ORIGINAL PAPER

Genetic control of soybean seed isoflavone content: importance of statistical model and epistasis in complex traits

Juan Jose Gutierrez-Gonzalez · Xiaolei Wu · Juan Zhang · Jeong-Dong Lee · Mark Ellersieck · J. Grover Shannon · Oliver Yu · Henry T. Nguyen · David A. Sleper

Abstract A major objective for geneticists is to decipher genetic architecture of traits associated with agronomic importance. However, a majority of such traits are complex, and their genetic dissection has been traditionally hampered not only by the number of minor-effect quantitative trait loci (QTL) but also by genome-wide interacting loci with little or no individual effect. Soybean (*Glycine max* [L.] Merr.) seed isoflavonoids display a broad range of variation, even in genetically stabilized lines that grow in a fixed environment, because their synthesis and accumulation are affected by many biotic and abiotic factors. Due to this complexity, isoflavone QTL mapping has often produced conflicting results especially with variable growing conditions. Herein, we comparatively mapped soybean seed isoflavones genistein, daidzein, and glycitein by using several of the most commonly used mapping approaches: interval mapping, composite interval mapping, multiple interval mapping and a mixed-model based composite interval mapping. In total, 26 QTLs, including many novel regions, were found bearing additive main effects in a population of RILs derived from the cross between Essex and PI 437654. Our comparative approach demonstrates that statistical mapping methodologies are crucial for QTL discovery in complex traits. Despite a previous understanding of the influence of additive QTL on isoflavone production, the role of epistasis is not well established. Results indicate that epistasis, although largely dependent on the environment, is a very important genetic component underlying seed isoflavone content, and suggest epistasis as a key factor causing the observed phenotypic variability of these traits in diverse environments.

Communicated by M. Sillanpaa.

J. J. Gutierrez-Gonzalez · X. Wu · J.-D. Lee ·
J. G. Shannon · H. T. Nguyen · D. A. Sleper (✉)
Division of Plant Sciences, National Center for Soybean
Biotechnology, University of Missouri,
Life Sciences Building, Columbia, MO 65201, USA
e-mail: Sleperd@missouri.edu

J. J. Gutierrez-Gonzalez
e-mail: juangutierrez@mizzou.edu

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

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

Introduction

A great proportion of the agronomic traits of economic importance are complex in nature as being controlled by many genes, each with little individual effects and highly influenced by environment. Soybean seed isoflavones, namely genistein, daidzein and glycitein, have long been implicated in human health (see Messina 1999; Ososki and Kennelly 2003; Usui 2006; Rochfort and Panozzo 2007, for reviews), and more recently in plant defense and nodulation (Subramanian et al. 2004, 2005, 2006, 2007; Zhang et al. 2007). They are considered complex traits because their levels are highly variable and regulated by multiple environmental and genetic factors (Hoeck et al. 2000; Mebrahtu et al. 2004). Induction of the phenylpropanoid pathway, a branch of which synthesizes isoflavonoids, by environmental factors has been extensively documented.
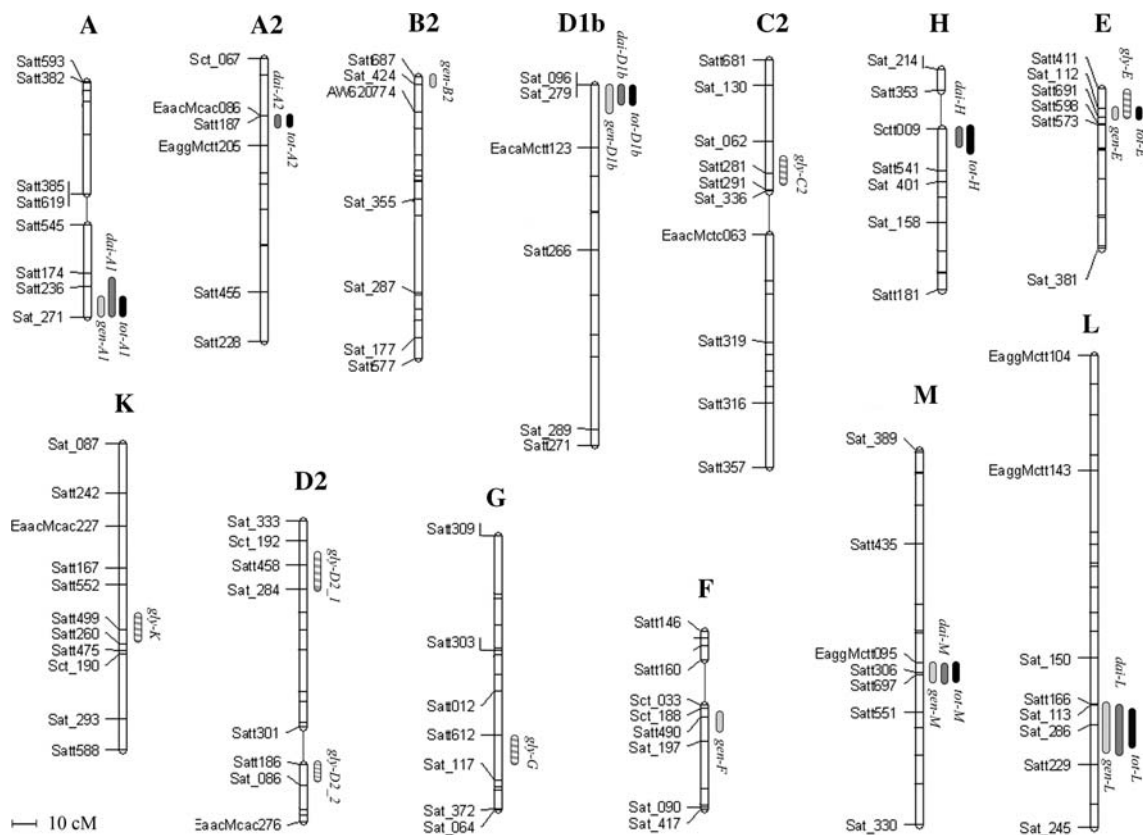
Fig. 1 Linkage group map summarizing QTL locations detected for all different models. *Light grey ovals* indicate genistein loci, *dark grey ovals* indicate daidzein loci, *black ovals* indicate total isoflavone loci, and *hatched ovals* indicate glycitein loci. Flanking and other key markers are depicted at the *right side* of the linkage group (LG).

Marks along LGs that do not have a name attached correspond to other marker positions that were used to genotype the population. The name of the QTL, shown aside each oval, is compound of the influenced trait: genistein (*gen*), daidzein (*dai*), glycitein (*gly*), total of isoflavones (*tot*) followed by the linkage group

For example, a particular genotype planted in different years in the same location can vary up to threefold for its seed isoflavone content (Eldridge and Kwolek 1983; Wang and Murphy 1994). Temperature is the most broadly studied influencing abiotic factor (Tsukamoto et al. 1995; Caldwell et al. 2005; Lozovaya et al. 2005), although water regime (Bennett et al. 2004; Caldwell et al. 2005; Lozovaya et al. 2005), soil fertility (Vyn et al. 2002), UV light (Beggs et al. 1985), and γ radiation (Variyar et al. 2004) are also influential factors. Among the biotic elicitors, the activation of enzymes in the pathway following rhizobial and pathogen infection is also well known (Peters et al. 1986; Lawson et al. 1996; Stafford 1997; McKhann et al. 1998; Dhaubhadel et al. 2003; Naoumkina et al. 2007).

