Theor Appl Genet (2011) 123:1375–1385
DOI 10.1007/s00122-011-1673-x

ORIGINAL PAPER

Major locus and other novel additive and epistatic loci involved in modulation of isoflavone concentration in soybean seeds

Juan J. Gutierrez-Gonzalez · Tri D. Vuong · Rui Zhong · Oliver Yu · Jeong-Dong Lee · Grover Shannon · Mark Ellersieck · Henry T. Nguyen · David A. Sleper

Received: 26 February 2011 / Accepted: 26 July 2011 / Published online: 18 August 2011

Abstract Seeds of soybean [*Glycine max* (L.) Merr.] accumulate more isoflavones than any tissue of any plant species. In other plant parts, isoflavones are usually released to counteract the effects of various biotic and abiotic stresses. Because of the benefits to the plant and positive implications that consumption may have on human health, increasing isoflavones is a goal of many soybean breeding programs. However, altering isoflavone levels through marker-assisted selection (MAS) has been impractical due to the small and often environmentally variable contributions that each individual quantitative trait locus (QTL) has on total isoflavones. In this study, we developed a Magellan × PI 437654 F_7-RIL population to construct a highly saturated non-redundant linkage map that encompassed 451 SNP and SSR molecular markers and used it to locate genomic regions that govern accumulation of isoflavones in the seeds of soybean. Five QTLs were found that contribute to the concentration of isoflavones, having single or multiple additive effects on isoflavone component traits. We also validated a major locus which alone accounted for up to 10% of the phenotypic variance for glycitein, and 35–37% for genistein, daidzein and the sum of all three soybean isoflavones. This QTL was consistently associated with increased concentration of isoflavones across different locations, years and crosses. It was the most important QTL in terms of net increased amounts of all isoflavone forms. Our results suggest that this locus would be an excellent candidate to target for MAS. Also, several minor QTLs were identified that interacted in an additive-by-additive epistatic manner, to increase isoflavone concentration.

Communicated by R. Johnson.

This work is dedicated to the memory of Elaine Sleper.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-011-1673-x) contains supplementary material, which is available to authorized users.

J. J. Gutierrez-Gonzalez · T. D. Vuong · G. Shannon · H. T. Nguyen · D. A. Sleper
Division of Plant Sciences and National Center for Soybean Biotechnology, University of Missouri, Columbia, MO 65211, USA

Present Address:
J. J. Gutierrez-Gonzalez
USDA-ARS Plant Science Research Unit and Department of Agronomy and Plant Genetics, University of Minnesota, St Paul, MN 55108, USA

R. Zhong · O. Yu
Donald Danforth Plant Science Center, 975 North Warson Road, Saint Louis, MO 63132, USA

J.-D. Lee
Division of Plant Biosciences, Kyungpook National University, Daegu 702-701, Republic of Korea

M. Ellersieck
Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211, USA

D. A. Sleper (✉)
Division of Plant Sciences and National Center for Soybean Biotechnology, Life Sciences Building, University of Missouri, Columbia, MO 65201, USA
e-mail: Sleperd@missouri.edu

Introduction

Isoflavones are released and concentrated in areas exposed to pathogens and distinct abiotic stresses (Graham et al.

2007; Moy et al. 2004; Subramanian et al. 2005). In addition, isoflavones have garnered much attention because of their potential use as antioxidants and in pharmaceuticals to prevent important human diseases such as cancer (Rochfort and Panozzo 2007; Usui 2006). Nevertheless, the effect that isoflavone consumption may have on humans appears to depend on factors such as racial group, sex and age. Thus, low-isoflavone lines may be more appropriate for certain human subgroups, for example infants and pregnant women (for review see Zhang and Yu 2009). Consequently, there might be value for developing cultivars with varying isoflavone concentrations and to understand the genetics of their regulation of synthesis and accumulation. Recently, loci and interactions among loci that affect isoflavone accumulation in soybean seeds were disclosed (Gutierrez-Gonzalez et al. 2009, 2010b). At least four other studies have reported on the genetic regions underlying isoflavone accumulation in soybean seeds (Kassem et al. 2006; Meksem et al. 2001; Primomo et al. 2005; Zheng et al. 2009). Despite the many loci reported, the presence of a major quantitative trait locus (QTL) that consistently affects isoflavone content across environments is yet to be identified.

Marker-assisted selection (MAS) is broadly used in crop plant breeding programs, because it can greatly accelerate development of new improved varieties as compared to phenotypic selection. Nevertheless, the use of MAS requires knowledge of reliable marker–trait associations that are relatively constant over environments. In the case of isoflavones, finding a stable QTL has been hampered, because isoflavone accumulation in soybean seeds is mainly governed by minor-effect QTLs that are often influenced by the environment. Although a linkage map is not strictly required for MAS, a dense marker genetic map greatly facilitates strong marker–gene correlations by permitting the utilization of improved QTL mapping approaches. Composite interval mapping (Zeng 1993; 1994) and multiple interval mapping (Kao et al. 1999; Zeng et al. 2000) have proved to be more accurate in determining QTL locations and effects than simple marker–trait associations (Kao et al. 1999). Markers utilized by soybean researchers include simple sequence repeats (SSR), restriction fragment length polymorphism (RFLP) and amplified fragment length polymorphisms (AFLP). Recently, single nucleotide polymorphisms (SNPs) have increasingly become available for the soybean genome, which because of their abundance open enormous opportunities to create high-density maps. Genotyping of segregating mapping populations is labor intensive and consequently has been a limiting factor in QTL studies. However, with the advent of efficient high-throughput techniques, comprehensive genome-wide marker studies can be conducted at a reasonable cost. One high-throughput method, involving multiplexing SNP assays using the Illumina GoldenGate system, has proven to be particularly effective even in discriminating paralogous genes in the complex soybean duplicated genome (Hyten et al. 2008).

Crop improvement is undeniably dependent on genetic variability. Regrettably, genetic diversity is reduced in soybean due to the intensive use of recurrent germplasm in breeding programs, which limits the introduction of new and desirable genes. The use of plant introductions (PIs) within plant species that have been introduced to areas outside their natural origin allows for the discovery of new sources of alleles that have been lost in the process of domestication and development of commercial varieties. This allele loss can be especially significant for traits previously not considered important in breeding programs. However, when the need arises, PIs can be used as a rich source of genetic variability for traits of emerging importance. One such soybean PI (PI 437654), originating from China and introduced to the USA from Russia, has up to fourfold more isoflavone concentration in comparison to commercial varieties (Gutierrez-Gonzalez et al. 2009), making it very suitable for isoflavone studies. In addition, this black seed-coated line is resistant to the soybean cyst nematode (*Heterodera glycines*) (SCN) races 1, 2, 3, 4, and 5 (Anand and Gallo 1984; Anand et al. 1988; Arelli et al. 1997) and has been used in the development of soybean varieties with broad resistance to SCN populations (Handoo and Anand 1993; Shannon et al. 2007; Diers et al. 2010).

