Research Paper

Single-base deletion in *GmCHR5* increases the genistein-to-daidzein ratio in soybean seed

Md. Abdur Rauf Sarkar[1,2,3], Wakana Otsu[2], Akihiro Suzuki[1,2], Fumio Hashimoto[1], Toyoaki Anai[1,2] and Satoshi Watanabe*[1,2]

[1] *The United Graduate School of Agricultural Sciences, Kagoshima University*, 1-21-24 Korimoto, Kagoshima 890-0065, Japan

[2] *Faculty of Agriculture, Saga University*, 1 Honjo-machi, Saga, Saga 840-8502, Japan

[3] *Department of Genetic Engineering and Biotechnology, Faculty of Biological Science and Technology, Jashore University of Science and Technology*, Jashore 7408, Bangladesh

Novel mutant alleles related to isoflavone content are useful for breeding programs to improve the disease resistance and nutritional content of soybean. However, identification of mutant alleles from high-density mutant libraries is expensive and time-consuming because soybean has a large, complicated genome. Here, we identified the gene responsible for increased genistein-to-daidzein ratio in seed of the mutant line F333ES017D9. For this purpose, we used a time- and cost-effective approach based on selective genotyping of a small number of F_2 plants showing the mutant phenotype with nearest-neighboring-nucleotide substitution–high-resolution melting analysis markers, followed by alignment of short reads obtained by next-generation sequencing analysis with the identified locus. In the mutant line, *GmCHR5* harbored a single-base deletion that caused a change in the substrate flow in the isoflavone biosynthetic pathway towards genistein. Mutated *GmCHR5* was expressed at a lower level during seed development than wild-type *GmCHR5*. Ectopic overexpression of *GmCHR5* increased the production of daidzein derivatives in both the wild-type and mutant plants. The present strategy will be useful for accelerating identification of mutant alleles responsible for traits of interest in agronomically important crops.

Key Words: soybean, isoflavone, daidzein, genistein, NGS, mutant line, gene identification.

Introduction

In leguminous plants such as soybean (*Glycine max* (L.) Merrill), isoflavones and their derivatives play important roles in microbe–plant interactions related to the establishment of symbiosis (Subramanian *et al.* 2006) and the biosynthesis of phytoalexins for defense against pathogens (Welle *et al.* 1991), and have health benefits as part of the human diet (Zaheer and Akhtar 2017). These compounds are synthesized through the phenylpropanoid pathway (**Supplemental Fig. 1**), which involves multiple biosynthetic enzymes (Hahlbrock and Scheel 1989). In the first step, chalcone synthase (CHS) catalyzes the conversion of *p*-coumaroyl-coenzyme A and malonyl-coenzyme A to naringenin chalcone (Ayabe *et al.* 1988). At the same time, CHS and chalcone reductase (CHR) catalyze the synthesis of isoliquiritigenin chalcone from the same substrates

(Bomati *et al.* 2005, Welle and Grisebach 1988). Next, naringenin chalcone and isoliquiritigenin chalcone are converted to the tricyclic flavanones naringenin and liquiritigenin, respectively, by the enzyme chalcone isomerase (CHI) (Ralston *et al.* 2005). Finally, naringenin and liquiritigenin are converted to the major aglycone isoflavones genistein and daidzein by an isoflavone-specific enzyme, isoflavone synthase, which catalyzes the conversion of flavanones to their corresponding isoflavones via 2,3-aryl ring migration (Akashi *et al.* 1999, Jung *et al.* 2000, Steele *et al.* 1999).

In soybean, duplication and inversion of *CHS* genes located in the *I* locus results in siRNA-mediated silencing of genes responsible for yellow seed coat (Kurauchi *et al.* 2009, Tuteja *et al.* 2009). Single-base deletion in the gene encoding flavonoid 3′-hydroxylase (*T* locus) changes pubescence color from tawny to gray (Toda *et al.* 2002). Soybean flower color is regulated by the genes encoding flavonoid 3′5′-hydroxylase (*W1* locus) (Zabala and Vodkin 2007), R2R3-type MYB transcription factor (*W2* locus) (Takahashi *et al.* 2013), dihydroflavonol 4-reductase 2 (*W4* locus) (Yan *et al.* 2014), flavonol synthase (*Wm* locus) (Takahashi *et al.* 2006), and flavanone 3-hydroxylase (*Wp*

locus) (Zabala and Vodkin 2005). Thus, many enzymes and transcription factors related to the flavonoid biosynthetic pathway are associated with the seed coat and flower color of soybean.

Traditionally, map-based cloning approaches use populations obtained from a cross between two parental lines with different genetic backgrounds. However, it is difficult to predict the genotype associated with a target trait in an individual because the number of segregated loci and the effect of each locus are often unknown. Therefore, additional materials such as backcross inbred lines and residual heterozygous lines where the target locus is segregated as a single Mendelian factor are often needed (Yamanaka *et al.* 2005). In contrast, mutant lines share almost the same genetic background with the wild-type line, which allows the use of mutant libraries to examine the function of genes related to specific biological traits. Using mutant libraries, many genes have been isolated from model plants, such as *Arabidopsis* (Azpiroz-Leehan and Feldmann 1997), *Lotus japonicus* (Urbański *et al.* 2013), and rice (Wang *et al.* 2013). However, constructing a high-density mutant library is time-consuming and labor-intensive for crops such as soybean due to their long growing seasons and the large area needed to cultivate a large number of individuals. To date, several mutant libraries for soybean have been reported (Anai 2012, Cooper *et al.* 2008, Li *et al.* 2017, Tsuda *et al.* 2015), but the use of such libraries has largely been restricted to identification of genes related to flowering (Cao *et al.* 2015, Watanabe *et al.* 2009, 2011, Xia *et al.* 2012) and seed fatty acid biosynthesis (Anai *et al.* 2008, 2012, Hoshino *et al.* 2010, 2014).

The genetic basis of the diversity of isoflavone content in soybean has been widely evaluated by quantitative trait locus (QTL) analysis using populations obtained from bi-parental lines displaying contrasting isoflavone content phenotypes or by genome-wide association study using germplasms (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). A total of 293 QTLs related to isoflavones (61 for daidzein, 69 for genistein, 72 for glycitein, 91 for total isoflavone content) are currently listed in the SoyBase database (April 2019; https://www.soybase.org/). We have reported that isoflavone accumulation is tissue dependent; we observed weak correlations between the isoflavone content of cotyledon and other parts of the seed (plumule, epicotyl, hypocotyl, and radicle), indicating that several genetic factors control isoflavone accumulation in soybean seed (Watanabe *et al.* 2019). The genes responsible for the reported QTLs are yet to be fully elucidated because a complicated range of environmental factors—including temperature during seed development, water availability in soil, and cultivation area—also affect the isoflavone content of soybean seed (Hoeck *et al.* 2000, Rasolohery *et al.* 2008, Tsukamoto *et al.* 1995).

Here, we characterized a mutant soybean line with altered isoflavone content in seed (i.e., increased genistein derivative content and decreased daidzein derivative content) compared with the wild-type line. We then identified the gene responsible for this mutation by using a combination of a rapid mapping procedure and next-generation sequencing (NGS) analysis. Finally, we discuss the usefulness of this mutant allele for improving the isoflavone content of soybean seed.