Distribution of isoflavone compounds within segregating populations suggests a quantitative genetic inheritance component (Supplemental Fig. 1). To elucidate the genomic regions underlying polygenic traits, or quantitative trait loci (QTL), several statistical methodologies have been used. In single-marker locus analysis (SMA), association between a QTL and a single marker is determined by analysis of variance (ANOVA) with significance tests or

simple linear regression. SMA is easy to implement because there is no requirement for a linkage map; however, it tends to confound QTL effect and distance in such a way that a QTL with little effect closely linked to a marker may be treated equal to a major QTL located distant from the marker. Simple interval mapping (IM) was proposed by Lander and Botstein (1989) to overcome the limitations of SMA. IM increases the probability of detection by considering two consecutive markers at each time and assessing the presence of a QTL by performing likelihood ratio tests to assign a probability (LOD score) at fixed points within the interval. A multiple regression variant of IM was proposed by Haley and Knot (1992) as a less computationally demanding approach. Composite interval mapping (CIM) (Zeng 1993, 1994) goes a step further and incorporates multiple regression analysis in IM, by considering some other key markers as cofactors (co-variables) to control external genetic variation when testing a particular interval. Neither IM nor CIM are able to detect multiple interacting QTL. Finally, multiple interval mapping (MIM) (Kao et al. 1999; Zeng et al. 2000) fits several marker intervals and their epistatic interactions

simultaneously in a mixed model approach by a multidimensional search. Every time a new factor is incorporated into a model the residual variance or error component (ε) decreases, thereby the model is able to explain more phenotypic variation, and the power and precision of QTL detection improve. Deciphering the genetic architecture of complex traits requires inclusion of as many of the factors accounting for their variation as possible. Conversely, lacking of a complete model may lead to a conflicting outcome. For instance, Meksem et al. (2001) and Kassem et al. (2004) reported eight loci associated with isoflavone amounts in soybean seeds by using IM. Intriguingly, only two of them were confirmed and 14 new ones were found when CIM was implemented running the same phenotypic data (Kassem et al. 2006).

Epistasis, or interactions between effects of alleles from two or more genetic loci, can account for a significant part of the variation for complex traits (Doerge 2001; Phillips 2008). It depends on the genetic background, and as a result, it is also believed to play an important role in certain biological processes such as adaptation, evolution and speciation (Whitlock 1995; Juenger et al. 2005). There is increasing evidence that epistatic models can improve the statistical power to detect QTL as well as the accuracy of their localization and the precision in determining the associated confidence intervals, because they might incorporate genetic variation that would not be associated with the trait otherwise (Yi and Xu 2002). Conversely, if interacting loci are not considered, variation caused by epistasis is removed from the tested model and absorbed into residual error. Nevertheless, detecting epistasis genome-wide is cumbersome and often hampered by either computationally demanding algorithms, a lack of appropriate statistical methodology, or a reduced mapping population size (Doerge 2001; Mao and Da 2004). Conventional single-QTL mapping models, like IM and CIM, are not able to deal with interacting loci, treating epistasis as background noise. Here, we use two software packages for the comparative. First, QTL-Cartographer has been extensively used for QTL mapping as it implements IM, CIM, and MIM among other methods. Second, in addition to incorporating epistasis in the model, the power of the other package, QTLNetwork, resides in that data from different locations and within-location replication are combined to perform an integrated analysis by a mixed linear model (MCIM) (Yang and Zhu 2005; Yang et al. 2007).

Herein we used a recombinant inbred line (RIL)-population developed from a cross between an isoflavone low-content line, Essex (1,236 ± 160 and 940 ± 227 μg/g, total converted isoflavone per gram of seed weight at Bradford Research and Extension Center, BREC, and Delta Research Center, DRC, respectively, see "Materials and methods") and the high-content PI 437654 (2,092 ± 54 and 1,623 ± 120) to elucidate the genetic architecture of isoflavones. The study proves that isoflavone accumulation in seeds is influenced by multiple interacting genetic loci, with epistasis playing an important role. Altogether, 26 additive genomic regions and many additive-by-additive epistatic interactions were found by comparative mapping using existing methodologies to affect isoflavone contents. We also discuss the importance of the statistical model and of considering epistatic interactions when mapping complex traits.

Materials and methods

Plant material and growing conditions

Eight soybean (*Glycine max* [L.] Merr.) accessions: RCAT-Angora, AC-756, PI 437654, Williams 82, PI 438489B, Magellan, Essex and Forest, were chosen for a comparative study to select appropriate parental lines for isoflavone studies. These were planted in 19-l pots following a randomized complete block design (RCBD) with four replications and grown between April and September 2005 in a greenhouse (GH) at the University of Missouri. Pots were weighted and adjusted to the same substrate amount. Growth conditions were 30/20°C day/night temperatures, and 16/8 h day/night lengths. All but PI 438489B were also planted in April 2005 in two-row plots following a RCBD with three replications in two field locations: University of Missouri Bradford Research and Extension Center (BREC, 38°9′N), located near Columbia, MO, and Grand Pass (GP, 39°19′N), Saline County, MO. Plants were allowed to grow under irrigated conditions until maturity, after which seed samples were collected from single plants (GH) or entire plot and pooled (BREC and GP). The pooled linear statistical model was a split plot design which contained the effect of environment (GH, BREC, GP), replication within environment, genotype (RCAT-Angora, AC-756, PI 437654, Williams 82, Magellan, Essex and Forest), and the interaction of environment × genotype. Mean differences were determined using Fisher's protected least significant difference according to the methods described by Chew 1977.

A mapping population comprised of 196 F$_7$-derived recombinant inbred lines (RILs) developed from a cross between Essex and the high-isoflavone content plant introduction PI 437654 were planted in two-row plots under irrigated conditions in a RCBD with three replications at two locations: BREC and the University of Missouri Delta Research Center (DRC, 36°44′N), located near Portageville, MO. Physiologically matured seed samples were harvested for isoflavone quantification from one plant

per plot at the DRC and from a pool of at least three plants at BREC. Effects were tested using PROC GLM in SAS STAT 9.1 (SAS Institute Inc., Cary, NC). The pooled linear model contained the effect of environment, replication (environment), genotype, and environment × genotype. Variance components to calculate heritability estimates were determined using PROC mixed.

Isoflavone extraction and quantification

Approximately 2.5 g (∼20 seeds) of soybean seeds were ground to a fine powder using a General Electric seed grinder (model 5XBG008, New York City, NY). 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). Samples were analyzed by reverse-phase HPLC on an Agilent 1100 high-performance liquid chromatography (HPLC) system (Santa Clara, CA). 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. Detection of metabolites was performed by photodiode array. Identification and quantification of each isoflavone component were based on available standards (Indofine Chemical Co., Somerville, NJ). Measurements are given as micrograms of isoflavone per gram of seeds plus/minus standard deviation or standard error, when corresponds (μg/g ± SD or SE).