QTL mapping and marker–trait associations are powerful tools for geneticists and plant breeders in marker-assisted breeding. An estimated 1,200 or more quantitative loci have been reported for 12 major crop species (Bernardo 2008). However, very few of them have been confirmed in different genetic backgrounds or for expression over environments. In this study, we developed a new soybean recombinant inbreed line (RIL) mapping population using two highly isoflavone contrasting lines: Magellan (low-isoflavone concentration) and PI 437654 (high-isoflavone concentration). We conducted a QTL study to: (i) confirm the higher seed isoflavone QTL from our previous study on the Essex × PI 437654 cross (Gutierrez-Gonzalez et al. 2010b), as well as check for consistency of previously identified QTL; and (ii) identify other novel QTLs for isoflavone concentration that may be useful in soybean breeding programs.

Materials and methods

Plant material and growing conditions

A recombinant inbred line (RIL) mapping population comprising 188 F_7 progeny was initially developed from a cross between the isoflavone contrasting varieties Magellan

(low-concentration) and PI437654 (high-concentration). These 188 RILs and parents were planted in 2007 in irrigated fields in two-row plots, 2-m long, with three replications. The experiment was conducted at two locations: the University of Missouri Bradford Research and Extension Center (BREC, 38°58′ N) and the University of Missouri Delta Research Center (DRC, 36° 44′ N). Detailed records of precipitation, temperatures and other climatological parameters during the growing period for both locations can be found at http://aes.missouri.edu/bradford/weather/ and http://aes.missouri.edu/delta/weather/, respectively. Seeds were harvested from mature plants, from a bulked sample, at least three plants from each plot and quantified for individual isoflavone concentration as detailed below.

Isoflavone extraction and quantification

Approximately 2.5 g (~20 seeds) of soybean seeds were ground to a fine powder using a commercial coffee grinder. The powder was extracted with 7 ml of 80% methanol at 55°C for 2 h, vortexing every 30 min. After centrifugation (5,000 rpm, 5 min), the supernatant was filtered using Fisherbrand 0.45-μm 25-mm nylon syringe filters (Fisher Scientific, Pittsburgh, PA, USA). Samples were analyzed by reverse-phase HPLC on an Agilent 1100 high-performance liquid chromatography (HPLC) system (Santa Clara, CA, USA). Separation and elution were accomplished using an 18-min linear gradient initiated with 20% methanol/80% 10 mM ammonium acetate (v/v) (pH 5.6) and completed with 100% methanol at a flow rate of 1 ml/min. An RP-C18 Lunar C2 column was used (Phenomix, La Jolla, CA, USA). Detection of metabolites was achieved by photodiode array. Identification and quantification of each isoflavone component were based on available standards (Indofine Chemical Co., Somerville, NJ, USA). Measurements are given as micrograms of isoflavone per gram of seeds plus/minus standard deviation (μg/g ± SD).

Molecular marker analysis

DNA extraction

DNA was extracted from young trifoliate leaves from each F_7 RIL using the AutoGenprep 960 robot following the manufacturer's instructions (AutoGen Inc., Holliston, MA, USA), which uses an automated adaptation of the CTAB method. Briefly, dried leaf tissues were ground to a fine powder with the GenoGrinder 2000 (BT&C Inc., Lebanon, NJ, USA). Approximately, 20 mg of the ground tissue was incubated in 1.5 Eppendorf tubes with 0.45 ml of CTAB at 65°C for 1:30 h. The tubes were spun at 5000 rpm for 5 min and 0.3 ml of the supernatant transferred to 96-deep well plates. The robot then completed the extraction by adding chloroform to the suspension followed by mixing and centrifugation. After RNase treatment of the aqueous layer, precipitated DNA was washed using 70% ethanol and dissolved in TE buffer, pH 8.0. DNA concentration and quality was measured using NanoDrop ND-1000 UV–Vis spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

Single nucleotide polymorphism (SNP) analysis

For SNP genotyping, a universal soybean linkage panel (USLP 1.0) containing 1,536 SNP markers (Hyten et al. 2010) was employed to genotype each RIL of the mapping population following an Illumina GoldenGate assay as described by Fan et al. (2006). All SNPs were placed onto the integrated molecular genetic linkage map as previously described (Hyten et al. 2010). Detailed specifications and protocols can be found in an earlier work (Vuong et al. 2010). Briefly, a multiple-step procedure was followed according to the manufacturer's instructions. This comprised genomic DNA activation by biotinylation, PCR amplification, oligonucleotide hybridization, washing and array imaging, performed using the Illumina BeadStation (Illumina, San Diego, CA, USA) to generate intensity data. Detection of each SNP for specific alleles was subsequently determined with the BeadStudio 3.0 software (Illumina, San Diego, CA, USA), based on the intensities detected. The clusters of homozygous and heterozygous genotypes for each SNP were manually checked for polymorphisms between the two parental lines. The 711 SNP loci found polymorphic were subsequently employed for genetic linkage map construction. Out of them, 446 were finally mapped and used for QTL mapping analysis.

Fluorescent simple sequence repeat (SSR) markers

For the PCR reaction, 20 ng/μl of template DNA was incubated with 1 × reaction buffer (20 mM Tris–HCl, 50 mM KCl, pH 8.0), 0.2 and 0.13 μM reverse primer (Applied BioSystems, Foster City, CA, USA) and fluorescently labeled forward primer (IDT Inc., Coralville, IA, USA), respectively, 2.5 mM $MgCl_2$, 0.2 mM of each dNTPs and 1U of *Taq* DNA polymerase (GenScript Corp., Piscataway, NJ, USA), in a final volume of 12.5 μl. Thermal cycler settings were 95°C for 5 min, followed by 5 min at 94°C (denaturation), 45 s at 47 or 52°C (annealing) and 1 min at 72°C (extension), repeated 35 times (cycles) and 7 min at 72°C after the last cycle. The resulting amplified and labeled DNA samples were separated and analyzed using the ABI 3100 DNA sequencer (Applied BioSystems, Foster City, CA, USA). Allele calling was accomplished with the GeneMapper v.3.7

software (Applied BioSystems, Foster City, CA, USA). A total of 20 SSR markers were initially chosen for their genomic localization as candidates to fulfill gaps on the SNP-base map, 11 of which were polymorphic and 5 were finally mapped.

