2014

Detection and Confirmation of Quantitative Trait Loci for Soybean Seed Isoflavones

Christopher J. Smallwood
University of Tennessee

Catherine N. Nyinyi
Monsanto, Harrisburg, SD

Dean A. Kopsell
University of Tennessee, dkopsell@utk.edu

Carl E. Sams
University of Tennessee, carlsams@tennessee.edu

Dennis R. West
University of Tennessee

See next page for additional authors

RESEARCH

Detection and Confirmation of Quantitative Trait Loci for Soybean Seed Isoflavones

Christopher J. Smallwood,* Catherine N. Nyinyi, Dean A. Kopsell, Carl E. Sams, Dennis R. West, Pengyin Chen, Stella K. Kantartzi, Perry B. Cregan, David L. Hyten, and Vincent R. Pantalone

ABSTRACT

Interest in soybean [*Glycine max* (L.) Merr.] isoflavones has increased in recent years owing to numerous reported health benefits. Consequently, quantitative trait loci (QTL) detection for marker-assisted breeding for isoflavones is being examined for genetic gains. This study sought to detect QTL for soybean isoflavones in a population of 274 recombinant inbred lines derived from a cross between 'Essex' and 'Williams 82' that were subdivided and tested by maturity (early, mid, and late). The field tests were conducted in three environments in 2009 (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). The population was genotyped with 480 polymorphic single nucleotide polymorphism markers. Isoflavones for each replicate were analyzed by near infrared reflectance spectroscopy, whose prediction equation was based on high performance liquid chromatography. Each maturity test, containing 91 or 92 recombinant inbred lines, was analyzed separately for QTL. In total, 21 QTL were detected: 7 for genistein (chromosomes 5, 6, 9, 13, 17, and 19), 5 for daidzein (chromosomes 5, 6, 9, 13, and 19), 3 for glycitein (chromosomes 6, 9, and 20), and 6 for total isoflavone content (chromosomes 5, 6, 9, 13, and 19). Of these 21 QTL, 12 were confirmed or positional confirmations from other studies. Utilization of these QTL could potentially lead to marker-assisted selection approaches for genetic gains in improving soybean isoflavones.

C.J. Smallwood, D.A. Kopsell, C.E. Sams, D.R. West, and V.R. Pantalone, Dep. of Plant Sciences, Univ. of Tennessee, 2431 Joe Johnson Drive, Knoxville, TN 37996; C.N. Nyinyi, Monsanto, 140 W. Industrial Drive, Harrisburg, SD 57032; P. Chen, Dep. of Crop, Soil, and Environmental Sciences, 115 Plant Science Building, Univ. of Arkansas, Fayetteville, AR 72701; S.K. Kantartzi, Dep. of Plant, Soil and Ag. Systems, Southern Illinois Univ., Carbondale, IL 62901; P.B. Cregan and D.L. Hyten, USDA, ARS, Soybean Genomics and Improvement Lab, Beltsville, MD 20705; D.L. Hyten, present address: Dupont Pioneer, 8305 NW 62nd Ave., PO Box 7060, Johnston, IA 50131. Received 23 May 2013. *Corresponding author (csmallw1@utk.edu).

Abbreviations: CIM, composite interval mapping; E, environment; ETREC, East Tennessee Research and Education Center; G, genotype; HPLC, high performance liquid chromatography; LOD, logarithm of odds; LSMEAN, least squares mean; MAS, marker-assisted selection; Mbp, mega base pair; NIRS, near-infrared reflectance spectroscopy; QTL, quantitative trait loci; RIL, recombinant inbred line; SNP, single nucleotide polymorphism; ST, stem termination.

The soybean [*Glycine max* (L.) Merr.] isoflavones of genistein, daidzein, and glycitein have gained considerable interest in recent years because of their potential benefits to human health. Soybean isoflavones are structurally similar to the hormone estrogen (Setchell, 1998) and can exist as aglycones, glucosides, or glucoside derivatives, including 6"-0-malonyl-esters and 6"-0-acetyl-esters (Wilson, 2004). After synthesis via the phenylpropanoid pathway (Bennett et al., 2004), isoflavones exist predominantly in the malonylglucoside form (Charron et al., 2005) and, on ingestion, are converted into aglycones (Brouns, 2002).