Linkage map and QTL analysis

DNA was extracted from young trifoliate leaves with the help of AutoGenprep 960 robot (AutoGen, Holliston, MA). Linkage map contained a total of 276 SSR and AFLP markers which were distributed on 26 linkage groups (LG), covering 2,406.5 cM and average distance of 8.7 cM (for more detailed information see Wu et al. 2009). Statistical analysis was performed using the SAS STAT 9.1 program, and heritability in the broad-sense over environments was calculated according to Hill et al. (1998):

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

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

QTLNetwork v2.0 (Institute of Bioinformatics, Zhejiang University, Hangzhou, China) was used for the mixed-model based composite interval mapping (MCIM) with individual observations from three replications, and

two-locations combined input data. Windows QTL Cartographer v.2.5 (Statistical Genetics, North Carolina State University) and trait averaged values for each environment were used for interval mapping (IM), composite interval mapping (CIM), and multiple interval mapping (MIM). For MCIM, critical F-value was assessed by permutation test using 1,000 permutations. QTL effects were estimated using Markov chain Monte Carlo method. Candidate interval selection, epistatic effects, and putative QTL detection were calculated with an experimental-wise type I error of α = 0.05, α = 0.001, and α = 0.001, respectively. Significance level for candidate intervals was relaxed to allow more genomic intervals to be included in the model. Genome scan was performed using 10 cM window size and 1 cM walk speed in all models. Forward and backward regression method was chosen to select markers for CIM analysis. To create MIM models, initial search for main QTLs was done by the MIM forward search method. Epistatic network outcome from QTLNetwork was considered as a starting point in the setting of epistatic interactions. LOD-score significance threshold values were estimated for each trait and location by permutation test analysis with 1,000 permutations and error α = 0.05, lending for IM: LOD 3.1 and 3.3 for genistein at BREC and the DRC, respectively, 3.3 and 3.1 for daidzein, 2.9 and 3.1 for glycitein, and 3.0 and 3.1 for total isoflavones. For CIM the threshold values were: 3.0 and 3.1, 3.2 and 3.0, 3.0 and 3.1, and 3.3 and 3.2, respectively. Because different models were compared, and for illustrative purposes, QTL of LOD score above 3.0 were included when they were significantly detected by other models and/or locations. R^2 (%) is the proportion of the phenotypic variation explained by the each particular QTL. The heritability of the additive effect, $h^2(a)$, or the proportion of the phenotypic variation explained by the additive component of the QTL, is calculated by the mixed linear model, and theoretically it should account for the same variance components as R^2, except for the fixed effects.

Results

Genotypic quantification

Phenotypic characterization of soybean lines over years and locations suggests that isoflavones are influenced by both genes and environment (Hoeck et al. 2000; Mebrahtu et al. 2004). In our studies, lines displayed a 2.8-fold range of variation within field locations and 3.0-fold in GH (Table 1). Among them, PI 437654 consistently accumulated the highest levels of isoflavones across all environments: 1,804 ± 102 μg/g at BREC, 2,052 ± 215 μg/g at GP, and 2,657 ± 247 μg/g at GH. On the other hand, the

Table 1 Isoflavone seed content of different genotypes

Lines	MG[a]	Genistein			Daidzein			Glycitein			Total isoflavones		
		BREC	GP	GH	BREC	GP	GH	BREC	GP	GH	BREC	GP	GH
PI-438489B	IV			2,264 ± 136			348 ± 53			146 ± 12			2,553 ± 198
PI-437654	III	1,487 ± 67	1,655 ± 177	1,909 ± 65	214 ± 13	268 ± 30	491 ± 67	102 ± 22	127 ± 12	143 ± 15	1,803 ± 102	2,051 ± 215	2,656 ± 247
RCAT Angora	II	1,024 ± 36	1,070 ± 125	1,395 ± 183	170 ± 27	181 ± 21	236 ± 24	31 ± 1	37 ± 2	39 ± 6	1,227 ± 60	1,288 ± 145	1,653 ± 343
AC 756	II	537 ± 103	635 ± 158	1,191 ± 88	88 ± 14	121 ± 28	229 ± 40	26 ± 2	33 ± 4	34 ± 6	652 ± 118	734 ± 93	1,455 ± 222
Essex	V	922 ± 8	847 ± 63	1,119 ± 93	177 ± 9	156 ± 26	232 ± 33	50 ± 1	46 ± 6	57 ± 11	1,150 ± 0	1,068 ± 98	1,409 ± 184
Forrest	V	986 ± 146	1,052 ± 72	835 ± 152	178 ± 46	193 ± 29	167 ± 32	40 ± 3	44 ± 5	37 ± 7	1,206 ± 194	1,290 ± 93	1,040 ± 308
Williams82	III	752 ± 100	798 ± 48	803 ± 94	118 ± 13	126 ± 6	107 ± 14	41 ± 1	45 ± 2	35 ± 4	903 ± 98	970 ± 55	946 ± 107
Magellan	IV	648 ± 19	674 ± 90	759 ± 102	102 ± 7	112 ± 12	92 ± 16	46 ± 3	51 ± 2	49 ± 4	797 ± 23	838 ± 97	901 ± 121

Measurements in μg per g of seeds ± standard deviation (SD). The assays and number of repeats are described in "Material and methods"

BREC Bradford Research and Extension Center, GP grand pass, GH greenhouse

[a] Maturity group

lowest values were obtained for AC756 in both field locations: 652 ± 128 μg/g (BREC) and 735 ± 93 μg/g (GP); and for Magellan (902 ± 121 μg/g) in the GH. Differences of least squares mean between field locations of each tested genotype were not significant (t-test; $P < 0.05$) (data not shown). However, PI 437654, RCAT-Angora, AC756, and Essex had significantly higher isoflavone contents in the GH than in the field ($P < 0.05$), and particularly AC756 as much as twofold.

An analysis of variance was conducted to establish the causes of such variation. Highly significant ($P < 0.0001$) genotypic and environmental effects were present for genistein, daidzein and total isoflavones. For glycitein, there were also significant genotypic ($P < 0.0001$) and environmental ($P < 0.05$) effects. Genotype × environment interaction (g × e) effect was an important component as well for genistein, daidzein and total isoflavone ($P < 0.0001$), and for glycitein ($P < 0.05$). As a whole, the results suggest that all genetic, environmental and g × e interaction effects contribute to the variation of isoflavone levels observed in genetically homogeneous lines.

Genetic and phenotypic variation within mapping population

Analysis of variance and heritability are often conducted in mapping studies to assess sources of variation within experimental populations. Heritability (H^2, $0 \leq H^2 \leq 1$) is an estimate of the genetic contributions to the phenotypic variance, and in its broad-sense includes additive, dominant, and epistatic genetic variance components. Within Essex × PI 437654 segregating mapping population, the heritability estimates were high for genistein, daidzein, and total isoflavones, 0.83, 0.89, and 0.86, respectively; and moderate for glycitein (0.43). ANOVA conducted over locations revealed a significant ($P < 0.0001$) genetic, environmental, and g × e effects (Table 2). The results confirm that the growing environment is crucial and that isoflavone accumulation in soybean seeds is not only controlled by both genetic and environmental factors but also there are important g × e interactions. Observations are also in agreement with other researchers, and consistent with the long-time consideration of these compounds as complex traits.