Materials and Methods

Mutant screening

Mutant line F333ES017D9 was identified from an ethyl methanesulfonate–treated 'Fukuyutaka' mutant library of 3331 lines reported previously for the isolation of soybean flowering time mutants (Xia *et al.* 2012). Among these lines, we choose 2720 lines that produced enough seeds (>20); five seeds from each of the 2720 lines were crushed and 100 mg of seed powder was used for extraction of isoflavones (see *Measurement of isoflavone content*). We selected F333ES017D9 as a candidate line with a high proportion of malonylgenistin in the total isoflavone content in its seeds; we propagated seeds of this line and confirmed the stability of the mutant phenotype.

Plant materials

Two Japanese soybean cultivars and one mutant line were used as parental lines: 'Fukuyutaka' (FUKU), 'Toyoshirome' (TOYO), and mutant line F333ES017D9 (MUT). FUKU and TOYO were bred at the National Agriculture Research Organization (Ibaraki, Japan). MUT was identified as described above in *Mutant screening*. To confirm the reliability of the mutant phenotype, FUKU and MUT (three plants per line) were grown in the summer of 2017 in a field owned by Saga University, Japan (33°14′N 130°17E′).

For mapping experiments and to estimate the effects of the mutated gene, three populations were obtained by crossing TOYO and MUT (POP1 and POP2) and by crossing FUKU and MUT (POP3). POP1 comprised 118 F_2 plants and the three parental lines (single plant per parental line was used for phenotyping) grown under natural daylength conditions in a field owned by Saga University. The seeds were sown on 19 July 2018 and the plants were harvested the following November. Seeds obtained from each F_2 plant were used for genotyping and phenotyping. POP2 was made from F3 seeds from POP1, and a single seed from each of 47 F_2 individuals of POP1 was used for evaluation of the effect of the mutated gene on tissue-specific isoflavone accumulation. POP3 comprised 62 F_2 plants grown in a greenhouse. The seeds were sown on 1 February 2019, with approximately 30 plants in a single 128-well nursery tray, and were grown under extended-light conditions (14.5 h) until the end of March to avoid the induction of flowering under short-day conditions. Seeds were

harvested from all 62 plants, and a single F_3 seed from each plant was used for genotyping and phenotyping.

DNA extraction and DNA marker analysis

Genomic DNA was extracted from seed powder (approx. 50 mg) by using the CTAB method (Murray and Thompson 1980) and an automated DNA extraction machine (GENE PREP STAR PI-480; Kurabo Industries Ltd., Osaka, Japan) in accordance with the manufacturer's instructions. All genotyping experiments were performed with DNA diluted 40 times in distilled water. To construct a linkage map of the mutated gene, we used nearest-neighboring-nucleotide substitution–high-resolution melting (NNNs-HRM) analysis and primers designed to detect single-nucleotide polymorphisms (SNPs), as reported (Yamagata *et al.* 2018). To identify the mutated locus, a modified selective genotyping method (Darvasi 1997) was used together with 376 primer pairs designed using SNPs between FUKU and TOYO (Watanabe *et al.* 2017). A LightCycler 96 system (Roche Diagnostics K.K., Tokyo, Japan) equipped with the manufacturer's software (LightCycler 96 Software ver. 1.1.0.1320) was used to detect polymorphisms in polymerase chain reaction (PCR) fragments containing SNPs for NNNs-HRM markers, and to determine the genotypes of individuals; the PCR conditions for genotyping were as described previously (Yamagata *et al.* 2018). We also designed 8 additional NNNS-HRM markers linked to the mutated locus from the SNP list between FUKU and TOYO for QTL analysis. The primers used for NNNs-HRM markers (Yamagata *et al.* 2018) are listed in **Supplemental Table 1**.

Measurement of isoflavone content

Isoflavones were extracted from whole seed and from isolated seed tissues. For screening of the mutant library, 5 freeze-dried whole seeds were crushed into a fine powder with a powder mill with a single replication. For analysis of FUKU and MUT, 10 freeze-dried whole seeds were crushed into a fine powder with three biological replications. For the analysis of POP1, 10 seeds from each F_2 individual, FUKU, TOYO, and MUT were mixed together. For the analysis of POP2 and POP3, a single F_3 seed from each plant was collected, the seed coat was removed, and the cotyledon and "hypocotyl" (plumule, epicotyl, hypocotyl, and radicle) were separated with a knife. Each freeze-dried cotyledon was crushed into fine powder with a powder mill. Samples of powdered cotyledons were used for DNA extraction (50 mg) and for isoflavone extraction (50 mg). Isoflavones were extracted from the hypocotyls of POP2 and POP3 seeds. Freeze-dried hypocotyls were weighed for calculation of isoflavone content by weight, placed in a 3-mL plastic tube provided by Yasuikikai (Osaka, Japan) with 5-mm metal beads (Yasuikikai), and then crushed into a fine powder with a Multibeads shocker (Yasuikikai). Then, powder obtained from whole seed (100 mg) or cotyledon (50 mg), or all of the powder obtained from hypocotyl, was

suspended in 1 mL of extraction buffer containing 70% ethanol (v/v) and 0.1% acetic acid (v/v). Extraction and separation by high-pressure liquid chromatography (Jasco Corp., Tokyo, Japan) were conducted as described previously (Watanabe *et al.* 2019). Extraction was performed three times, and the combined extract (approx. 3 mL) was filtered through a 0.45-μm nitrocellulose membrane (GE Healthcare UK Ltd., Buckinghamshire, England). Filtrate samples (3 μL) were analyzed by high-pressure liquid chromatography on a Hydrosphere C18 column (YMC Co., Ltd., Kyoto, Japan). External standard curves for daidzein, daidzin, malonyldaidzin, glycitein, glycitin, malonylglycitin, genistein, genistin, and malonylgenistin (all purchased from Wako, Osaka, Japan) were constructed, and peak areas for the extracted isoflavones were calculated according to their retention times and the areas of the standards. To evaluate mutant phenotype, G/D score was used, which was defined as the ratio of genistein-type isoflavones (genistin [Ge] and malonygenistin [Mge]) to daidzein-type isoflavones (daidzin [Da] and malonyldaidzin [Mda]). If the amount of an isoflavone was under the detectable level, we excluded that component from the statistical analysis.

NGS analysis

DNA used for whole-genome sequencing analysis was extracted from young leaves before full expansion, for parental lines FUKU and TOYO or from combined powdered whole seed from the 16 POP1 individuals with the highest G/D scores (mixed-MUT) . NGS analysis was performed on our behalf by Novogene (Beijing, China). Approximately 17-Gbp sequences were obtained for each sample. A soybean reference sequence (data version, Glyma 1.0) was obtained from a public database (https://phytozome.jgi.doe.gov). FASTQ files containing sequence data were used to align short reads to a candidate region (58.4–61.4 Mbp region of Gm18) by using the Bowtie 2 program (ver. 2.3.5.1, Langmead and Salzberg 2012). We used the SAMtools suite (ver. 0.1.18, Li *et al.* 2009) to detect polymorphisms among the three extracted DNAs. Default parameters were used for short read alignment and to extract polymorphisms. We used IGV viewer (ver. 2.4.10, Robinson *et al.* 2017) to manually confirm each polymorphism between the sequences of FUKU, TOYO, and mixed-MUT.