Statistical, linkage and QTL analysis

Environments (e), replications (environments), genotypes (g), and environments × genotypes (g × e) interaction effects were assumed to be random in an analysis of variance using PROC GLM in SAS STAT 9.1 (SAS Institute Inc., Cary, NC). Variance components to calculate heritability estimates were determined using PROC Mixed model, considering g, e and g × e as random effects. Heritability in the broad sense was calculated according to Hill et al. (1998):

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{E} + \frac{\sigma_\varepsilon^2}{ER}}$$

with σ_g^2, σ_{ge}^2 and σ_ε^2 the genotype, genotype × environment, and error components of variance, respectively; E the number of environments; and R the number of replications.

After eliminating redundant markers, assessment of linkage between marker pairs and marker groupings were performed by means of two software packages: JoinMap v.3.0 (van Ooijen and Voorrips 2001) and Carthagene v.1.0 (de Givry et al. 2005) using the initial threshold parameters of minimum two-point LOD score of 9.0 and a maximum two-point distance of 30 cM. Marker ordering and spacing in each linkage group (LG) was accomplished in an iterative process with the algorithm *buildfw* of Carthagene v.1.0, which uses an incremental insertion method that allows constructing robust framework maps. The distance function used was *Kosambi*. The quality of the maps was further checked with the *annealing* algorithm in the Carthagene package. Marker order was compared to the integrated genetic linkage map of soybean (Hyten et al. 2010) and with the recently released whole genome sequence Glyma1 assembly for Williams 82 (Schmutz et al. 2010), as that information became available. Markers in conflict with either were removed before each iteration. The final version of the linkage map contained a total of 451 SNP and SSR markers, which were distributed on 24 linkage groups (LG) covering 2,152 cM, with an average distance between markers of 4.8 cM.

Mixed models can improve the statistical power for QTL detection by performing an integrated analysis with data from each replicate, which includes different locations and within-location replications (Wang et al. 1999). QTLNetwork v2.0 (Yang et al. 2007) was used to perform mixed model-based composite interval mapping for the QTL mapping analysis with random environments and replications (environments), and run with the two locations input data and three replications per location all combined in a 2 × 3 scheme. In addition, data were run separately for each environment for comparison. Critical F values were assessed by permutation test using 1,000 permutations rendering 9.0, 9.0, 8.0 and 8.9 for genistein, daidzein, glycitein and total isoflavones, respectively. Individual replicate data and the two environments together were used to run the QTLNetwork. QTL effects were estimated using Markov Chain Monte Carlo (MCMC) method. Candidate interval selection, epistatic effects and putative QTL detection were calculated with an experimental-wise Type I error of $\alpha = 0.05$, $\alpha = 0.05$ (and $\alpha = 0.001$) and $\alpha = 0.05$ (and $\alpha = 0.001$), respectively. Genome scan was performed using a 10-cM window size and 1-cM walk speed.

Results

Genetic and phenotypic variation within mapping population

Parental lines Magellan and PI 437654 that consistently contrasted in isoflavone concentrations across different environments (Gutierrez-Gonzalez et al. 2009) were crossed to develop a recombinant inbred line (RIL) mapping population. In the present study, the average seed isoflavone across two field trials was threefold higher for PI 437654 (1880 ± 253 μg/g, mean ± SD) than for Magellan (603 ± 103 μg/g). Despite the low-isoflavone concentration of Magellan, transgressive segregation was observed among RILs (Fig. 1 and Supplemental Fig. 1), suggesting that both parental lines possess positive-effect alleles. In both locations, a remarkable range of variation was evident for each individual isoflavone among RILs of the mapping population. For instance, seed genistein concentration varied between 58 and 1297 μg/g, and daidzein ranged between 49 and 2013 μg/g (Table 1). The analysis of variance (ANOVA) conducted over environments (Table 1) determined the variation in the accumulation of individual and total soybean seed isoflavone concentrations as affected by genetic (G), environmental (E) and G × E interaction effects ($P < 0.0001$, for all the effects). Heritability in the broad sense (H^2) was high for genistein, daidzein and total isoflavones (0.90–0.95), and intermediate for glycitein (0.62), which were similar to values reported by Gutierrez-Gonzalez et al. (2009, 2010b). Only the distribution of glycitein at the Delta Research Center (DRC) appeared to be non-normal.

Black seed color may represent a barrier to acceptance in commercial channels, and since there may be some

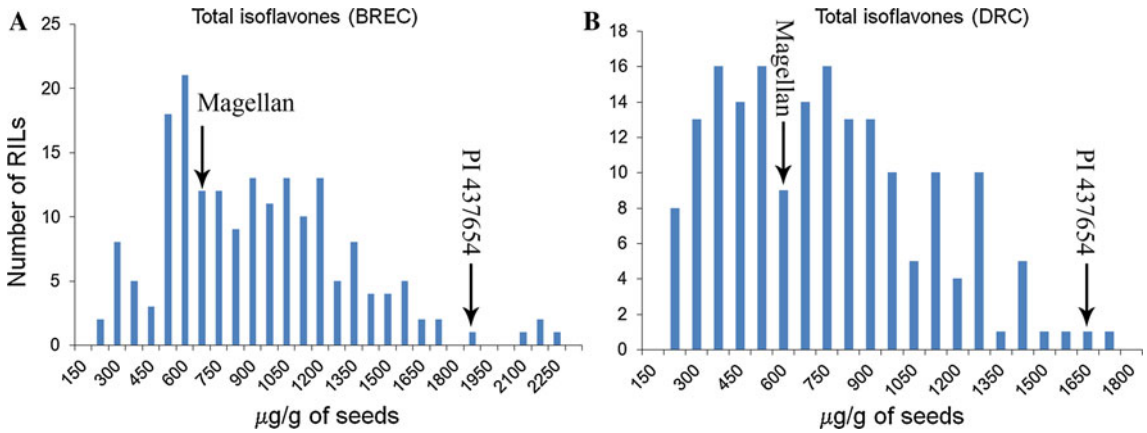

Fig. 1 Distribution of total isoflavone concentrations from the analysis of 188 RILs of the Magellan × PI 437654 population across three replications in each of two field locations, BREC (**a**) and DRC (**b**), in 2007. *Arrows* indicate the position of the two parental lines. *Horizontal axis* shows seed isoflavone concentration in μg/g

Table 1 ANOVA, effect mean squares and heritability of genistein, daidzein, glycitein and total isoflavone concentration analyzed over two environments for 188 RILs from Magellan × PI 437654

Trait	Mean ± SD[a]	Range	CV[b](%)	Effect mean squares[c]				H^{2d}	Skewness	Kurtosis
				V(G)	V(E)	V(GE)	V(ε)			
Genistein	409 ± 202	58–1297	49.4	173114	1748824	18261.7	5964.3	0.90	0.70	0.37
Daidzein	385 ± 212	49–2013	55.1	189042	2129377	19026.8	7061.3	0.95	1.30	4.49
Glycitein	20 ± 13	0–277	65.5	286.3	374.7	323	103.5	0.62	9.97	171.32
Total Isoflav	814 ± 413	124–3174	50.8	727045	7867853	71600.5	24486.7	0.95	0.94	1.71

[a] Mean ± standard deviation (SD) and range units in micrograms of isoflavones per gram of seeds

[b] Coefficient of variation within the segregating mapping population over environments

[c] Variance of genetic V(G), environmental V(E), genotype-by-environment interaction V(GE), and residual V(e) effects

[d] Heritability of the traits in the broad sense over environments

reason to believe that black-seeded soybeans are characterized by having higher isoflavone concentrations than yellow varieties, segregation in both traits was compared. No co-segregation was found between isoflavone concentrations and seed color in the mapping population (correlation coefficient $r = 0.045$, 0.094, 0.039, 0.071 for genistein, daidzein, glycitein and total isoflavones, respectively).