Health benefits associated with soybean isoflavones have been reported, including cancer prevention (Birt et. al., 2001), reduced

Published in Crop Sci. 54:595–606 (2014).
doi: 10.2135/cropsci2013.05.0340

risk for coronary heart disease (Sagara et al., 2003), reduced problems with diabetes and obesity (Bhathena and Velasquez, 2002), improved bone and cardiovascular health, and decreased menopausal symptoms (Brouns, 2002). Soybean isoflavones are also positively correlated with other important traits. Among the traits positively associated with isoflavones are yield (Primomo et al., 2005; Morrison et al., 2008) and aphid resistance (Meng et al., 2011). Linolenic acid, an undesirable fatty acid found in soybean oil, is negatively associated with isoflavone content (Wilson, 2004), suggesting that reduced linolenic acid genotypes would have the added benefit of increased isoflavone levels. The soybean plant also benefits in several areas from isoflavones, including stimulation of soil microbe rhizobia in root nodule formation for nitrogen fixation and antifungal and antipathogenic activity (Ogbuewu et al., 2010).

Improving soybean isoflavone content can be difficult because of its quantitative inheritance nature (Meksem et al., 2001; Kassem et al., 2004; Kassem et al., 2006; Primomo et al., 2005; Zeng et al., 2009). Heritability estimates for soybean isoflavones are inconsistent, ranging from moderate (Primomo et al., 2005; Zeng et al., 2009) to high (Gutierrez-Gonzalez et al., 2009). Gaining a better understanding of the genomic regions that control isoflavone content would be beneficial for making genetic improvements. Several studies have sought to identify quantitative trait loci (QTL), which control isoflavone content. Meksem et al. (2001) identified seven QTL for genistein, daidzein, and glycitein on five chromosomes in a population of 100 recombinant inbred lines (RIL) derived from the cross between 'Essex' and 'Forrest'. Following up on that study, Kassem et al. (2004) used an increased number of molecular markers to screen the same population of 100 RIL and confirmed 6 of the previous QTL, while detecting 2 new QTL. Kassem et al. (2006) revisited the Essex × Forrest population using the composite interval mapping (CIM) technique and only confirmed two QTL from the previous mapping study, while detecting 14 new QTL. Primomo et al. (2005) detected 17 QTL for genistein, daidzein, glycitein, and total isoflavones on nine chromosomes. Of particular interest were five QTL located on nearby genetic regions to those previously detected by Kassem et al. (2004), who used a different population grown in different environments. In a similar study, Zeng et al. (2009) detected 11 QTL for genistein, daidzein, glycitein, and total isoflavones on nine chromosomes. That population of 136 RIL, derived from the Chinese cultivars Zhongdou 27 (high isoflavone parent) and Jiunong 20 (low isoflavone parent), was grown in seven different environments in China. A similar analysis of a cross between 'AC756' (low isoflavone parent) and 'RCAT Angora' (high isoflavone parent) grown at two locations in Ontario, Canada, detected QTL for genistein, glycitein, and total isoflavones in a similar region on chromosome 7 (Primomo et al., 2005). Additional QTL have been detected by Gutierrez-Gonzalez et al. (2009, 2010b, 2011), who examined major and minor QTL for soybean isoflavones as well as epistatic interactions.

Further challenges for soybean isoflavone improvement include the significant impact of environment on seed isoflavone content (Eldridge and Kwolek, 1983; Gutierrez-Gonzalez et al., 2009; Murphy et al., 2009). Environmental factors influencing isoflavone content include temperature and moisture (Lozovaya et al., 2005) as well as soil type (Barion et al., 2010). These challenges reinforce the need to select QTL that are stable across multiple target environments (Bernardo, 2008). Additionally, increasing the efficiency of phenotypic analysis is an important consideration for breeders interested in making selections among a large number of progeny of a population. Another important consideration for soybean isoflavones is QTL detection based on maturity, as genistein, daidzein, and total isoflavones are positively correlated with maturity (Primomo et al., 2005). Once detected, QTL may be used for marker-assisted selection (MAS), especially when confirmed from previous studies. There are several advantages in using MAS as a selection method for some traits where that technology is effective, including increased reliability, improved use of time, and reduced cost in comparison with conventional methods such as selecting on the basis of phenotype (Torres et al., 2010).