Expression analysis using developing seed from parental lines

Parental lines FUKU and MUT were grown in a greenhouse (February to June 2018) under a long-day photoperiod (14.5 h/9.5 h) until the end of March. Developing seed was collected randomly from several plants at 40, 55, and 70 days after flowering (DAF). Total RNA was isolated from developing seed (50–100 mg) with a Total RNA Extraction Kit Mini Plant (SciTrove, Tokyo, Japan) with additional rDNase I (Takara Bio Inc., Shiga, Japan)

treatment. cDNA was synthesized from 1 μg total RNA with ReverTra Ace (Toyobo, Osaka, Japan). A 5-μL aliquot (approx. 25 ng) of cDNA was used as a template. Quantitative real-time PCR was conducted as follows in a LightCycler 96 system: 95°C for 5 min for activation of enzyme, 95°C for 15 s, 60°C for 15 s, 72°C for 20 s, for a total of 40 amplification cycles. EvaGreen Dye (20× in water; Biotium, Inc., Fremont, CA, USA) was used as the fluorescent dye, and dNTPs (Takara), homemade recombinant *Taq* polymerase, the PCR buffer mentioned in a previous study (Yamagata *et al.* 2018), and 1 pmol of each gene-specific primer were used to evaluate transcript levels of target genes. Expression relative to *GmACT2/7* (*Glyma. 19G147900*), the internal reference gene, was calculated by using the 2-ΔΔCt method (Schmittgen and Livak 2008). Data from four technical replicates were analyzed. The primers used are listed in **Supplemental Table 1**.

Construct preparation for ectopic expression of GmCHR5

The full-length coding sequence of *GmCHR5* (*Glyma. 18G285800*) was amplified from FUKU seed cDNA by PCR using KOD FX Neo polymerase (Toyobo). The PCR product was cloned into the entry vector pENTERTM/D-TOPO via the TOPO Cloning reaction (Thermo Fisher Scientific, Waltham, MA, USA) and sequenced. The cloned fragment was transferred into the destination vector pUB-GW-GFP (Maekawa *et al.* 2008) by the Gateway LR recombination reaction (Thermo Fisher Scientific) and designated GmCHR5-OE. GmCHR5-OE and the pUB-GW-GFP vector were introduced into *Agrobacterium rhizogenes* strain LBA1334 by a freeze–thaw method (Holsters *et al.* 1978). The primers for cloning of the *GmCHR5* gene are listed in **Supplemental Table 1**.

Soybean hairy root transformation

Soybean transgenic hairy roots were generated as reported previously (Chen *et al.* 2018) with some modifications. Briefly, FUKU and MUT seeds were sterilized with chlorine gas for 7 to 8 h, transferred to MS medium without sucrose (Murashige and Skoog 1962), and germinated for 5 to 7 days under a 16 h/8 h photoperiod at 22 to 25°C. LBA1334 harboring GmCHR5-OE or pUB-GW-GFP vector were re-suspended in sucrose-free MS liquid medium containing freshly prepared 100 μM acetosyringone (Sigma-Aldrich, St. Louis, MO, USA) to a final OD$_{600}$ of 0.6 and used for inoculation. Dark-green, healthy cotyledons were separated from the hypocotyl to make two cotyledons. The point where the cotyledon and hypocotyl join was gently wounded with a sterile blade and soaked in the medium containing LBA1334 for several minutes. Co-cultivation was conducted in the dark at 22 to 25°C for 3 days. Cotyledons were then incubated on the induction culture medium in the dark at 25°C for 3 to 4 weeks. Transgenic hairy roots with strong green fluorescent protein (GFP) signal were selected, weighed, placed in a 3-mL plastic tube with 5-mm metal beads (Yasuikikai), frozen in liquid nitrogen, crushed with a Multibeads shocker, and used for isoflavone extraction as described above.

QTL analysis and statistical analysis

A linkage map of Gm18 was constructed in AntMap software (ver. 1.2, Iwata and Ninomiya 2006) using the default parameters for 96 plants from POP1 (selected in the order of their sample ID numbers). QTLs were identified by single-interval mapping in R/qtl software (ver. 1.44-9, Broman *et al.* 2003). Linear regression analysis was used to confirm the effects of the QTLs by using the genotype of the closest DNA marker in 96 individuals of POP1 and their isoflavone content. Linear regression analysis was also used to estimate the effect of the mutated locus by using the genotype of the DNA marker targeting the polymorphism detected in the candidate gene (GmCHR5_1bp_del; see results section) in 96 individuals in POP1, 47 individuals in POP2, and 62 individuals in POP3, and their isoflavone content. Statistical analysis for the ectopic expression of *GmCHR5* was conducted with single-factor ANOVA and Tukey's post-hoc test, with $P = 0.05$ considered the cut-off for statistical significance. All statistical analyses were performed in R software (ver. 3.6.1, R Development Core Team 2008).

Results

### Identification of a mutant line with altered seed isoflavone content

We screened 2720 out of 3331 lines in an ethyl methane-sulfonate–treated FUKU mutant library (Xia *et al.* 2012) and obtained several lines with a unique isoflavone phenotype. One of those lines, F333ES017D9 (MUT), showed a much higher malonylgenistin (Mge) content (72.5% of total isoflavone context) in whole seed compared with the average value of the mutant library (50.7 ± 6.92%, **Supplemental Fig. 2**). MUT showed a very high G/D score (4.31) compared with the average value of the mutant library (1.4 ± 0.44, **Supplemental Fig. 2**). We confirmed the mutant phenotype by propagating seeds obtained from plants grown in the 2017 summer season and found that MUT also showed a significantly decreased daidzin (Da), malonyldaidzin (Mda), glycitin, and malonylglycitin (Mgly) content in whole seed (**Table 1**). The G/D score of MUT (9.1) was significantly higher than that of FUKU (2.4; **Table 1**). However, there was no significant difference in total isoflavone content between MUT and FUK (**Table 1**).

Identification of the locus of the mutated gene

We used selective genotyping and QTL analysis to identify the locus of the mutated gene in an F$_2$ population obtained from the cross between MUT and TOYO (POP1). We expected to find segregation of genetic loci related to isoflavone content, besides the mutated gene, within the population because of the difference in genetic background

Table 1. Phenotypes of FUKU and MUT for whole-seed isoflavone content (μg/g)

Isoflavone	FUKU ($n=3$)		MUT ($n=3$)		*P*-value
	Average	SD[a]	Average	SD	Student's *t*-test
Daidzin	10.0	0.10	3.4	1.95	$P < 0.01$
Genistin	33.4	6.43	44.5	10.05	ns[c]
Glycitin	14.1	1.77	6.7	1.36	$P < 0.01$
Malonyldaidzin	394.2	34.71	148.4	76.70	$P < 0.05$
Malonylgenistin	917.7	114.10	1132.1	210.11	ns
Malonylglycitin	52.4	4.73	17.0	2.59	$P < 0.01$
G/D score[b]	2.4	0.30	9.1	2.53	$P < 0.05$
Total isoflavone	1421.8	133.84	1352.1	295.27	ns

[a] SD, standard deviation.
[b] G/D score, ratio of genistein-type isoflavones (genistin + malonyl-genistin) to daidzein-type isoflavones (daidzin + malonyldaidzin).
[c] ns, not significant.
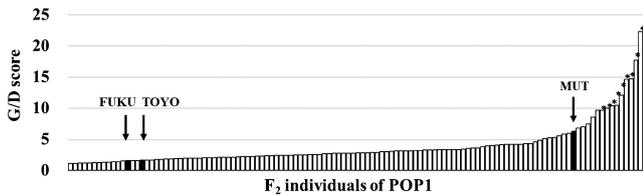
Fig. 1. Distribution of G/D scores in POP1. Each POP1 plant is ordered according to its G/D score. G/D scores of the parental lines FUKU, TOYO, and MUT are shown in black. Asterisks indicate the 10 plants with the highest G/D score, which are discussed in the main text.