Molecular recombination mapping

The universal soybean linkage panel 1.0 (the USLP 1.0), the base of the GoldenGate platform used, has 1,536 SNPs uniformly distributed across the soybean genome and are mapped onto the integrated molecular linkage map (Hyten et al. 2010). In addition to these SNPs, another 20 SSR markers were assayed for polymorphism between the parental lines to fill the gaps created in the initial drafts of the map. A total of 842 SNPs and 11 SSRs, respectively, were polymorphic between the parents and subsequently used on the segregating population and for linkage analysis. Based on the latest version of the consensus map recently published by Hyten et al. (2010), the addition of the 11 SSR markers enabled us to join groups of SNP markers on the same chromosome. Out of the total 853 polymorphic markers assayed for linkage, 451 (53%) were ultimately mapped at the established threshold levels (446 SNPs and 5 SSR) and placed in 24 linkage groups (LG). Markers that exceeded 15% of missing observations were removed prior to the analysis (8%). In addition, we eliminated markers that were redundant or mapped to the same location (19%), and markers that exhibited segregation distortion ($P < 0.001$) (2%), to build a robust framework linkage map that would allow more accurate QTL mapping. As much as 14% of the initial markers remained unlinked or linked in short LGs. The final molecular linkage map total length was 2152 cM, with an average distance between markers of 4.8 cM and estimated genome coverage of 90.7%. The complete map is displayed in Supplemental Fig. 3.

Accuracy of the molecular map was iteratively checked by three methods. First, pairwise recombination fractions

across the genome were calculated with the R/qtl package (Broman et al. 2003) (Supplemental Fig. 2). Graphical representation of pairwise recombination is a quick, visual way to detect possible genotyping errors. The presence of a well-defined diagonal, in red, indicates that consecutive markers have the highest log of odds (LOD) ratio or the smallest recombination fraction, thereby suggesting a correct linkage among markers. Second, marker order was compared to the integrated genetic linkage map of soybean (Hyten et al. 2010). Third, when the information was available, marker order was also compared with the recently released whole genome sequence Glyma1 assembly for Williams 82 (Schmutz et al. 2010), through the generic genome browser on http://www.soybase.org. Markers that contributed to conflicts among results from the three methods were removed prior to the next round of linkage analysis. Interactions were continued until a stable marker order was found. A total of 3% of markers were discarded through this quality filter. A map with consistent order, as compared to the integrated map, was obtained except for one of the telomeric regions of the linkage group Gm05, comprising the region between the SNP markers BARC-060051-16321 and BARC-059035-15581 (Fig. 2a). According to our linkage analysis, this region appears inverted relative to the integrated consensus map and the soybean genome sequence.

QTL main effects and interactions

Isoflavones particularly fit mixed linear models because of their large variation over environments (Gutierrez-Gonzalez et al. 2009, 2010b). These models are especially appropriate for RIL-mapping populations because they allow testing the same genetic material in varying environments. The practicality of this approach is detection of an increased number of minor QTLs, genetic effect interactions and genetic interactions with the environment. A total of five QTLs were identified on four linkage groups with a main effect on one or more of the traits (genistein, daidzein, glycitein or total isoflavone seed concentrations), with a genome-wise type I error of $\alpha = 0.001$ (Table 2 and Fig. 2a).

One of those is a major QTL found on Gm05 (*Glycine max* chromosome 5), on former linkage group (LG) A1, which had a significant effect on each individual isoflavone plus total seed isoflavone accumulation (named *qGEN5*, *qDAI5*, *qGLY5* and *qTOT5*, respectively). This locus was previously reported to affect the four traits in a different set of RILs from the cross Essex × PI 437654 (Gutierrez-Gonzalez et al. 2009, 2010b). However, the confidence interval of the QTL was significantly reduced in this present work to 3.7 cM (approximately 1.3 Mb). Importantly, the additive component of this locus alone was able to explain around one-third the percentage of the phenotypic variance, $h^2(a)$, found for most of the individual isoflavone forms. That is, for genistein, daidzein and glycitein, the locus accounted for about 36, 35 and 12% of that variance, respectively, and for 37% when the sum of all isoflavones was considered. In addition, this major locus on chromosome 5 displayed a very significant additive-by-environment (A × E) interaction effect, which was found to influence glycitein amounts. Contribution of this interaction to the proportion of the variance explained for glycitein was substantial ($h_{A\times E} = 11\%$) with an estimated effect value of 4 ± 0.7 (μg/g ± SD). The magnitude of the variance explained by the interaction was similar to that attributable to the additive effects alone. Overall, our results strongly support the presence of a major QTL on Gm05 that controls the accumulation of all known isoflavones in soybean seeds. We not only validate this QTL in a second population, but also report a larger stable contribution than ever described for any isoflavone form or for total isoflavone. On that same chromosome (Gm05), a minor locus (*qGLY5_2*) explained glycitein amounts ($\alpha = 0.05$, $h^2(a) = 2\%$).

Another locus was also found to influence daidzein and total isoflavone (*qDAI6*, and *qTOT6*, respectively) with a significance level of $\alpha = 0.001$. This locus was located on Gm06 (LG C2) and accounted for 8 and 5% of the phenotypic variance of daidzein, and total isoflavone concentrations, respectively. A total of 4% of the variance in genistein was also attributable to this QTL (*qGEN6*). However, the significance was only at $\alpha = 0.05$. In addition, two more loci affecting isoflavone levels were detected. The first identified locus, on Gm11, affected all four traits examined: *qGEN11*, *qDAI11*, *qGLY11* and *qTOT11*, explaining 7% of the phenotypic variance for genistein, daidzein, and total isoflavones, and 3% for glycitein. The second was a minor-effect QTL on Gm08 that accounted for 2% of both genistein (*qGEN8*) and total isoflavones (*qTOT8*) variation.