Isoflavone QTL mapping

The apparently conflicting results obtained by some research groups when using different mapping models (Meksem et al. 2001; Kassem et al. 2004, 2006) caused us to consider a multi-model strategy by testing some of the most broadly used mapping algorithms on the same phenotypic data. Some models, such as the one implemented in

Table 2 Mean and range, ANOVA, random effects, and heritability over environments of Essex × PI 437654 mapping population

Trait	Mean ± SD[a]	Range[a]	CV (%)	Effects[b]			H^{2c}	Skewness	Kurtosis
				g	e	g × e			
Genistein	844 ± 296	107–1,980	18.8	<0.0001	<0.0001	<0.0001	0.83	0.15	−0.15
Daidzein	463 ± 198	73–1,384	22.7	<0.0001	<0.0001	<0.0001	0.89	0.63	0.59
Glycitein	88 ± 48	14–342	26.4	<0.0001	<0.0001	<0.0001	0.43	1.65	3.35
Total isoflavones	1,393 ± 478	220–3,008	18.3	<0.0001	<0.0001	<0.0001	0.86	0.18	−0.17

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

[b] P values of random effects over environments: genetic (g), environmental (e), and g × e interaction

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

MCIM, are able to combine data sets from different locations, years, and replications by means of a mixed model. Others, on the other hand, analyze each dataset independently. A total of seven loci displaying additive effect were found to explain genistein seed levels by a combination of all models, six of which were detected by MCIM (Table 3; Fig. 1). Among all, there was a major QTL on linkage group (LG) A1, named *gen-A1* for genistein locus at LG A1. A QTL in the proximity was previously reported to influence genistein, daidzein and total isoflavones (Primomo et al. 2005). Other important QTL were found in linkage groups D1b (*gen-D1b*), E (*gen-E*), and M (*gen-M*). In addition, minor loci were detected in groups B2 (*gen-B2*), also reported by Kassem et al. (2004), and F (*gen-F*). Lastly, a locus in LG L (*gen-L*) was not revealed by MCIM but it was detected by CIM and MIM although only in one location. Surprisingly, out of the total seven discovered QTL, IM model was not able to localize any single locus at BREC, and only *gen-A1* at DRC.

Six additive genomic regions were also discovered for daidzein, the second most abundant soy isoflavone, one of which was exclusively identified by MIM and only at DRC (*dai-L*). The other five were: *dai-A1*, *dai-A2*, *dai-D1b*, *dai-M*, and *dai-H* (Table 4; Fig. 1). The latter also displayed an additive × environment (a × e) interaction effect when MCIM was implemented. Importantly, only the major QTL (*dai-A1*) was detected by all models at both locations. This region was also found to explain genistein accumulation (*gen-A1*) as previously shown. The *dai-H* locus was identified by all models as well, but only at DRC.

Glycitein is the most variable isoflavone in soybean (Table 2). This complicates QTL discovery because it requires larger mapping populations and advanced statistical methods that can take into account most of that variation. Six loci were found to control glycitein accumulation when considering all mapping algorithms together (Table 5; Fig. 1). Two of them viz. one in LG C2 (*gly-C2*), and the other in LG E (*gly-E*), were revealed by MCIM model with data from BREC and DRC combined. Locus *gly-C2* was the major glycitein QTL reported herein.

However, IM could not map a single QTL at BREC, and just *gly-C2* at DRC. This same region was also detected with CIM in DRC, and with MIM in both locations. Two other loci in LG D2 (*gly-D2_1* and *gly-D2_2*) were found to be significant only in BREC. On the other hand, *gly-G* and *gly-K* were detected at DRC solely. Interestingly, no one of the QTLs for glycitein discovered in the present study was previously reported.

Total isoflavone (sum of genistein, daidzein and glycitein) seed content was also considered as a trait and mapped. As a result, seven QTLs were identified. All but one were found by MCIM with two locations combined data (Table 6; Fig. 1). The major QTL was located in LG A1 (*tot-A1*), and matched also the major QTL for genistein and daidzein, *gen-A1* and *dai-A1*. All mapping approaches identified this as a major locus. In addition to its important additive component, *tot-A1* also showed a significant a × e interaction effect revealed by MCIM when analyses were performed over both locations. The other identified QTLs were located on LGs A2 (*tot-A2*), D1b (*tot-D1b*), E (*tot-E*), H (*tot-H*), M (*tot-M*), and L (*tot-L*), also showing a × e interaction effect. IM only detected two out of the seven, *tot-A1* and *tot-H*, although the later only in one environment. With CIM, *tot-D1b* and *tot-E* regions were also added. Finally, model MIM incorporated *tot-A2* and *tot-L* at BREC and DRC, respectively. Taken together, the results of all four traits suggest that the accumulation of isoflavones in soybean seeds is largely under additive genetic control, and that there can also be an a × e interaction effect.

Epistatic interactions

The structure of the segregating population allowed us to map only the additive-by-additive (aa) epistatic network. However, aa is not only the main interaction but also the easiest to detect because its statistical power of detection is higher than for the other components, i.e. additive-by-dominant, dominant-by-additive, and dominant-by-dominant (Mao and Da 2004). Among all mapping methods

Table 3 QTL bearing additive effect detected for genistein seed content by different models

				MCIM (combined)	IM				CIM				MIM			
					BREC		DRC		BREC		DRC		BREC		DRC	
QTL[a]	Interval	Range[b]	A[c]	h²(a)	LOD	R²	LOD	R²	LOD	R²	LOD	R²	LOD	R²	LOD	R²
gen-A1	SATT236-SAT_271	29.4–36.4	95.8 ± 9.0	5.8			6.1	25.3	3.5	9.4	6.3	22.1			9.4	21.3
gen-B2	SATT687-SAT_424	0.0–3.0	49.0 ± 7.7	1.5					3.1	6.0			3.2	4.5		
gen-D1b	SAT_279-EACAMCTT123	0.0–10.0	59.5 ± 7.8	2.3					3.8	8.4			8.4	9.6		
gen-E	SAT_112-SATT691	8.0–11.0	71.0 ± 7.7	2.9									7.6	7.7	3.0	3.5
gen-F	SATT490-SAT_197	3.4–10.0	40.4 ± 7.0	1.7									4.5	5.4		
gen-M	SATT697-SATT551	87.4–93.4	54.8 ± 7.0	2.6												
gen-L	SAT_113-SAT_286	141.4–160.9	−66.5								3.0	6.9			4.0	6.4
Sum of contributions of all QTLs in the model				16.8				25.3		23.8		29.0		27.2		31.2