between FUKU and TOYO. The G/D scores were FUKU (1.5), TOYO (1.6), and MUT (6.2) for plants grown in the 2018 season, and we concluded that G/D score was a suitable trait for evaluation of the effect of the mutant gene. The G/D score in POP1 ranged from 1.1 to 23.0 (**Fig. 1**), indicating that the MUT allele responsible for a high G/D score was possibly segregated in POP1, and that G/D score

was affected by other segregating loci between FUKU and TOYO or by environmental factors because of the continuous distribution of G/D score. We considered that QTL analysis would be useful for identifying the mutated locus in POP1 and applied a selective genotyping approach to find the chromosomal region associated with the mutated locus in MUT. We selected the 10 plants in POP1 with the highest G/D score for selective genotyping, and screened the SNP loci linked to the mutant locus based on the genotype of the selected plants. A homozygous FUKU genotype determined with NNNs-HRM markers would indicate a strong link between the analyzed DNA marker and the mutated loci. In contrast, DNA markers not linked to the mutated locus would show a random genotype (homozygous for FUKU or TOYO, or heterozygous). In total, 348 of 376 (92.6%) NNNs-HRM markers covering the whole soybean genome showed polymorphisms between FUKU and TOYO. Among them, we identified only two markers, Gm18_60238623 and Gm18_60832874 (marker names are based on chromosome number and physical position), showing complete coincidence with the FUKU genotype in all selected plants.

We then constructed a linkage map for soybean chromosome 18 (Gm18) with additional NNNs-HRM markers. Linkage between the mutated gene and these markers was confirmed by QTL analysis using 96 plants from POP1. An LOD (logarithm of the odds) peak was detected at 29 cM (closest marker, Gm18_61024346 at 35.2 cM; **Fig. 2A**). We predicted the genetic effects of this mutated gene by linear regression analysis with the closest marker (Gm18_61024346) and found that this locus could explain 54.9% of the phenotypic variance of G/D score in POP1 (**Fig. 2B**). The mutated gene was located within an interval between Gm18_58420729 and Gm18_61355250.

Identification of the responsible gene by NGS analysis

Next, we identified the gene responsible for the target
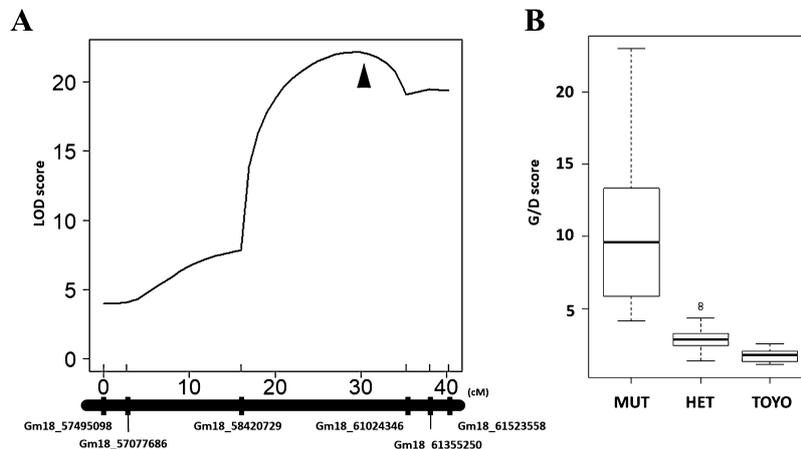
Fig. 2. Identification of mutant locus by QTL analysis of POP1. (a) LOD score curve for the linkage map of part of Gm18. Marker names are presented as the combination of chromosome name and physical position. (B) Boxplot showing phenotypic classification with the DNA marker closest to the mutated locus. HET, heterozygous.

mutation by using the MutMap technique (Abe *et al.* 2012) for a targeted chromosomal region and NGS whole-genome sequence data for FUKU, TOYO, and mixed DNA of POP1 plants showing the MUT phenotype (mixed-MUT). A total of 72.1 K short reads (150-bp paired-end sequences) for FUKU, 71.3 K for TOYO, and 63.8 K for mixed-MUT were obtained. The estimated sequence volume for each sample was 17.4 (mixed-MUT), 19.4 (TOYO), and 19.7 times (FUKU) that of the reference soybean genome (1.1 Gbp; Schmutz *et al.* 2010). From the result of QTL analysis of Gm18, we extracted short reads mapped to a 58.4- to 61.4-Mbp region covering the mutated gene locus. Three nucleotide substitutions and two deletions were polymorphic between FUKU and MUT (**Supplemental Table 2**). Among these mutated sites, a single-base deletion was located in the third exon of *Glyma18g52250* (gene ID in Wm82.a2.v1 database, *Glyma.18G285800*). In the MUT allele, the guanine nucleotide at position 620 from the start codon of the coding sequence was deleted, resulting in a protein of 212 amino acids (aa) instead of 394 aa for the wild-type protein (**Fig. 3**). This gene was annotated in the database (https://phytozome.jgi.doe.gov/pz/portal.html) as 6′-deoxychalcone synthase; the product of this gene (*GmCHR5*) shows chalcone reductase activity and is responsible for control the relative concentrations of daidzein and genistein derivatives in soybean seed (Mameda *et al.* 2018). We hypothesized that the mutation of *GmCHR5* resulted in reduced substrate flow for daidzein synthesis and increased substrate flow for genistein synthesis.

Genetic effects of the 1-bp deletion in GmCHR5 on isoflavone content

We next confirmed the reliability of the NGS data and the effect of the mutant allele by using several mapping populations. We developed additional HRM markers targeting the 1-bp deletion in *GmCHR5* (hereafter GmCHR5_1bp_del). The melting temperature of the PCR product with this deletion is lower than that of the wild type (**Supplemental Fig. 3**). GmCHR5_1bp_del was mapped to the locus between Gm18_ Gm18_58420729 and Gm18_61024346. Only two recombinations were observed between GmCHR5_1bp_del and Gm18_61024346.

We next used linear regression analysis to further examine the genetic effects of GmCHR5_1bp_del on isoflavone phenotype in POP1, POP2, and POP3 (**Tables 2–4**). GmCHR5_1bp_del explained 59.7% (G/D scores), 46.7% (Mda), 37.2% (Da), 33.7% (Mge), 17.0% (Ge), and 13.7% (Mgly) of phenotypic variances for isoflavones in POP1 (**Table 2**). In dissected seed tissue from POP2, GmCHR5_1bp_del increased the G/D score of cotyledon and hypocotyl tissues, and 42.7% of the cotyledon G/D score variance and 59.9% of that of the hypocotyl could be explained by this mutation (**Table 3**). Increased production of genistein derivatives (Ge and Mge) and decreased production of daidzein derivatives (Da and Mda) was clear in both cotyledon and hypocotyl tissues in the MUT compared with the TOYO genotype (**Table 3**). We also confirmed the genetic effects of GmCHR5_1bp_del in the backcross population POP3, which was grown in a greenhouse in winter. We detected a significant association between G/D score in cotyledon (24.0%) and in hypocotyl (28.3%) (**Table 4**).