Additive-by-additive epistatic interactions were found for several individual and total isoflavones (Table 3). Four of the aforementioned additive-effect loci were also involved in pairwise epistatic interactions, which appears to reinforce the important role of these QTLs affecting isoflavones. The first of these interactions influenced genistein, between the loci BARC-063661-18416-BARC-066175-19800 and BARC-042999-08498-BARC-016279-02316 (*qGEN6–qGEN11*), daidzein between the intervals BARC-040475-07751-BARC-023517-05442 and BARC-016279-02316 (*qDAI6–qDAI11*), and total isoflavones between BARC-063661-18416-BARC-066175-19800 and BARC-042999-08498-BARC-016279-02316 (*qTOT6–qTOT11*). The percentage of the variation that this epistatic effect accounted for ranged between 1 (genistein)

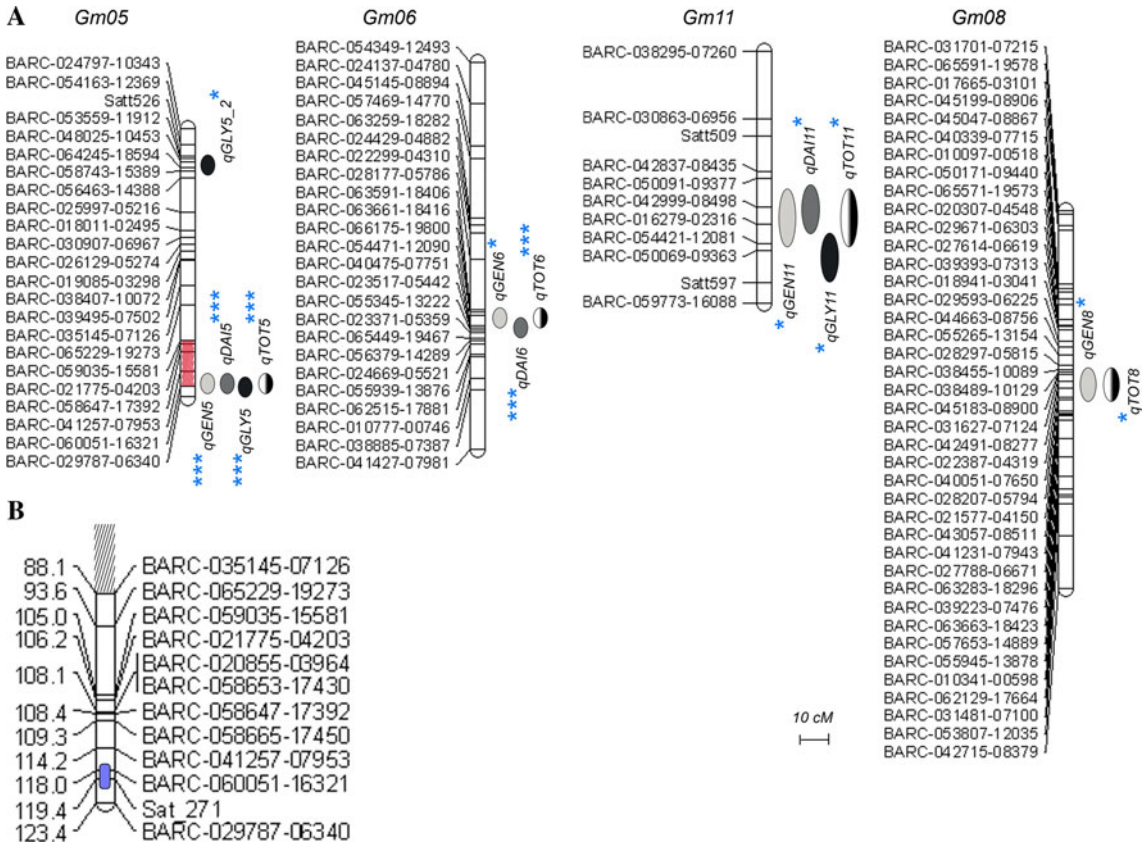

Fig. 2 **a** A linkage map constructed from the analysis of 188 RILs from the Magellan × PI 437654 cross. SNP markers are shown on the *left side* and isoflavone QTL are depicted at the *right side* of each linkage group. *Light gray ovals*, *dark gray ovals*, *hatched ovals* and *black ovals* indicate loci associated with genistein, daidzein, glycitein and total isoflavone, respectively. The name of the QTL, shown near each oval, is a composite of the influenced trait: genistein (*gen*), daidzein (*dai*), glycitein (*gly*) and total of isoflavones (*tot*), followed by the chromosome number. Hatched area in chromosome Gm05 highlights the inverted region. *Asterisks* indicate the level of significance at which the QTL was detected: *0.05 and ***0.001. **b** Details of the Gm05 telomeric region with the addition of four more markers. *A rectangle* delimits the major QTL interval of confidence, centered on the SNP marker BARC-060051-16321

and 2% (daidzein). Another important epistatic interaction influenced genistein concentrations (*qGEN5–qGEN8* at intervals BARC-041257-07953-BARC-060051-16321 and BARC-031627-07124-BARC-042491-08277) and total isoflavone concentration (*qGEN5–qGEN8* at intervals BARC-041257-07953-BARC-060051-16321 and BARC-042491-08277-BARC-022387-04319). The last epistatic interaction, contributing 2% to the variability of daidzein, was detected between two non-additive loci in chromosomes Gm02 and Gm10, between the intervals BARC-032525-08992-BARC-018187-02537 and BARC-063079-18232-BARC-041457-07999.

Discussion

The influence of genes and environment on isoflavone synthesis has been widely reported (Caldwell et al. 2005; Eldridge and Kwolek 1983; Gutierrez-Gonzalez et al. 2009, 2010a, 2010b; Hoeck et al. 2000; Lozovaya et al. 2005; Mebrahtu et al. 2004; Tsukamoto et al. 1995; Wang and Murphy 1994). The external factors influencing isoflavones are so numerous that even subtle changes at the micro-environmental level could lead to significant differences in isoflavone concentration. The bell-shaped distribution of frequencies (Fig. 1 and Supplemental Fig. 1) and the normality-associated parameters, skewness and kurtosis on Table 1, confirm that these traits are inherited in a quantitative manner, with many genes exerting small individual effects. Nevertheless, concentrations of each isoflavone component and total isoflavone in PI 437654 ranked at the 90% percentile. This indicates that the PI parent of the population had the majority of positive alleles for the QTL. Our results support the contention that PI 437654 is the major contributor to increased isoflavones since cv. Magellan had only one allele at *qGEN8* and *qTOT8* that increased genistein and total isoflavone concentrations. There was little variation in glycitein concentration at one of the locations (DRC), which suggests that the expression of variation was masked in this particular

Table 2 Additive effects found for the five QTLs, ordered by trait: genistein, daidzein, glycitein and total isoflavone concentration in soybean seeds