Only QTLs displaying R² or h²(a) values in the corresponding column were detected by that particular model. Locations: Bradford Research and Extension Center (BREC), and Delta Research Center (DRC)

h²(a) Heritability of the additive effect or percentage of variation that is explained by the additive component of the QTL, R² percentage of variance accounted for by the QTL, LOD Log of odds (LOD score) associated to the QTL detected by each method, MCIM mixed model-based composite interval mapping combining data from two locations, IM interval mapping, CIM composite interval mapping

[a] Name given to a particular QTL, gen for genistein content, followed by the linkage group and a number when more than one in the same LG

[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 when applicable. Mean effect of substituting both Essex alleles by PI437654 alleles. Thus, positive values indicate that the PI437654 allele increase the phenotypic value

Table 4 QTL bearing additive effect detected for daidzein seed content by different models

					IM				CIM				MIM			
MCIM (combined)					BREC		DRC		BREC		DRC		BREC		DRC	
QTLᵃ	Interval	Rangeᵇ	Aᶜ	h^2(a)	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2
dai-A1	SATT236-SAT_271	22.5–36.4	69.5 ± 5.7	9.2	5.3	20.3	5.7	19.7	5.7	13.9	6.7	18.8	7.8	18.0	7.5	15.5
dai-A2	SATT187-EAGGMCTT205	23.0–27.1	−18.2 ± 4.6	2.4					3.2	5.4						
dai-D1b	SAT_279-EACAMCTT123	0.0–8.0	37.1 ± 4.8	2.2					3.0	5.7			5.4	5.0		
dai-H	SCTT009-SATT541	0.0–7.0	45.1 ± 5.0	3.8			4.0	9.4			3.2	5.8			5.8	7.9
dai-M	SATT697-SATT551	87.4–94.4	42.4 ± 4.7	3.0									4.3	4.9		
dai-L	SAT_113-SAT_286	142.8–161.9	−27.54 ± 5.13												5.0	6.5
Sum of contributions of all QTLs in the model				20.7		20.3		29.1		25.0		24.6		27.9		29.9

Only QTLs displaying R^2 or h^2(a) values in the corresponding column were detected by that particular model. Locations: Bradford Research and Extension Center (BREC), and Delta Research Center (DRC)

h^2(a) Heritability of the additive effect or percentage of variation that is explained by the additive component of the QTL, R^2 percentage of variance accounted for by the QTL, LOD log of odds (LOD score) associated to the QTL detected by each method, MCIM for mixed model-based composite interval mapping combining data from two locations, IM interval mapping, CIM composite interval mapping

[a] Name given to a particular QTL, dai for daidzein content, followed by the linkage group and a number when more than one in the same LG

[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 when applicable. Mean effect of substituting both Essex alleles by PI437654 alleles by PI437654 alleles. Thus, positive values indicate that the PI437654 allele increase the phenotypic value

Table 5 QTL bearing additive effect detected for glycitein seed content by different models

	MCIM (combined)				IM				CIM				MIM			
					BREC		DRC		BREC		DRC		BREC		DRC	
QTL^a	Interval	Range^b	A^c	h²(a)	LOD	R²	LOD	R²	LOD	R²	LOD	R²	LOD	R²	LOD	R²
gly-C2	SATT281-SATT291	41.0–48.6	−12.2 ± 1.3	4.4			3.0	7.8			4.1	7.6	4.0	5.6	3.3	5.7
gly-E	SAT_112-SATT691	1.0–11.0	8.9 ± 1.4	2.9					4.3	7.9			3.1	4.2	3.1	3.6
gly-K	SATT552-SATT499	68.6–78.3	10.48													
gly-G	SATT612-SAT_117	81.0–92.5	−13.51												4.7	7.3
gly-D2_1	SATT458-SAT_284	13.6–26.8	8.9						4.1	6.8			4.4	5.8		
gly-D2_2	SATT186-SAT_086	0.0–3.2	−7.59										3.2	4.4		
Sum of contributions of all QTLs in the model				7.3				7.8		14.7		7.6		20.0		16.6

Only QTLs displaying R² or h²(a) values in the corresponding column were detected by that particular model. Locations: Bradford Research and Extension Center (BREC), and Delta Research Center (DRC)

h²(a) Heritability of the additive effect or percentage of variation that is explained by the additive component of the QTL, LOD log of odds (LOD score) associated to the QTL detected by each method, MCIM mixed model-based composite interval mapping combining data from two locations, IM interval mapping, CIM composite interval mapping

^a Name given to a particular QTL, gly for glycitein content, followed by the linkage group and a number when more than one in the same LG

^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 when applicable. Mean effect of substituting both Essex alleles by PI437654 alleles. Thus, positive values indicate that the PI437654 allele increase the phenotypic value

tested herein, only models MCIM and MIM include variation due to epistasis. Four epistatic interactions were found by MCIM to be significant for genistein when data were combined over both locations, and seven and five at BREC and DRC by MIM, respectively (Fig. 2; Supplemental Table S-I). Importantly, no single epistatic interaction was maintained over environments. Daidzein usually had more significant epistatic interactions and accounted for more variance than genistein. For example, at BREC eight ($R^2 = 37.7\%$) interactions were mapped by MIM (Supplemental Table S-II), and just one of them accounted for up to 10.1% of the variation. The importance of considering epistasis when dealing with complex traits was revealed greatest for glycitein, for which the proportion of its variance due to epistatic effects was up to 2.8 times the variance of the additive QTL. Strikingly, a single interaction was able to explain as much as 22.0% of the phenotypic variation and had an aa additive value of 22.53 µg/g (at BREC by MIM). In this environment, epistasis was responsible for 54.1% of the variance (Supplemental Table S-III). Finally, epistatic networks for the sum of all isoflavones showed ten statistically significant interacting pairs at BREC ($R^2 = 28.3\%$), and nine at DRC (27.7%) (MIM, Supplemental Table S-IV). The overall results show that epistasis can account for an important part of the isoflavone phenotypic variation and suggest that epistasis might be one of the main causes of the variable isoflavone amounts observed in soybean genotypes over certain environments.

Discussion

Fluctuation in isoflavone content and composition

Basal levels of certain secondary metabolites may be quite variable even in genetically stabilized lines, particularly if their production is affected by a set of biotic and abiotic factors that can act as inducers or repressors. Eight genotypes were tested in field and controlled conditions to gain more knowledge on what might be the origin of the reported tremendous oscillations observed for isoflavone concentrations over environments (Eldridge and Kwolek 1983; Wang and Murphy 1994; Tsukamoto et al. 1995; Hoeck et al. 2000; Mebrahtu et al. 2004; Caldwell et al. 2005; and Lozovaya et al. 2005). In general, plants tended to accumulate more isoflavones when growing in a controlled environment, probably because temperature can be kept in a more desired range. A series of factors, such as temperature, water, and soil nutrient condition are known to play an important role in isoflavone production, and could explain this difference. Although average values were maintained similar, 25.6°C in GH versus 24.4°C in