These mapping experiments indicated that the 1-bp deletion in *GmCHR5* was responsible for the altered production of daidzein and genistein derivatives in soybean seed tissues and that the effect of this mutation on G/D score was affected by the differences in growth conditions between the field (2018 summer season) and greenhouse (2019 winter season).
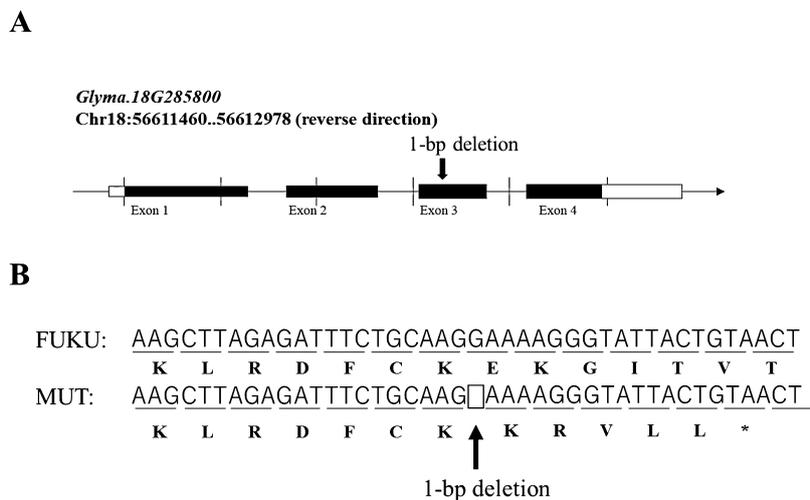

A

Glyma.18G285800
Chr18:56611460..56612978 (reverse direction)

1-bp deletion

Exon 1   Exon 2   Exon 3   Exon 4

B

FUKU: AAGCTTAGAGATTTCTGCAAGGAAAAGGGTATTACTGTAACT
            K   L   R   D   F   C   K   E   K   G   I   T   V   T

MUT: AAGCTTAGAGATTTCTGCAAG☐AAAAGGGTATTACTGTAACT
            K   L   R   D   F   C   K   K   R   V   L   L   *

1-bp deletion

Fig. 3. Gene structure and location of the single-base deletion in MUT *GmCHR5* (*Glyma.18G285800*). (A) Exons and untranslated regions of *GmCHR5* are shown as black and white rectangles, respectively. The interval between two vertical lines is 250 bp. (B) Nucleotide sequence around the 1-bp deletion (white box) and the amino acid sequence (single capital letters) from residue 201 from the start codon of *GmCHR5* mRNA.

Table 2. Effect of GmCHR5_1bp_del on isoflavone content (µg/g) in whole seed of POP1

	Genotype								
	MUT homozygous		Heterozygous		TOYO homozygous				
Number of plants	19		53		24				
Isoflavone	Average	SD[a]	Average	SD	Average	SD	*F*-value	PVE[b]	*P*-value
Daidzin	9.4	5.2	26.8	9.7	39.1	19.8	27.0	37.2%	$P < 0.001$
Genistin	149.6	52.9	131.3	32.8	100.3	32.7	9.5	17.0%	$P < 0.001$
Malonyldaidzin	395.7	205.5	970.4	291.1	1307.9	451.6	40.7	46.7%	$P < 0.001$
Malonylgenistin	3035.8	555.3	2622.9	456.3	2091.3	309.8	23.7	33.7%	$P < 0.001$
Malonylglycitin	50.0	27.5	69.9	26.5	85.1	35.7	7.4	13.7%	$P < 0.01$
G/D score[c]	10.3	5.6	3.0	0.8	1.8	0.4	68.9	59.7%	$P < 0.001$

[a] SD, standard deviation.
[b] PVE, proportion of phenotypic variance explained.
[c] G/D score, ratio of genistein-type isoflavones (genistin + malonylgenistin) to daidzein-type isoflavones (daidzin + malonyldaidzin).

Table 3. Effect of GmCHR5_1bp_del on isoflavone content (µg/g) in dissected seed tissues from POP2

	Genotype								
	MUT homozygous		Heterozygous		TOYO homozygous				
Number of plants	13		14		20				
Isoflavone	Average	SD[a]	Average	SD	Average	SD	*F*-value	PVE[b]	*P*-value
In cotyledon									
Daidzin	7.7	5.8	34.8	19.6	48.5	31.3	12.1	35.4%	$P < 0.001$
Genistin	164.4	83.3	153.4	58.0	123.7	52.6	1.8	ns[c]	ns
Malonyldaidzin	254.9	151.5	1012.6	439.2	1322.9	653.5	19.7	46.7%	$P < 0.001$
Malonylgenistin	3043.9	821.1	2649.1	720.4	2283.7	625.3	4.7	17.3%	$P < 0.05$
Malonylglycitin	10.6	3.0	10.0	1.4	9.5	3.8	0.5	ns	ns
G/D score[d]	16.5	14.1	3.1	1.1	2.3	1.4	16.4	42.7%	$P < 0.001$
In hypocotyl									
Daidzin	134.7	78.2	306.0	114.1	310.0	82.3	17.8	44.2%	$P < 0.001$
Genistin	350.3	143.3	203.1	107.4	129.3	65.7	18.3	44.8%	$P < 0.001$
Glycitin	584.1	433.9	1306.5	720.0	1476.2	699.7	8.4	27.3%	$P < 0.001$
Malonyldaidzin	5542.3	2344.0	10548.7	3469.7	10965.1	2970.9	15.6	41.0%	$P < 0.001$
Malonylgenistin	10087.1	3203.8	4877.9	1787.3	3565.9	1520.0	37.9	62.8%	$P < 0.001$
Malonylglycitin	2294.6	1103.3	4402.1	1960.5	4854.9	1867.8	9.8	30.4%	$P < 0.001$
G/D score	2.3	1.2	0.6	0.3	0.4	0.2	33.6	59.9%	$P < 0.001$

[a] SD, standard deviation.
[b] PVE, proportion of phenotypic variance explained.
[c] ns, not significant.
[d] G/D score, ratio of genistein-type isoflavones (genistin + malonylgenistin) to daidzein-type isoflavones (daidzin + malonyldaidzin).

Expression of *GmCHR5* during seed development in FUKU and MUT

A search of the RNA atlas in SoyBase (https://www.soybase.org) revealed that *GmCHR5* is expressed during seed development at 25, 28, 35, and 42 DAF. We analyzed the expression profile of *GmCHR5* by extracting RNA from developing seeds of FUKU and MUT at 40, 55, and 70 DAF. In both lines, constitutive *GmCHR5* expression was detected at 40 and 55 DAF, but expression was reduced at 70 DAF (**Fig. 4**). *GmCHR5* expression was much higher in FUKU (about 7.0 times at 40 and 55 DAF; 3.7 times at 70 DAF) than in MUT, suggesting that the mutation in *GmCHR5* reduced its transcription or mRNA stability.

Ectopic expression of *GmCHR5* in transgenic hairy roots

Control hairy roots of FUKU and MUT transformed with the pUB-GW-GFP vector showed a low G/D score (average G/D score for all transformants, 0.29 ± 0.13 [SD]) and Mda was the major isoflavone component of these hairy roots. FUKU (average G/D score, 0.27) showed a low G/D score compared with MUT (0.30, **Fig. 5**). We found a significant decrease in G/D score in 10 independent transgenic FUKU (0.08) and MUT (0.16) hairy roots showing a strong GFP signal and therefore overexpressing *GmCHR5* (**Fig. 5**). These results indicated that ectopic expression of *GmCHR5* increased the production of daidzein derivatives.