Effects[a]	QTL	Interval	IC[b]	A ± SE[c]	$\alpha = 0.001$		$\alpha = 0.05$	
					[d]*F* value	$h^2(a)$	*F* value	$h^2(a)$
Genistein								
qGEN5	*qGm05*	BARC-041257-07953-BARC-060051-16321	80.0–83.7	120 ± 8	98.4	36	93.5	36
qGEN6	*qGm06*	BARC-063661-18416-BARC-066175-19800	79.2–84.0	42 ± 7			12.6	4
qGEN8	*qGm08*	BARC-031627-07124-BARC-042491-08277	51.7–58.9	−34 ± 7			12.1	2
qGEN11	*qGm11*	BARC-042999-08498-BARC-016279-02316	45.6–61.2	32 ± 7			10.7	7
Daidzein								
qDAI5	*qGm05*	BARC-041257-07953-BARC-060051-16321	80.0–83.7	123 ± 8	103.2	35	112.2	35
qDAI6	*qGm06*	BARC-040475-07751-BARC-023517-05442	84.5–86.2	56 ± 8	21.8	8	23.9	8
qDAI11	*qGm11*	BARC-042999-08498-BARC-016279-02316	44.6–58.2	31 ± 8			9.7	7
Glycitein								
qGLY5	*qGm05*	BARC-060051-16321-BARC-029787-06340	81.0–84.7	4 ± 1	57.0	12	57.0	12
qGLY5_2	*qGm05_2*	BARC-048025-10453-BARC-064245-18594	6.2–26.8	1 ± 1			9.5	2
qGLY11	*qGm11*	BARC-054421-12081-BARC-050069-09363	59.2–72.9	1 ± 1			8.7	3
Total								
qTOT5	*qGm05*	BARC-041257-07953-BARC-060051-16321	80.0–83.7	244 ± 16	105.1	37	97.5	37
qTOT6	*qGm06*	BARC-063661-18416-BARC-066175-19800	79.2–84.5	105 ± 15	17.2	5	17.2	5
qTOT8	*qGm08*	BARC-042491-08277-BARC-022387-04319	51.7–59.9	−61 ± 15			9.8	2
qTOT11	*qGm11*	BARC-042999-08498-BARC-016279-02316	45.6–60.2	63 ± 15			10.9	7

Results based on the analysis of 188 RILs derived from the cross Magellan × PI 437654, mapped on the framework map

[a] Name given to a particular effect of each QTL, *DAI*, *GEN*, *GLY* and *TOT* for genistein, daidzein, glycitein and total isoflavone concentration, respectively, followed by the chromosome number and a number when more than one in the same chromosome

[b] Interval of confidence in centiMorgans with respect to the first marker in the LG

[c] Main additive effect in μg/g plus/minus standard error. Mean effect of substituting both Magellan alleles by PI437654 alleles. Thus, positive values indicate that the PI437654 allele increases the phenotypic value. $h(a)^2$ is the heritability of the additive effect or percentage of variation that is explained by the additive component of the QTL

[d] *F* values of significance of each QTL. Threshold *F* values were 9.0, 9.0, 8.0 and 8.9 for genistein, daidzein, glycitein and total isoflavones, respectively, for the two-environment combined analysis

environment. Effectively at this location, significant variation for only *qGLY5* was detected ($p = 0.001$) (data not shown). Heritability in the broad sense (H^2) exhibited high numbers (>90%) for genistein, daidzein and total isoflavones (Table 1), indicating that despite the strong impact environmental effects may have on isoflavone concentration, the effect is fairly uniform across all genotypes and that genotype × environment interaction does not seriously compromise genotypic main effects, making progress from selection feasible.

The use of a reliable high-density marker framework map allowed us not only to increase the precision of locating the QTL, but also to narrow down confidence intervals. The comprehensive consensus molecular marker map (Hyten et al. 2010) is undoubtedly a valuable tool to assist in the construction of robust population-specific linkage maps. Many of the SNPs included in the universal soybean linkage panel 1.0 (the USLP 1.0), the platform used in our study, have been also located on the soybean whole genome sequence, which provides additional authenticity to marker order. Our final version of the linkage framework map had a total of 451 markers (SNPs and SSR), which were distributed on 24 linkage groups covering 2,152 cM, with an average distance between markers of 4.8 cM. Marker order was in full agreement with the integrated soybean genetic linkage map (Hyten et al. 2010) and with the recently released whole genome sequence, except for one of the telomeric regions on chromosome Gm05. This distal part of Gm05, shown as inverted with regard to the other sources, spanned 14.6 cM, between the SNP markers BARC-060051-16321 and BARC-059035-15581. Notably, the interval of confidence of the major QTL described above appears to cover one of the boundaries of this inverted region (Fig. 2a). This major QTL was also detected in a different recombinant map that was constructed from a cross that shared the PI 437654 parent (Gutierrez-Gonzalez et al. 2010b), where the authors also reported the same misaligned region, as compared to the integrative map. Nevertheless, most likely this inverted region found in both maps does not correspond to a real

Table 3 Additive-by-additive epistatic interactions for genistein, daidzein, glycitein and total isoflavone concentration from analyses of 188 RILs from Magellan × PI437654, mapped on the framework map

Interval_i[a]	Interval_j[a]	AA ± SE[b]	*P* value	h^2(aa)[c]
Genistein				
BARC-041257-07953-BARC-060051-16321	BARC-031627-07124-BARC-042491-08277	−29 ± 8	0.000112	2
BARC-063661-18416-BARC-066175-19800	BARC-042999-08498-BARC-016279-02316	21 ± 8	0.005752	1
Daidzein				
BARC-040475-07751-BARC-023517-05442	BARC-042999-08498-BARC-016279-02316	28 ± 8	0.000411	2
BARC-032525-08992-BARC-018187-02537	BARC-063079-18232-BARC-041457-07999	26 ± 8	0.000448	2
Total isoflavones				
BARC-041257-07953-BARC-060051-16321	BARC-042491-08277-BARC-022387-04319	−58 ± 15	0.000139	2
BARC-063661-18416-BARC-066175-19800	BARC-042999-08498-BARC-016279-02316	43 ± 15	0.004530	1

P values represent the significance of each effect

[a] Marker interval within each epistatic interaction occurs

[b] Estimated additive-by-additive effect in μg/g plus/minus standard error; a positive sign for epistatic effects indicates that parental allele combinations at the two loci involved in epistasis increase total concentration of isoflavones, while a negative sign indicates that recombinant allele combinations increase total isoflavone values

[c] Heritability of the additive-by-additive interaction effect (%). *P* values represent the significance of each effect

genomic inversion, because if there was a real inversion then the three parental lines would have to have the inversion to be detected on the genetic maps. This is unlikely to be the case due to the different crosses and origins of those lines; it could be just the reflection of a mis-assembly in the consensus map and the current version of the Williams 82 whole genome sequence (Dr David Hyten, personal communication). An improved consensus map is currently being developed with the addition of two large populations with about 900 RILs and more than 17,000 markers, which would provide greater resolution and help to resolve the Gm05 telomeric region (Dr. Perry Cregan, personal communication).