Table 6 QTL bearing additive effect detected for total isoflavone seed content by different models

| MCIM (combined) | | | | | IM | | | | CIM | | | | MIM | | | |
| | | | | | BREC | | DRC | | BREC | | DRC | | BREC | | DRC | |
QTL[a]	Interval	Range[b]	A[c]	h^2(a)	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2	LOD	R^2
tot-A1	SATT236-SAT_271	29.4–36.4	200.4 ± 14.4	8.1	3.5	14.9	7.3	30.3	3.5	9.5	8.2	26.3	9.9	16.6	17.4	30.0
tot-D1b	SAT_279-EACAMCTT123	0.0–7.0	110.4 ± 11.0	2.4					5.0	9.5			6.9	6.7	5.6	3.2
tot-E	SAT_112-SATT691	8.0–11.3	76.8 ± 12.0	1.9					3.3	6.4			3.1	3.8	3.7	3.1
tot-H	SCTT009-SATT541	0.0–10.0	59.1 ± 13.0	2.1			3.2	8.0			3.2	6.3				
tot-L	SAT_286-SATT229	143.8–159.9	−91.3 ± 12.8	2.3											7.4	7.4
tot-M	SATT697-SATT551	87.4–93.4	102.1 ± 11.2	3.3					4.1	8.6			3.6	4.0		
tot-A2	SATT187-EAGGMCTT205	23.0–27.1	54.13 ± 10.75										3.1	3.4		
Sum of contributions of all QTLs in the model				20.0		14.9		38.3		34.0		32.6		31.1		43.7

Only QTLs displaying R^2 or h^2(a) values in the corresponding column were detected by that particular model. Locations: Bradford Research and Extension Center (BREC), and Delta Research Center (DRC)

h^2(a) Heritability of the additive effect or percentage of variation that is explained by the additive component of the QTL, R^2 percentage of variance accounted for by the QTL, *LOD* log of odds (LOD score) associated to the QTL detected by each method, *MCIM* mixed model-based composite interval mapping combining data from two locations, *IM* interval mapping, *CIM* composite interval mapping

a Name given to a particular QTL, *tot* for total isoflavones, followed by the linkage group and a number when more than one in the same LG

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 when applicable. Mean effect of substituting both Essex alleles by PI437654 alleles. Thus, positive values indicate that the PI437654 allele increase the phenotypic value

Fig. 2 Peaks of epistasis throughout the genome for genistein, **a** at BREC and **b** at DRC; for daidzein **c** at BREC and **d** at DRC; glycitein **e** at BREC and **f** at DRC; and for the sum of all isoflavones **g** at BREC and **h** at DRC. Linkage groups of interacting loci are depicted on the *x*- and *y*-axes. The statistic *F*-value is taken as the *z*-axe. *F*-values have also a *color key* representation along peaks. Marks at *multicolor bars* at the *left* of each figure show the threshold *F*-value above which the epistasis interaction is significant
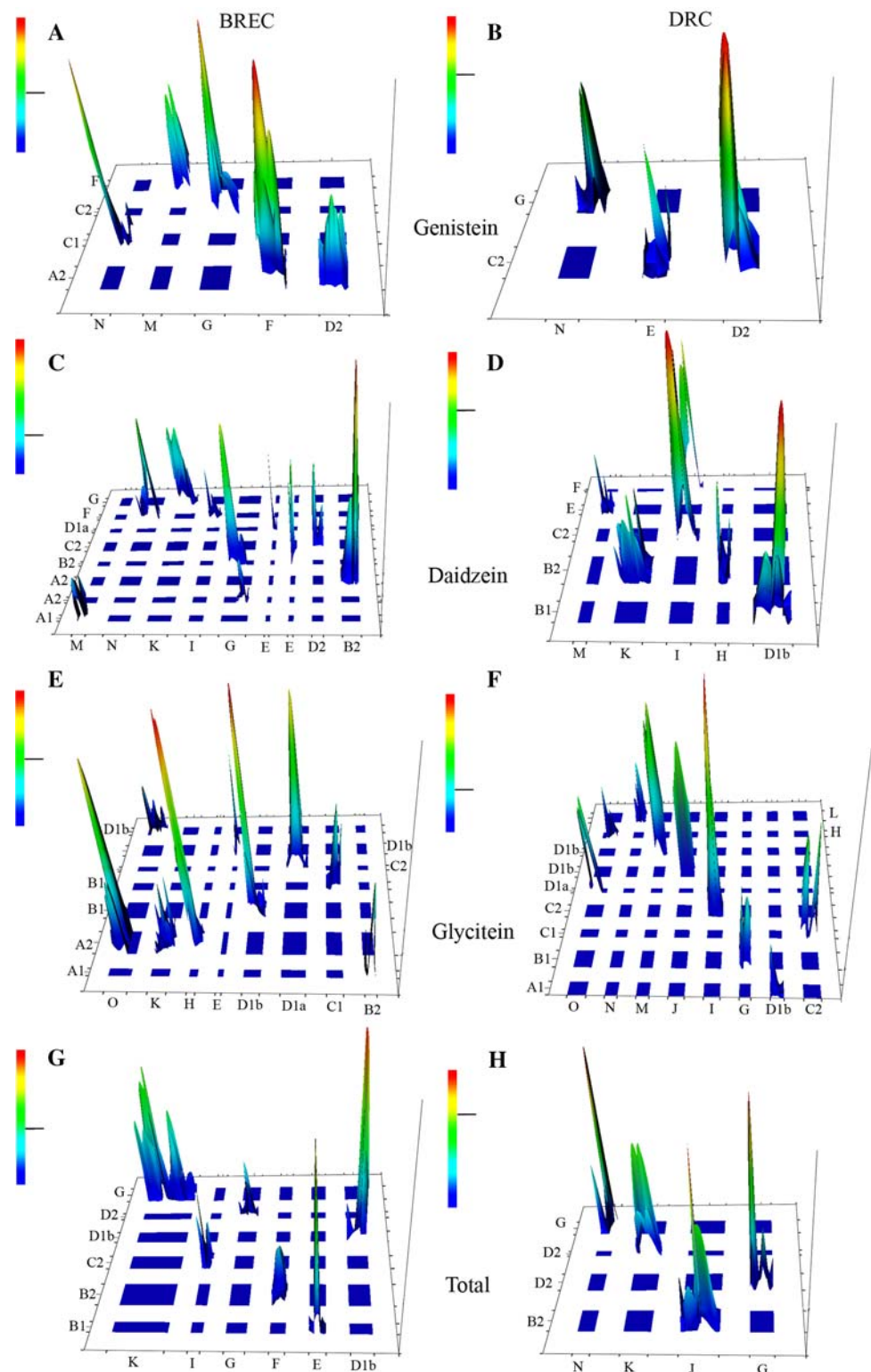
both field locations, recorded maximums were 29.4, 32.2 and 32.8°C in a 3-month period (July–September) in GH, 38.7, 37.1 and 32.2°C at BREC, and 40.0, 37.8 and 33.3°C at DRC, suggesting that cooler temperatures exert a favorable effect in isoflavone seed accumulation, as previously reported (Tsukamoto et al. 1995; Caldwell et al. 2005; and Lozovaya et al. 2005). Not to be obviated is also the fact that the growth cycle in the greenhouse was longer than in the field, thus plants had more time to synthesize and accumulate metabolic compounds.