Table 4. Effect of GmCHR5_1bp_del on isoflavone content (µg/g) in a backcross population grown in a greenhouse (POP3)

	Genotype								
	MUT homozygous		Heterozygous		FUKU homozygous				
Number of plants	23		21		18				
Isoflavone	Average	SD[a]	Average	SD	Average	SD	F-value	PVE[b]	P-value
In cotyledon									
Daidzin	1.7	2.8	3.6	2.9	6.8	4.1	10.4	29.0%	$P < 0.001$
Genistin	18.1	4.1	17.6	3.8	22.0	8.1	3.8	11.4%	$P < 0.05$
Glycitin	3.7	2.1	5.4	2.9	3.8	1.6	3.6	10.9%	$P < 0.05$
Malonyldaidzin	47.5	65.7	106.6	83.1	179.3	105.5	12.3	29.3%	$P < 0.001$
Malonylgenistin	946.2	1155.2	1078.7	413.8	358.9	316.9	1.6	ns[c]	ns
G/D score[d]	48.1	38.2	23.7	31.8	8.2	6.7	9.3	24.0%	$P < 0.001$
In hypocotyl									
Daidzin	34.6	30.5	53.0	34.2	95.3	45.1	14.4	32.7%	$P < 0.001$
Genistin	452.7	176.5	409.7	124.5	308.6	124.7	5.0	14.6%	$P < 0.01$
Glycitin	107.1	65.7	53.5	69.5	51.0	56.7	3.0	ns	ns
Malonyldaidzin	1370.6	1051.2	1925.7	1053.4	3318.2	1243.5	16.0	35.2%	$P < 0.001$
Malonylgenistin	10838.2	3249.8	9346.5	2103.9	7855.9	2457.6	6.3	17.6%	$P < 0.01$
Malonylglycitin	252.1	202.6	372.0	243.7	643.4	304.7	12.7	30.1%	$P < 0.001$
G/D score	10.7	6.1	7.1	6.2	2.7	1.3	11.7	28.3%	$P < 0.001$

[a] SD, standard deviation.
[b] PVE, proportion of phenotypic variance explained.
[c] ns, not significant.
[d] G/D score, ratio of genistein-type isoflavones (genistin + malonylgenistin) to daidzein-type isoflavones (daidzin + malonyldaidzin).
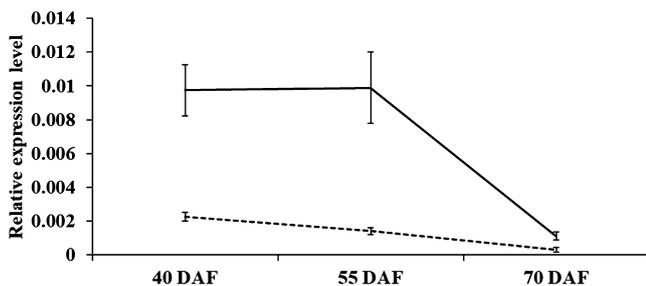

Fig. 4. Expression of *GmCHR5* in developing seed from FUKU and MUT. Expression levels are shown relative to the expression of the internal control gene *GmACT2/7*. *P*-value < 0.001 (two-way ANOVA) for the difference in relative expression between FUKU (solid line) and MUT (dashed line) at all three of the development stages indicated. DAF, days after flowering.
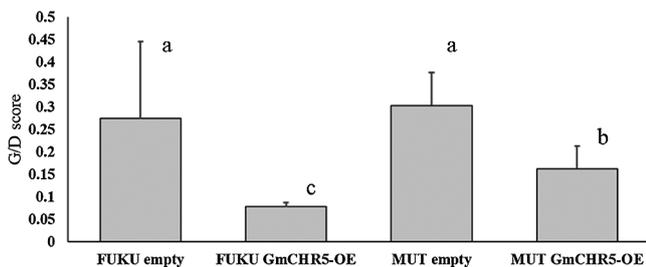

Fig. 5. Ectopic expression of *GmCHR5* in FUKU and MUT transgenic hairy roots. Values are means ± SD (10 independent hairy root lines). Different letters show significant differences by single-factor ANOVA and Tukey's post-hoc test, *P* < 0.05.

Discussion

Here, we identified a gene, *GmCHR5*, mutated in an isoflavone soybean mutant with increased genistein derivative content and reduced daidzein derivative content in whole seed. Single-base deletion in *GmCHR5* resulted in a truncated protein and caused the substrate flow to move away from the production of daidzein derivatives and toward the production of genistein derivatives in the isoflavone biosynthetic pathway. To identify the gene responsible for the mutant phenotype, we applied NNNs-HRM genotyping analysis to a small number (10) of F$_2$ plants with the mutant phenotype; this analysis was followed by alignment of short reads obtained by NGS analysis to the identified QTL region. This procedure is time- and cost-effective in comparison with traditional linkage analysis using a linkage map covering the whole genome. The rapid mapping procedure demonstrated in the present study will be useful for accelerating the identification of genes in agronomically important crops that have large or complicated genomes.

In soybean, 14 *CHR* genes have been identified based on sequence homology, 11 of which are expected to be functional and have tissue-specific expression patterns (Sepiol *et al.* 2017). The expression of *Glyma.18G285800* (*GmCHR5*), *Glyma.14G005700* (*GmCHR1*, named by Mameda *et al.* 2018), and *Glyma.02G307300* (*GmCHR6*) are induced by inoculation with the soybean pathogen

Phytophthora sojae (Sepiol *et al.* 2017). Physical interaction with isoflavone synthase was detected for GmCHR5 but not for GmCHR1 or 6 (Mameda *et al.* 2018, Nakayama *et al.* 2019). The increased expression of *GmCHR5* during seed development correlated well with the isoflavone accumulation pattern in developing seed (Mameda *et al.* 2018). Together, these findings suggest that GmCHR5 plays a major role in the synthesis of daidzein derivatives in soybean seed.

CHR of *Medicago sativa* (312 aa, AAB41555 in a public database; https://www.ncbi.nlm.nih.gov/) is monomeric and has a $(\alpha/\beta)_8$-barrel structure (Bomati *et al.* 2005). A BLAST comparison showed that aa 12–314 of GmCHR5 had a similar sequence to aa 11–312 of *M. sativa* CHR (262/303; 86% similarity). The mutation in the third exon of *GmCHR5* resulted in production of a truncated protein (212 aa) lacking the C-terminal region containing aa residues related to NADP(H) binding (e.g., Pro-301 and Ile-298) (Bomati *et al.* 2005). Further studies to evaluate the enzymatic activity of the MUT allele using expression vectors are necessary to confirm that the allele is null for CHR activity.

The formation of a metabolon comprising GmCHR5, isoflavone, and CHS on the surface of the endoplasmic reticulum is essential for rapid substrate transfer from CHS to CHR (Mameda *et al.* 2018, Waki *et al.* 2016). Therefore, we can expect that GmCHR5 deficiency would increase naringenin production and decrease liquiritigenin production, which is supported by the high proportion of phenotypic variance explained (PVE) for G/D score we obtained for the mutant line. The loss-of-function of *GmCHR5*, however, did not lead to complete daidzein-derivative deficiency, likely because other CHR-encoding genes (e.g., *GmCHR6*, *Glyma.16G219600*, and *Glyma.12G006800*) are also expressed during early seed maturation (Sepiol *et al.* 2017). The PVE for G/D score for the mutant allele varied among the experimental populations (**Tables 2**–**4**) and some POP1 plants showed G/D scores higher than that of MUT (**Fig. 1**). Together, these results indicate that other genetic and environmental factors affect the expression and/or enzymatic activity of GmCHR5.