To help in clarifying the order of markers on this telomeric region, framework-map conditions were slightly relaxed to include markers with up to 20% missing observations. Four more markers were added to the region, further reducing the interval of confidence to 3.3 cM (Fig. 2b and Supplemental Fig. 4). The *qGm05* was placed between the markers BARC-060051-16321 and Sat_271 (BARCSOYSSR_05_1260) (Supplemental Fig. 4). Molecular breeders are continuously in search of significant marker–QTL associations that would allow successful implementation of MAS techniques. Nevertheless, MAS approaches have been difficult to apply in the case of isoflavones, because individual isoflavone QTLs have small additive effects which in many cases are also environmentally modulated. Validation and narrowing the interval of the major locus in Gm05 constitutes an important achievement in eliminating these obstacles. First, the QTL was responsible for the variation in the amounts of all individual soybean isoflavones. Second, the percentage of that variation was as high as 36, 35, 12 and 37% for genistein, daidzein, glycitein and total isoflavones, respectively, which, to our knowledge, is a QTL with the greatest consistent effect ever reported in soybean. Other large-effect loci have been described (Kassem et al. 2006; Primomo et al. 2005); however, either their effects were drastically diminished or not even detected when the location and/or year changed. Third, the Gm05 QTL was detected in at least two different crosses and evaluated over several environments. These observations strongly support the conception that the telomeric region of Gm05 bears a principal gene for isoflavone synthesis, most likely an undiscovered kinase-related gene or a trans-acting factor, because no known phenylpropanoid pathway gene is located in the vicinity of the QTL (Gutierrez-Gonzalez et al. 2010b). The confidence interval of the locus has now been considerably narrowed and centered on the SNP marker BARC-060051-16321. To facilitate selection of possible markers for MAS, computer-based identified potential SSR markers (Song et al. 2010) that lay within the QTL flanking markers were downloaded from Soybase (www.soybase.org) and summarized in Supplementary Table 1. Fine mapping of the region is needed to further reduce the confidence interval and number of postulated candidate genes. Although significant genotype-by-environment interactions (A × E interaction) of some of the flavones were detected in this study and by Gutierrez-Gonzalez et al. (2010b), the magnitude of the main effect (A) of *qGm05* suggests that the QTL should be quite useful in MAS.

Apart from the validation of this key locus, other novel additive effects that affect seed isoflavone accumulations

were uncovered in four different genomic regions. First, a second locus on Gm05 (*qGLY5_2*) was found to explain glycitein amounts, constituting a different effect from that found by Kassem et al. (2006) on daidzein. The locus on Gm06 (*qGEN6*, *qDAI6* and *qTOT6*) is the second in importance of the ones reported in this study, contributing up to 8% in the variance of the trait. These QTLs appear to overlap the *qGLY6_2* described earlier by our group (Gutierrez-Gonzalez et al. 2010b), however, while in the earlier report it was found to affect only glycitein amounts. This study was involved in controlling genistein, daidzein and total seed isoflavone concentrations, but not glycitein. Candidate genes do not appear to be associated to either this locus or to *qGLY5_2*. The third locus with novel effects is located on Gm08 and is responsible for genistein (*qGEN8*) and total isoflavones (*qTOT8*). A QTL in the vicinity that influenced daidzein and total isoflavones was found in the same study by Gutierrez-Gonzalez et al. (2010b), in which the authors reported that the chalcone synthase (CHS) cluster, with up to as many as 12 CHS-related genes, was located within the confidence interval of the QTL. Nevertheless, it is not clear whether both regions described the same QTL, because the loci reported here are about 20 cM apart from the CHS cluster and Magellan is the bearer of the positive allele of the gene underlying the QTL, while in the previous study the PI 437654 allele had the positive effect. The last of the newly detected additive effects were found on Gm11. The importance of this locus was that it accounted for an increase in all three individual isoflavones plus the sum of all of them (*qGEN11*, *qDAI11*, *qGLY11* and *qTOT11*). This QTL was reported to affect only glycitein in an earlier study (Kassem et al. 2006). The three novel additive effects added, on genistein, daidzein and total isoflavones, are in consonance with the presence of a 4-coumarate:CoA ligase (*4CL*) gene within the QTL interval, as this enzyme catalyzes one of the first reactions in the phenylpropanoid pathway and is required to synthesize all isoflavones.

Several epistatic interactions were found involving the major locus on Gm05, reflecting the central role of this locus. Epistatic interactions have been proven to be a key genetic factor in determining the final seed isoflavone concentrations (Gutierrez-Gonzalez et al. 2009); however, they are difficult to confirm because of their usually small effects and environmental interactions. Importantly, we could validate a relevant epistatic interaction involving two additive QTLs: the major locus at Gm05 (*qGm05*) and *qGm08*. This interaction was previously described for a different population with the PI 437654 parental line in common (Gutierrez-Gonzalez et al. 2010b). However, in that report, epistatic interaction significantly influenced genistein and daidzein, and in the present work it accounted for genistein and total isoflavones. The combined result of the two studies suggests that the epistatic interaction involves all three metabolites.

Overall, our results suggest that the use of markers at the *qGm05* locus in MAS will be quite effective in increasing isoflavone concentrations in soybeans and that PI 437654 can be a good source for the positive allele, since the co-segregation between seed color and isoflavone concentrations appears to be insignificant. We also highlighted some putative PCR-based SSR markers for easy implementation in breeding programs. Other novel isoflavone-synthesizing regions were also discovered that may assist in genetic characterization of the isoflavone pathway. Additionally, a list containing the primer sequence information for the BARC_SNP markers flanking the identified additive QTL and epistatic interaction effects has been condensed in Supplemental Table 2.

Acknowledgments The authors thank Dr Jason D. Gillman for his contribution and fruitful discussions. This work was supported in part by grants from DOE (DE-SC0001295), NSF (MCB-0923779) and USDA (2010-65116-20514) to O.Y.