Genetic control of isoflavone content

The variability observed for seed isoflavones in the evaluation of soybean lines was attributable not only to genetic causes but also to both the environment and g × e interaction. Highly significant environmental and g × e effects could also be found within the mapping population (Table 2). Remarkably, broad-sense heritability estimates (H^2) were intriguingly high (80–90%) for most of the isoflavone components, which indicates that genetic factors are the main contributors to the phenotypic variability. Elevated heritability values for isoflavones have been previously reported (Meksem et al. 2001; Chiari et al. 2004). With high H^2 estimates, one should not expect the variation observed on genetically identical individuals. In fact, in a population of RILs the only components contributing to H^2 are the additive and *aa* epistatic genetic variances. Thus, it is tempting to speculate that either the additive loci or the epistatic interactions or both might change their effects along with environmental changes, but then at the same time, the sum of these effects should be kept high to reach an elevated H^2. It could also be possible that the changes occur such that additive loci and epistasis counteract to affect seed isoflavone content, at least in part. Effectively, we observed that a decrease in phenotypic variance accounted for by the sum of additive QTL could be compensated with an increase in the epistatic contribution, and vice versa, as described below.

Nineteen quantitative trait loci with additive effects were found to affect genistein, daidzein and glycitein production. Among them, a major QTL was located in LG A1 (*gen-A1*, *dai-A1* and *tot-A1*) which accounted for up to 30.3% of isoflavone variation, depending on the model and trait (Tables 3, 4, 5, 6). It was the only QTL that was detected by all the tested models and locations. Other important genomic regions for the synthesis of genistein are *gen-E*, *gen-L*, *gen-D1b*, and *gen-M*. Daidzein shares a great number of enzymes in the phenylpropanoid metabolic pathway with genistein. Moreover, their relative seed content within the mapping population were 62% for genistein and 33% for daidzein, on average. Therefore, it is not surprising that four total-isoflavone genomic regions were coincident for both genistein and daidzein, i.e. at linkage groups A1, D1b, L, and M. Glycitein, however, is very distinct in many aspects, suggesting that a different branch of the pathway is probably involved in its synthesis, with distinct enzymes and/or differences in activation (Yu and McGonigle 2005). Glycitein is present in small amounts but displays more variability, the additive QTL are less numerous and usually with very little individual effects, and epistatic contributions are consistently high (Table 5, Supplemental Table S-III). This variability could be explained because it exists mainly in the tiny embryo

section of the seed, with practically no presence in cotyledons (Kim et al. 2007). Despite those barriers for QTL mapping, a major (*gly-E*) and five minor QTLs were found, suggesting that even minor QTL can be detected with adequate combinations of prediction models and population increased size.

Epistatic network of interactions

The ability for detecting aa epistatic interactions depends not only on the sample size but also on the heritability of the aa effect (h^2aa) (Mao and Da 2004). Traditionally, there have been two approaches to determine interacting loci. In the first method, the whole genome must be screened for the candidate intervals and then all possible interacting pairs are tested for significance. It is comprehensive but computationally demanding. The alternative approach relies on checking only a known set of loci introduced into the model. By using both approaches we mapped epistatic connections for genistein, daidzein, glycitein, and total isoflavones at each location separately and combined. Genistein had fewer epistatic interactions and accounted for less phenotypic variation than for any other isoflavone, but still the contribution to the total variance (R^2) was 15.2 and 27.4% at DRC and BREC, respectively. Glycitein, on the other hand, had as much as 54.1% of its variation explained by interacting loci. Daidzein and total isoflavones showed intermediate values, with about one-third of the trait variability due to epistasis. Remarkably, only one epistatic interaction was significant at both locations for all traits, proving that epistatic effects must be analyzed in particular environments (Fig. 2). In fact, regardless of the number and the magnitude of the effect, specific epistatic interactions are rarely consistent over environments. Epistasis has been considered as an important source of genetic variation in the adaptation to a changing environment (Whitlock 1995). We actually found that epistasis was very sensitive to environment, and epistatic networks were largely changed when plants grew in variable environments.

Importance of the mapping model

Variability observed in complex traits has overwhelmingly hindered mapping studies, especially if the trait is affected by many genes with little individual effects. Powerful statistical methodologies and advanced software are required to precisely isolate the loci involved. Epistatic networks of interacting loci with no additive effect by themselves add even more complexity. Therefore, it is not surprising that different mapping approaches may lead to very distinct sets of isoflavone QTL, even when the same phenotypic data were used (Kassem et al. 2004, 2006;

Primomo et al. 2005), because only considering all possible sources of variation together with a large population the chance of recovering QTL increases. The wide range of variation displayed by isoflavones suggests that only models that include most sources of variation will detect minor-effect QTL. Statistical methodologies have been developed to improve detection of loci with minor effects. For example, Mayer (2005) compared MIM and regression interval mapping and he found that MIM was more powerful and produced more accurate QTL positions and effect estimates. Similar results were obtained by Kao et al. (1999) when comparing IM, CIM, and MIM. We also proved that the use of prediction models able to include most sources of variation were more effective in detecting QTL with minor effects. Every time that some variation was removed from the error component and introduced in the model, a larger number of QTL were detected. This was particularly true for the minor QTL. For example, *gen-E*, *gen-F*, *dai-M*, and *tot-L* were discovered only by MCIM and MIM, the two models which include variation due to interacting loci. Conversely, out of seven total isoflavone QTL, a single marker analysis only detected four and six significant marker-trait associations ($P < 0.05$) at DRC and BREC, respectively (data not shown). More importantly, as many as ten associations were found which turned out not to be relevant by other prediction methods. The importance of the statistical method is further remarked, for example, for *tot-D1b*, which was not detected by IM at any location (Table 6). When CIM was implemented, it was revealed, although only at BREC. Finally, it was detected in both locations when including epistasis as a source of variation by MIM.

In addition to inclusion of epistatic effects, best results were obtained when a larger sample size with data from different locations and within-location replications were entered in the analyses (MCIM). It can be hypothesized that a QTL would remain undiscovered if it had minor effects at both locations but did not reach the statistical significance when individual locations were considered alone. Consequently, it would only be detected because of the increased statistical power due to a larger sample size. Effectively, including replicated data from all locations as individual observations might have a major impact on QTL detection for isoflavone concentration. For example, *gen-M* was only detected by MCIM in analysis of the combined data across locations. However, MCIM run for each location independently failed to detect the QTL at either location. This was repeatedly observed for other loci (data not shown). Inversely, if a QTL has a significant small effect at one location but has no effect at the other location, the significance of the QTL might be minimized when pooling the entries of both locations. This could be the case of *dai-L*, which was significant at DRC (MIM) but no

longer remained detected by the combined MCIM. Alternatively, it might also be possible that MCIM, with a model considering a × e effects, was able to discern more clearly this environmental interaction than MIM, and thus not regarding it as significant. However, caution must be taken when interpreting mapping results, especially when the analysis is aimed at determining the significance of epistatic and environmental effects, due to the limited population size inherent to a practical development of mapping populations.