In transgenic FUKU and MUT hairy roots, we observed a reduction in G/D score by GmCHR5-OE, indicating that exogenous GmCHR5 can contribute to the production of daidzein derivatives. In the mock (control vector only)-infected hairy roots, G/D scores were similar between FUKU and MUT, and low (0.27–0.3; **Fig. 5**) compared with seed (2.4–9.1, **Table 1** and 1.5–6.2, **Fig. 1**). Endogenous daidzein was already synthesized in hairy roots, which may have decreased the difference between the G/D scores of FUKU and MUT. Soybean roots grown under a controlled environment showed a low G/D score and a significant difference between FUKU and MUT (0.037 for FUKU and 0.15 for MUT). These results indicate that the effect of the MUT allele in transgenic hairy roots was unstable and that multiple copies of soybean *GmCHR* would probably

affect the biosynthesis of daidzein in a spatial or temporal manner. The silencing of *GmCHR* by RNAi in hairy roots decreases the production of daidzein derivatives and suppresses race-specific resistance against *P. sojae* (Graham *et al.* 2007) because of reduced synthesis of the daidzein derivatives phytoalexin and glyceollin I. In the present study, we did not observe any negative effects of the *GmCHR5* mutation on plant growth; however, whether this mutation decreases resistance to soybean diseases needs further examination.

Two QTLs have been reported around *GmCHR5*. The first, *q*Sifc-a-18-3 (for total isoflavone content), detected by genome-wide association analysis using 366 Chinese soybean landraces, is located close to the SNP Gm18_57350313 and explains 2% of the PVE (Meng *et al.* 2016). The second, *q*CDZ|proG_1, is a minor QTL that controls daidzein content and was detected between SSR (simple sequence repeat) markers Satt288 and Satt191 (Gm18, 55.4–58.7 Mbp in Glyma 1.0) by bi-parental QTL analysis of two Chinese soybean varieties (Han *et al.* 2015). It is possible that the cause of these QTLs is a mutation of *GmCHR5*. Comparison of the sequences of *GmCHR5* in soybean germplasms of wide geographical origin will be necessary to evaluate the genetic effects of *GmCHR5* on the phenotypic diversity of isoflavone content within soybean. Further study using the mutant line identified in the present study will also be useful to uncover the biological activities of GmCHR5, because this mutant line has a uniform genetic background with the wild-type line.

The allele carrying a mutation of *GmCHR5* found in the present study may be useful as a source allele to produce soybean lines that produce seeds with a high genistein-derivative content. Such lines would have high breeding value because of the potentially beneficial effects of genistein on inflammation, angiogenesis, and the development of metabolic diseases in humans (Mukund *et al.* 2017).

In conclusion, we identified a novel mutant allele, *GmCHR5*, responsible for increased production of major aglycones and their derivatives in a soybean mutant line. This mutant allele could be a useful genetic material for future breeding programs. Our approach, selective genotyping with NNNs-HRM markers followed by NGS analysis of a restricted chromosomal region, is a time- and cost-effective means of identifying the genes responsible for a target phenotypic trait from a dense mutant library, especially for crops with large or complicated genomes.

Literature Cited

Abe, A., S. Kosugi, K. Yoshida, S. Natsume, H. Takagi, H. Kanzaki, H. Matsumura, K. Yoshida, C. Mitsuoka, M. Tamiru *et al.* (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. Nat. Biotechnol. 30: 174–178.

Akashi, T., T. Aoki and S. Ayabe (1999) Cloning and functional expression of a cytochrome P450 cDNA encoding 2-hydroxyisoflavanone synthase involved in biosynthesis of the isoflavonoid skeleton in Licorice. Plant Physiol. 121: 821–828.

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.

Anai, T., T. Yamada, R. Hideshima, T. Kinoshita, S.M. Rahman and Y. Takagi (2008) Two high-oleic-acid soybean mutants, M23 and KK21, have disrupted microsomal omega-6 fatty acid desaturase, encoded by *GmFAD2-1a*. Breed. Sci. 58: 447–452.

Anai, T. (2012) Potential of a mutant-based reverse genetic approach for functional genomics and molecular breeding in soybean. Breed. Sci. 61: 462–467.

Anai, T., T. Hoshino, N. Imai and Y. Takagi (2012) Molecular characterization of two high-palmitic-acid mutant loci induced by X-ray irradiation in soybean. Breed. Sci. 61: 631–638.

Ayabe, S.-I., A. Udagawa and T. Furuya (1988) NAD(P)H-dependent 6′-deoxychalcone synthase activity in *Glycyrrhiza echinata* cells induced by yeast extract. Arch. Biochem. Biophys. 261: 458–462.

Azpiroz-Leehan, R. and K.A. Feldmann (1997) T-DNA insertion mutagenesis in *Arabidopsis*: going back and forth. Trends Genet. 13: 152–156.

Bomati, E.K., M.B. Austin, M.E. Bowman, R.A. Dixon and J.P. Noel (2005) Structural elucidation of chalcone reductase and implications for deoxychalcone biosynthesis. J. Biol. Chem. 280: 30496–30503.

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.

Cao, D., Y. Li, S. Lu, J. Wang, H. Nan, X. Li, D. Shi, C. Fang, H. Zhai, X. Yuan *et al.* (2015) *GmCOL1a* and *GmCOL1b* function as flowering repressors in soybean under long-day conditions. Plant Cell Physiol. 56: 2409–2422.

Chen, L., Y. Cai, X. Liu, C. Guo, S. Sun, C. Wu, B. Jiang, T. Han and W. Hou (2018) Soybean hairy roots produced *in vitro* by *Agrobacterium rhizogenes*-mediated transformation. Crop J. 6: 162–171.

Cooper, J.L., B.J. Till, R.G. Laport, M.C. Darlow, J.M. Kleffner, A. Jamai, T. El-Mellouki, S. Liu, R. Ritchie, N. Nielsen *et al.* (2008) TILLING to detect induced mutations in soybean. BMC Plant Biol. 8: 9.

Darvasi, A. (1997) The effect of selective genotyping on QTL mapping accuracy. Mamm. Genome 8: 67–68.

Graham, T.L., M.Y. Graham, S. Subramanian and O. Yu (2007) RNAi silencing of genes for elicitation or biosynthesis of 5-deoxyisoflavonoids suppresses race-specific resistance and hypersensitive cell death in *Phytophthora sojae* infected tissues. Plant Physiol. 144: 728–740.

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.

Hahlbrock, K. and D. Scheel (1989) Physiology and molecular biology of phenylpropanoid metabolism. Annu. Rev. Plant Physiol. Plant Mol. Biol. 40: 347–369.

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.

Holsters, M., D. de Waele, A. Depicker, E. Messens, M. van Montagu and J. Schell (1978) Transfection and transformation of *Agrobacterium tumefaciens*. Mol. Gen. Genet. 163: 181–187.

Hoshino, T., Y. Takagi and T. Anai (2010) Novel *GmFAD2-1b* mutant alleles created by reverse genetics induce marked elevation of oleic acid content in soybean seeds in combination with *GmFAD2-1a* mutant alleles. Breed. Sci. 60: 419–425.

Hoshino, T., S. Watanabe, Y. Takagi and T. Anai (2014) A novel *GmFAD3-2a* mutant allele developed through TILLING reduces α-linolenic acid content in soybean seed oil. Breed. Sci. 64: 371–377.