References

Anand SC, Gallo KM (1984) Identification of additional soybean germplasm with resistance to race 3 of the soybean cyst nematode. Plant Dis 68:593–595

Anand SC, Gallo KM, Baker IA, Hartwig EE (1988) Soybean plant introductions with resistance to races 4 or 5 of soybean cyst nematode. Crop Sci 28:563–564

Arelli APR, Wilcox JA, Meyers O Jr, Gibson PT (1997) Soybean germplasm resistant to Races 1 and 2 of *Heterodera glycines*. Crop Sci 37:1367–1369

Bernardo R (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci 48:1649–1664

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890

Caldwell CR, Britz SJ, Mirecki RM (2005) Effect of temperature, elevated carbon dioxide, and drought during seed development on the isoflavone content of dwarf soybean [Glycine max (L.) Merrill] grown in controlled environments. J Agric Food Chem 53:1125–1129

de Givry S, Bouchez M, Chabrier P, Milan D, Schie T (2005) CARTHAGENE: multipopulation integrated genetic and radiated hybrid mapping. Bioinformatics 21(8):1703–1704

Diers BW, Cary T, Thomas D, Colgrove A, Niblack T (2010) Registration of LD00-2817P soybean germplasm line with resistance to soybean cyst nematode from PI 437654. J Plant Registrations 4:141–144

Eldridge A, Kwolek W (1983) Soybean isoflavones: effect of the environment and variety on composition. J Agric Food Chem 31:394–396

Fan J, Gunderson K, Bibikova M, Yeakley J, Chen J, Garcia EW, Lebruska L, Laurent M, Shen R, Barker D (2006) Illumina universal bead arrays. Methods Enzymol 410:57–73

Graham TL, Graham MY, Subramanian S, Yu O (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 JJ, Wu X, Zhang J, Lee JD, Ellersieck M, Shannon JG, Yu O, Nguyen HT, Sleper DA (2009) Genetic control of soybean seed isoflavone content: importance of statistical model and epistasis in complex traits. Theor Appl Genet 119(6):1069–1083

Gutierrez-Gonzalez JJ, Guttikonda S, Aldrich DL, Tran LSP, Zhong R, Yu O, Nguyen HT, Sleper DA (2010a) Differential expression of isoflavone biosynthetic genes in soybean during water deficits. Plant Cell Physiol 51(6):936–948

Gutierrez-Gonzalez JJ, Wu X, Gillman JD, Lee JD, Zhong R, Shannon JG et al. (2010b) Intricate environment-modulated genetic networks control isoflavone accumulation in soybean seeds. BMC Plant Biol 10:105

Handoo ZA, Anand SC (1993) Biological manifestation of resistance to soybean cyst nematode development in 'Hartwig' soybean. Crop Prot 12:371–372

Hill J, Becker HC, Tigerstedt PMA (1998) Quantitative and ecological aspects of plant breeding. Chapman and Hall, London

Hoeck JA, Fehr WR, Murphy PA, Welke GA (2000) Influence of genotype and environment on isoflavone contents of soybean. Crop Sci 40:48–51

Hyten D, Song Q, Choi I, Yoon M, Specht J, Matukumalli L, Nelson R, Shoemaker R, Young N, Cregan P (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet 116:945–952

Hyten DL, Choi I-Y, Song Q, Specht JE, Carter TE, Shoemaker RC, Hwang E-Y, Matukumalli LK, Cregan P (2010) A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for QTL mapping. Crop Sci 50:960–968

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152:1203–1216

Kassem MA, Shultz J, Meksem K, Cho Y, Wood AJ, Iqbal MJ, Lightfoot DA (2006) An updated 'Essex' by 'Forrest' linkage map and first composite interval map of QTL underlying six soybean traits. Theor Appl Genet 113:1015–1026

Lozovaya VV, Lygin AV, Ulanov AV, Nelson RL, Dayde J, Widhohm JM (2005) Effect of temperature and soil moisture status during seed development on soybean seed isoflavone concentration and composition. Crop Sci 45:1934–1940

Mebrahtu T, Mohamed A, Wang CY, Andebrhan T (2004) Analysis of isoflavone contents in vegetable soybeans. Plant Foods Hum Nutr 59:55–61

Meksem K, Njiti VN, Banz WJ, Iqbal MJ, Kassem MyM, Hyten DL, Yuang J, Winters TA, Lightfoot DA (2001) Genomic regions that underlie soybean seed isoflavone content. J Biomol Biotechnol 1:38–44

Moy P, Qutob D, Chapman BP, Atkinson I, Gijzen M (2004) Patterns of gene expression upon infection of soybean plants by *Phytophthora sojae*. Mol Plant Microbe Interact 17:1051–1062

Primomo VS, Poysa V, Ablett GR, Jackson CJ, Gijzen M, Rajcan I (2005) Mapping QTL for individual and total isoflavone content in soybean seeds. Crop Sci 45:2454–2462

Rochfort S, Panozzo J (2007) Phytochemicals for health, the role of pulses. J Agric Food Chem 55:7981–7994

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178–183

Shannon JG, Wrather JA, Sleper DA, Robbins RT, Nguyen HT, Anand SC (2007) Registration of 'Jake' soybean. J of Plant Registrations 1:29–30

Song QJ, Jia GF, Zhu YL, Hwang EY, Hyten DL, Cregan PB (2010) Abundance of SSR Motifs and Development of Candidate Polymorphic SSR Markers (BARCSOYSSR_1.0) in Soybean. *Crop Sci*. 50: ahead of print

Subramanian S, Graham MY, Yu O, Graham TL (2005) RNA interference of soybean isoflavone synthase genes leads to silencing in tissues distal to the transformation site and to enhanced susceptibility to *Phytophthora sojae*. Plant Physiol 137:1345–1353

Tsukamoto C, Shimada S, Igita K, Kudou S, Kokubun M, Okubo K, Kitamura K (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

Usui T (2006) Pharmaceutical prospects of phytoestrogens. Endocr J 53(1):7–20

van Ooijen JH, Voorrips RE (2001) JoinMap 3.0 software for the calculation of genetic linkage maps. Plant Res Int, Wageningen

Vuong TD, Sleper DA, Shannon JG, Nguyen HT (2010) Novel quantitative trait loci for broad-based resistance to soybean cyst nematode (*Heterodera glycines* Ichinohe) in soybean PI 567516C. Theor Appl Genet 121:1253–1266

Wang H, Murphy PA (1994) Isoflavone composition of American and Japanese soybeans in Iowa: effects of variety, crop year, and location. J Agric Food Chem 42:1674–1677

Wang DL, Zhu J, Paterson AH (1999) Mapping QTLs with epistatic effects and QTL × environment interactions by mixed linear model approaches. Theor Appl Genet 99:1255–1264

Yang J, Zhu J, Williams RW (2007) Mapping the genetic architecture of complex traits in experimental populations. Bioinformatics 23:1527–1536

Zeng ZB (1993) Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci. Proc Natl Acad Sci USA 90:10972–10976

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics 136:1457–1468

Zeng ZB, Kao CH, Basten CJ (2000) Estimating the genetic architecture of quantitative traits. Genet Res 74:279–289

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

Zhang J, Yu O (2009) Metabolic engineering of isoflavone biosynthesis in seeds. In: Modification of seed composition to promote health and nutrition. Krishnan H (ed). Agronomy Monograph Series, pp 151–177