A common feature of the comparison was the increasing phenotypic variance explained by the models as more parameters were included. Thus, CIM explained a higher percentage of the variance than IM. Considering epistasis in the model increased the percentage dramatically. For example, daidzein variation explained with IM at BREC was 20.3%. By switching to CIM the percentage was 25.0%. Finally, MIM was able to raise the value to 65.6% by including variation due to epistasis. It might also happen that epistatic and main effects compensate each other to the point of maintaining total isoflavone content between certain levels. That would explain the elevated H^2 observed for the traits. In fact, the sum of the variance that is explained for the additive model for genistein, daidzein and glycitein had a similar value in both locations indicating that even though the set of QTL may differ in a changing environment, the total variance that they account for remains constant. The sum of epistatic interactions, on the other hand, tends to produce very different values when the growing conditions change, which points to epistasis as the main cause of the variation experienced by isoflavones over environments.

Conclusions

The range of values of isoflavones is overwhelming, even for homozygous genotypes growing in the same year and location, which greatly complicates genetic studies. In this sense, model selection was shown to exert a tremendous influence for the detection of main and interacting QTL for complex traits. Consistently, when more source of variance is included, the model is able to highlight more QTL. In addition, although epistasis detection requires a large sample size, our results indicate that all measurements should be taken in a given environment because of the dependency of epistatic interactions to growing conditions. On the other hand, despite that environmental interaction is also possible for main QTL and that QTL effects depend on environment, combining data sources for several locations and within-location replications under the same model (MCIM) may increase the chance for their detection due to the increased statistical power. As a final conclusion, only

considering exhaustive statistical methodologies that include epistasis, together with large mapping populations growing in different environments (years and/or locations), and a dense genetic map with evenly distributed markers, the chance of recovering QTL improves for complex traits.

References

Beggs CJ, Stolzer-Jehle A, Wellmann E (1985) Isoflavonoid formation as an indicator of UV stress in bean (Phaseolus vulgaris L.) leaves: the significance of photorepair in assessing potential damage by increased solar UV-B radiation. Plant Physiol 79(3):630–634

Bennett JO, Yu O, Heatherly LG, Krishnan HB (2004) Accumulation of genistein and daidzein, soybean isoflavones implicated in promoting human health, is significantly elevated by irrigation. Agric Food Chem 52:7574–7579

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

Chew V (1977) Comparisons among treatment means in an analysis of variance. USDA ARS/H/6

Chiari L, Piovesan ND, Naoe LK, Jose IC, Viana JMS, Moreira MA, Barros EG (2004) Genetic parameters relating isoflavone and protein content in soybean seeds. Euphytica 138:55–60

Dhaubhadel S, McGarvey BD, Williams R, Gijzen M (2003) Isoflavonoid biosynthesis and accumulation in developing soybean seeds. Plant Mol Biol 53:733–743

Doerge RW (2001) Mapping and analysis of quantitative trait loci in experimental populations. Nat Rev 3:43–52

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

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait in line crosses using flanking markers. Heredity 69:315–324

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

Juenger TE, Sen S, Stowe KA, Simms EL (2005) Epistasis and genotype–environment interaction for quantitative trait loci affecting flowering time in Arabidopsis thaliana. Genetica 123:87–105

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

Kassem MA, Meksem K, Iqbal MJ, Njiti VN, Banz WJ, Winters TA, Wood A, Lightfoot DA (2004) Definition of soybean genomic regions that control seed phytoestrogen amounts. J Biol Biotechnol 1:52–60

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

Kim JA, Hong SB, Jung WS, Yu CY, Ma KH, Gwag JG, Chung IM (2007) Comparison of isoflavones composition in seed, embryo, cotyledon and seed coat of cooked-with-rice and vegetable soybean (Glycine max L.) varieties. Food Chem 102:738–744

Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLPs linkage maps. Genetics 121:185–199

Lawson CGR, Rolfe BG, Djordjevic MA (1996) Rhizobium inoculation induces condition-dependent changes in the flavonoid composition of root exudates from Trifolium subterraneum. Aust J Plant Physiol 23:93–101

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

Mao Y, Da Y (2004) Statistical power for detecting epistasis QTL effects under the F-2 design. Genet Sel Evol 37:129–150

Mayer M (2005) A comparison of regression interval mapping and multiple interval mapping for linked QTL. Heredity 94:599–605

McKhann HI, Paiva NL, Dixon RA, Hirsch AM (1998) Expression of genes for enzymes of the flavonoid biosynthetic pathway in the early stages of the rhizobium–legume symbiosis. Adv Exp Med Biol 439:45–54

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, MyM Kassem, Hyten DL, Yuang J, Winters TA, Lightfoot DA (2001) Genomic regions that underlie soybean seed isoflavone content. J Biomol Biotechnol 1(1):38–44

Messina MJ (1999) Legumes and soybeans overview of their nutritional profiles and health effects. Am J Clin Nutr 70(Suppl):439-S–450-S

Naoumkina M, Farag MA, Sumner LW, Tang Y, Liu CJ, Dixon RA (2007) Different mechanisms for phytoalexin induction by pathogen and wound signals in Medicago truncatula. Proc Natl Acad USA 104:17909–17915

Ososki AL, Kennelly EJ (2003) Phytoestrogens: a review of the present state of research. Phytother Res 17:845–869

Peters NK, Frost JW, Long SR (1986) A plant flavones, luteolin, induces expression of Rhizobium meliloti nodulation genes. Science 233:977–980

Phillips PC (2008) Epistasis, the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev 9:855–867

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

Stafford HA (1997) Roles of flavonoids in symbiotic and defense functions in legume roots. Bot Rev 63:27–39

Subramanian S, Hu X, Lu G, Odelland JT, Yu O (2004) The promoters of two isoflavone synthase genes respond differentially to nodulation and defense signals in transgenic soybean roots. Plant Mol Biol 54:623–639

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

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

Subramanian S, Stacey G, Yu O (2007) Distinct, crucial roles of flavonoids during legume nodulation. Trends Plant Sci 12:282–285

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

Variyar PS, Limaye A, Sharma A (2004) Radiation-induced enhancement of antioxidant contents of soybean (*Glycine max* Merrill). J Agric Food Chem 52:3385–3388

Vyn TJ, Yin X, Bruulsema TW, Jackson CC, Rajcan I, Brouder SM (2002) Potassium fertilization effects on isoflavone concentrations in soybean (*Glycine max* L. Merr.). J Agric Food Chem 50:3501–3506

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

Whitlock MC (1995) Multiple fitness peaks and epistasis. Annu Rev Ecol Syst 26:601–629

Wu X, Blake S, Sleper DA, Shannon JG, Cregan P, Nguyen HT (2009) QTL, additive and epistatic effects for SCN resistance in PI 437654. Theor Appl Genet 118:1093–1105

Yang J, Zhu J (2005) Methods for predicting superior genotypes under multiple environments based on QTL effects. Theor Appl Genet 110:1268–1274

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

Yi N, Xu S (2002) Mapping quantitative trait loci with epistatic effects. Genet Res Camp 79:185–198

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

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

Zhang J, Subramanian S, Zhang Y, Yu O (2007) Flavone synthases from *Medicago truncatula* are flavanone-2-hydroxylases and are important for nodulation. Plant Physiol 144:741–751