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. 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.

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.

Kurauchi, T., T. Matsumoto, A. Taneda, T. Sano and M. Senda (2009) Endogenous short interfering RNAs of chalcone synthase genes associated with inhibition of seed coat pigmentation in soybean. Breed. Sci. 59: 419–426.

Langmead, B. and S.L. Salzberg (2012) Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357–359.

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.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Li, Z., L. Jiang, Y. Ma, Z. Wei, H. Hong, Z. Liu, J. Lei, Y. Liu, R. Guan, Y. Guo *et al.* (2017) Development and utilization of a new chemically-induced soybean library with a high mutation density. J. Integr. Plant Biol. 59: 60–74.

Maekawa, T., M. Kusakabe, Y. Shimoda, S. Sato, S. Tabata, Y. Murooka and M. Hayashi (2008) Polyubiquitin promoter-based binary vectors for overexpression and gene silencing in *Lotus japonicus*. Mol. Plant Microbe Interact. 21: 375–382.

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.

Meng, S., J. He, T. Zhao, G. Xing, Y. Li, S. Yang, J. Lu, Y. Wang and J. Gai (2016) Detecting the QTL-allele system of seed isoflavone content in Chinese soybean landrace population for optimal cross design and gene system exploration. Theor. Appl. Genet. 129: 1557–1576.

Mukund, V., D. Mukund, V. Sharma, M. Mannarapu and A. Alam (2017) Genistein: Its role in metabolic diseases and cancer. Crit. Rev. Oncol. Hematol. 119: 13–22.

Murashige, T. and F. Skoog (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol. Plant. 15: 473–497.

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

Nakayama, T., S. Takahashi and T. Waki (2019) Formation of flavonoid metabolons: Functional significance of protein-protein interactions and impact on flavonoid chemodiversity. Front. Plant Sci. 10: 821.

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.

Ralston, L., S. Subramanian, M. Matsuno and O. Yu (2005) Partial reconstruction of flavonoid and isoflavonoid biosynthesis in yeast using soybean type I and Type II chalcone isomerases. Plant Physiol. 137: 1375–1388.

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.

Robinson, J.T., H. Thorvaldsdóttir, A.M. Wenger, A. Zehir and J.P. Mesirov (2017) Variant review with the integrative genomics viewer. Cancer Res. 77: e31–e34.

Schmittgen, T.D. and K.J. Livak (2008) Analyzing real-time PCR data by the comparative C(T) method. Nat. Protoc. 3: 1101–1108.

Schmutz, J., S.B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D.L. Hyten, Q. Song, J.J. Thelen, J. Cheng *et al.* (2010) Genome sequence of the palaeopolyploid soybean. Nature 463: 178–183.

Sepiol, C.J., J. Yu and S. Dhaubhadel (2017) Genome-wide identification of chalcone reductase gene family in soybean: Insight into root-specific *GmCHRs* and *Phytophthora sojae* resistance. Front.

Plant Sci. 8: 2073.

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.

Subramanian, S., G. Stacey and O. Yu (2006) Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. Plant J. 48: 261–273.

Takahashi, R., S.M. Githiri, K. Hatayama, E.G. Dubouzet, N. Shimada, T. Aoki, S. Ayabe, T. Iwashina, K. Toda and H. Matsumura (2006) A single-base deletion in soybean flavonol synthase gene is associated with magenta flower color. Plant Mol. Biol. 63: 125–135.

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

Toda, K., D. Yang, N. Yamanaka, S. Watanabe, K. Harada and R. Takahashi (2002) A single-base deletion in soybean flavonoid 3′-hydroxylase gene is associated with gray pubescence color. Plant Mol. Biol. 50: 187–196.

Tsuda, M., A. Kaga, T. Anai, T. Shimizu, T. Sayama, K. Takagi, K. Machita, S. Watanabe, M. Nishimura, N. Yamada *et al.* (2015) Construction of a high-density mutant library in soybean and development of a mutant retrieval method using amplicon sequencing. BMC Genomics 16: 1014.

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.

Tuteja, J.H., G. Zabala, K. Varala, M. Hudson and L.O. Vodkin (2009) Endogenous, tissue-specific short interfering RNAs silence the chalcone synthase gene family in *glycine max* seed coats. Plant Cell 21: 3063–3077.

Urbański, D.F., A. Małolepszy, J. Stougaard and S.U. Andersen (2013) High-throughput and targeted genotyping of *Lotus japonicus* LORE1 insertion mutants. Methods Mol. Biol. 1069: 119–146.

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, N., T. Long, W. Yao, L. Xiong, Q. Zhang and C. Wu (2013) Mutant resources for the functional analysis of the rice genome. Mol. Plant 6: 596–604.

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., R. Hideshima, Z. Xia, Y. Tsubokura, S. Sato, Y. Nakamoto, N. Yamanaka, R. Takahashi, M. Ishimoto, T. Anai *et al.* (2009) Map-based cloning of the gene associated with the soybean maturity locus *E3*. Genetics 182: 1251–1262.

Watanabe, S., Z. Xia, R. Hideshima, Y. Tsubokura, S. Sato, N. Yamanaka, R. Takahashi, T. Anai, S. Tabata, K. Kitamura *et al.* (2011) A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. Genetics 188: 395–407.

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.

Watanabe, S., R. Yamada, H. Kanetake, A. Kaga and T. Anai (2019) Identification and characterization of a major QTL underlying soybean isoflavone malonylglycitin content. Breed. Sci. 69: 564–572.

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.

Welle, R., G. Schröder, E. Schiltz, H. Grisebach and J. Schröder (1991) Induced plant responses to pathogen attack. Analysis and heterologous expression of the key enzyme in the biosynthesis of phytoalexins in soybean (*Glycine max* L. Merr. cv. Harosoy 63). Eur. J. Biochem. 196: 423–430.

Xia, Z., S. Watanabe, T. Yamada, Y. Tsubokura, H. Nakashima, H. Zhai, T. Anai, S. Sato, T. Yamazaki, S. Lü *et al.* (2012) Positional cloning and characterization reveal the molecular basis for soybean maturity locus *E1* that regulates photoperiodic flowering. Proc. Natl. Acad. Sci. USA 109: E2155–E2164.

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.

Yamanaka, N., S. Watanabe, K. Toda, M. Hayashi, H. Fuchigami, R. Takahashi and K. Harada (2005) Fine mapping of the *FT1* locus for soybean flowering time using a residual heterozygous line derived from a recombinant inbred line. Theor. Appl. Genet. 110: 634–639.

Yan, F., S. Di, F. Rojas Rodas, T. Rodriguez Torrico, Y. Murai, T. Iwashina, T. Anai and R. Takahashi (2014) Allelic variation of soybean flower color gene *W4* encoding dihydroflavonol 4-reductase 2. BMC Plant Biol. 14: 58.

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.

Zabala, G. and L.O. Vodkin (2005) The *wp* mutation of *Glycine max* carries a gene-fragment-rich transposon of the CACTA superfamily. Plant Cell 17: 2619–2632.

Zabala, G. and L.O. Vodkin (2007) A rearrangement resulting in small tandem repeats in the *F3′5′H* gene of white flower genotypes is associated with the soybean *W1* locus. Crop Sci. 47: S-113–S-124.

Zaheer, K. and M.H. Akhtar (2017) An updated review of dietary isoflavones: Nutrition, processing, bioavailability and impacts on human health. Crit. Rev. Food Sci. Nutr. 57: 1280–1293.

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.