. Theor. Appl. Genet. 123: 1375–1385.

Han, Y., W. Teng, Y. Wang, X. Zhao, L. Wu, D. Li and W. Li (2015) Unconditional and conditional QTL underlying the genetic interrelationships between soybean seed isoflavone, and protein or oil contents. Plant Breed. 134: 300–309.

Hoeck, J.A., W.R. Fehr, P.A. Murphy and G.A. Welke (2000) Influence of genotype and environment on isoflavone contents of soybean. Crop Sci. 40: 48–51.

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

Jung, W., O. Yu, S.-M.C. Lau, D.P. O'Keefe, J. Odell, G. Fader and B. McGonigle (2000) Identification and expression of isoflavone synthase, the key enzyme for biosynthesis of isoflavones in legumes. Nat. Biotechnol. 18: 208–212.

Kaga, A., T. Shimizu, S. Watanabe, Y. Tsubokura, Y. Katayose, K. Harada, D.A. Vaughan and N. Tomooka (2012) Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections. Breed. Sci. 61: 566–592.

Kassem, M.A., K. Meksem, M.J. Iqbal, V.N. Njiti, W.J. Banz, T.A. Winters, A. Wood and D.A. Lightfoot (2004) Definition of soybean genomic regions that control seed phytoestrogen amounts. J. Biomed. Biotechnol. 2004: 52–60.

Kim, H.J., J.-S. Lim, W.-K. Kim and J.-S. Kim (2012) Soyabean glycollins: biological effects and relevance to human health. Proc. Nutr. Soc. 71: 166–174.

Kreuzaler, F. and K. Hahlbrock (1975) Enzymic synthesis of an aromatic ring from acetate units. Partial purification and some properties of flavanone synthase from cell-suspension cultures of *Petroselinum hortense*. Eur. J. Biochem. 56: 205–213.

Kunihiro, S., D. Tanabe, Y. Niwa, K. Kitamura, J. Abe and T. Yamada (2017) Isolation and molecular characterization of a *Lotus japonicus R2R3-MYB* subgroup 7 transcription factor gene. Plant Biotechnol. 34: 45–49.

Latunde-Dada, A.O., F. Cabello-Hurtado, N. Czittrich, L. Didierjean, C. Schopfer, N. Hertkorn, D. Werck-Reichhart and J. Ebel (2001) Flavonoid 6-hydroxylase from soybean (*Glycine max* L.), a novel plant P-450 monooxygenase. J. Biol. Chem. 276: 1688–1695.

Laursen, T., B.L. Møller and J.-E. Bassard (2015) Plasticity of specialized metabolism as mediated by dynamic metabolons. Trends Plant Sci. 20: 20–32.

Le, B.H., J.A. Wagmaister, T. Kawashima, A.Q. Bui, J.J. Harada and R.B. Goldberg (2007) Using genomics to study legume seed development. Plant Physiol. 144: 562–574.

Li, B., L. Tian, J. Zhang, L. Huang, F. Han, S. Yan, L. Wang, H. Zheng and J. Sun (2014) Construction of a high-density genetic map based on large-scale markers developed by specific length amplified fragment sequencing (SLAF-seq) and its application to QTL analysis for isoflavone content in Glycine max. BMC Genomics 15: 1086.

Lygin, A.V., O.V. Zernova, C.B. Hill, N.A. Kholina, J.M. Widholm, G.L. Hartman and V.V. Lozovaya (2013) Glyceollin is an important component of soybean plant defense against *Phytophthora sojae* and *Macrophomina phaseolina*. Phytopathology 103: 984–994.

Mameda, R., T. Waki, Y. Kawai, S. Takahashi and T. Nakayama (2018) Involvement of chalcone reductase in the soybean isoflavone metabolon: identification of GmCHR5, which interacts with 2-hydroxyisoflavanone synthase. Plant J. 96: 56–74.

Meksem, K., V.N. Njiti, W.J. Banz, M.J. Iqbal, M.M. Kassem, D.L. Hyten, J. Yuang, T.A. Winters and D.A. Lightfoot (2001) Genomic regions that underlie soybean seed isoflavone content. J. Biomed. Biotechnol. 1: 38–44.

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

Nwachukwu, I.D., F.B. Luciano and C.C. Udenigwe (2013) The inducible soybean glyceollin phytoalexins with multifunctional health-promoting properties. Food Res. Int. 54: 1208–1216.

Pei, R., J. Zhang, L. Tian, S. Zhang, F. Han, S. Yan, L. Wang, B. Li and J. Sun (2018) Identification of novel QTL associated with soybean isoflavone content. Crop J. 6: 244–252.

Primomo, V.S., V. Poysa, G.R. Ablett, C.-J. Jackson, M. Gijzen and I. Rajcan (2005) Mapping QTL for individual and total isoflavone content in soybean seeds. Crop Sci. 45: 2454–2464.

R Development Core Team (2008) R: A Language and Environment for Statistical Computing.

Rasolohery, C.A., M. Berger, A.V. Lygin, V.V. Lozovaya, R.L. Nelson and J. Daydé (2008) Effect of temperature and water availability during late maturation of the soybean seed on germ and cotyledon isoflavone content and composition. J. Sci. Food Agric. 88: 218–228.

Salvo, V.A., S.M. Boue, J.P. Fonseca, S. Elliott, C. Corbitt, B.M. Collins-Burow, T.J. Curiel, S.K. Srivastav, B.Y. Shih, C. Carter-Wientjes *et al.* (2006) Antiestrogenic glyceollins suppress human breast and ovarian carcinoma tumorigenesis. Clin. Cancer Res. 12: 7159–7164.

Shelton, D., M. Stranne, L. Mikkelsen, N. Pakseresht, T. Welham, H. Hiraka, S. Tabata, S. Sato, S. Paquette, T.L. Wang *et al.* (2012)

Transcription factors of Lotus: regulation of isoflavonoid biosynthesis requires coordinated changes in transcription factor activity. Plant Physiol. 159: 531–547.

Steele, C.L., M. Gijzen, D. Qutob and R.A. Dixon (1999) Molecular characterization of the enzyme catalyzing the aryl migration reaction of isoflavonoid biosynthesis in soybean. Arch. Biochem. Biophys. 367: 146–150.

Tsukamoto, C., S. Shimada, K. Igita, S. Kudou, M. Kokubun, K. Okubo and K. Kitamura (1995) Factors affecting isoflavone content in soybean seeds: Changes in isoflavones, saponins, and composition of fatty acids at different temperatures during seed development. J. Agric. Food Chem. 43: 1184–1192.

van de Weijer, P.H.M. and R. Barentsen (2002) Isoflavones from red clover (Promensil) significantly reduce menopausal hot flush symptoms compared with placebo. Maturitas 42: 187–193.

Waki, T., D. Yoo, N. Fujino, R. Mameda, K. Denessiouk, S. Yamashita, R. Motohashi, T. Akashi, T. Aoki, S. Ayabe *et al.* (2016) Identification of protein–protein interactions of isoflavonoid biosynthetic enzymes with 2-hydroxyisoflavanone synthase in soybean (*Glycine max* (L.) Merr.). Biochem. Biophys. Res. Commun. 469: 546–551.

Wang, Y., Y. Han, X. Zhao, Y. Li, W. Teng, D. Li, Y. Zhan and W. Li (2015) Mapping isoflavone QTL with main, epistatic and QTL × environment effects in recombinant inbred lines of soybean. PLoS ONE 10: e0118447.

Watanabe, S., C. Tsukamoto, T. Oshita, T. Yamada, T. Anai and A. Kaga (2017) Identification of quantitative trait loci for flowering time by a combination of restriction site-associated DNA sequencing and bulked segregant analysis in soybean. Breed. Sci. 67: 277–285.

Welle, R. and H. Grisebach (1988) Isolation of a novel NADPH-dependent reductase which coacts with chalcone synthase in the biosynthesis of 6′-deoxychalcone. FEBS Lett. 236: 221–225.

Yamagata, Y., A. Yoshimura, T. Anai and S. Watanabe (2018) Selection criteria for SNP loci to maximize robustness of high-resolution melting analysis for plant breeding. Breed. Sci. 68: 488–498.

Yi, J., M.R. Derynck, X. Li, P. Telmer, F. Marsolais and S. Dhaubhadel (2010) A single-repeat MYB transcription factor, GmMYB176, regulates *CHS8* gene expression and affects isoflavonoid biosynthesis in soybean. Plant J. 62: 1019–1034.

Yoshikawa, T., Y. Okumoto, D. Ogata, T. Sayama, M. Teraishi, M. Terai, T. Toda, K. Yamada, K. Yagasaki, N. Yamada *et al.* (2010) Transgressive segregation of isoflavone contents under the control of four QTLs in a cross between distantly related soybean varieties. Breed. Sci. 60: 243–254.

Yu, O. and B. McGonigle (2005) Metabolic engineering of isoflavone biosynthesis. Adv. Agron. 86: 147–190.

Zeng, G., D. Li, Y. Han, W. Teng, J. Wang, L. Qiu and W. Li (2009) Identification of QTL underlying isoflavone contents in soybean seeds among multiple environments. Theor. Appl. Genet. 118: 1455–1463.