To clarify phrasing, confirmed QTL and positional confirmation of QTL will be defined as used in this paper. A QTL is considered confirmed following the standards of the soybean genetics committee (SoyBase and the Soybean Breeder's Toolbox, 2007). Briefly, a confirmed QTL must occur in a population resulting from an original set of meiotic events consisting of one or both parents from the original study, in which the experiment-wise error rate is 0.01 or less. A positional confirmation of QTL in this paper will be considered following the same standards as the confirmed QTL, except that for a positional confirmation of QTL, both parents will be different than the parents from the original detection study.

The objectives of the current research were to: (i) identify soybean isoflavone QTL stable over multiple environments, based on maturity; (ii) confirm or positionally confirm isoflavone QTL from previous studies using the near-infrared reflectance spectroscopy (NIRS) measurement technique; and (iii) examine for phenotypic correlations between isoflavones and other important agronomic and seed quality traits for soybean.

MATERIAL AND METHODS

Plant Materials

A population of 274 RIL was developed from the cross between Essex and 'Williams 82'. Essex is a maturity group V soybean cultivar with a determinate growth habit, purple flower, and gray pubescence (Smith and Camper, 1973), while Williams 82 is a

maturity group III soybean cultivar with indeterminate growth habit, white flower, and tawny pubescence (Bernard and Cremeens, 1988). The seed of Essex and Williams 82 were obtained from the USDA soybean germplasm collection (www.ars-grin.gov, accessed 27 Dec. 2013), and a random single plant of each parental line was intentionally selfed for two generations to provide highly homozygous parental lines to be crossed for RIL development. The initial cross was made in the summer of 2005 at the East Tennessee Research and Education Center (ETREC) in Knoxville, TN. The hybrid seed resulting from the cross were harvested in the fall of 2005 and grown as F_1 single plants in Puerto Rico at the Tropical Agricultural Research Station in Isabela, Puerto Rico, in the winter of 2005–2006. Following the single seed descent method (Brim, 1966), the population was advanced from the F_2 to the F_5 generation. Plant rows from the $F_{5:6}$ generation were grown at ETREC in the summer of 2008 in 3.1-m rows for agronomic data collection and leaf collection for DNA extraction. Each of the 274 $F_{5:7}$ lines was increased in the winter of 2008–2009. In the summer of 2009 the 274 $F_{5:8}$ RIL were grown in replicated yield trials. The 274 RIL were divided into three different tests (early, mid, and late) on the basis of maturity (10 d) from the 2008 ETREC plant row data, with each test containing 91 or 92 RIL. Both parents, along with four other check cultivars or lines, were included in each test. The check lines for the early test ('IA4004', 'LD00-2817P', 'LD00-3309', and 'Macon'), mid test (TN05-4008, TN06-189, TN06-196, and '5002T'), and late test ('JTN-5203', 'Osage', 5002T, and '5601T') were selected to correspond with the appropriate maturity group for their respective test. The 2009 field trials were grown using a randomized complete block design with three replications and grown in three locations (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). Each entry was planted in a plot consisting of two adjacent rows 6.1-m in length, with the rows spaced 0.8-m apart.

Gas Chromatography

Measurements for palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid were done for each replicate and location of the 2009 field tests using gas chromatography with a procedure described by Spencer et al. (2004), using a Hewlett-Packard HP 6890 series gas chromatograph (Agilent Technologies) system equipped with a 7683 auto sampler, a 7673 flame ionization detector, an immobilized 30 m × 0.53 mm inner diameter, and an Agilent DB-23 capillary column with 0.5-μm fused stationary phase.

Near Infrared Reflectance Spectroscopy

Following harvest from each location of the 2009 growing season, approximately 25 g of seed from each plot were ground for 20 sec in a Knifetec 1095 Sample Mill (FOSS Tecator) to produce ground whole soybean with a uniform consistency and particle size. Samples were analyzed using the NIRS instrument (NIR 6500, FOSS North America) as described by Panthee et al. (2006), except that for this study the ground samples were scanned using updated ISIscan software v. 2.85. The equation used for NIRS prediction of genistein, daidzein, and glycitein was derived from high performance liquid chromatography (HPLC) analysis using 497, 499, and 492 samples with R^2 values of 0.85, 0.85, and 0.65, respectively. This produced sample estimates for the soybean isoflavones genistein, daidzein, and glycitein in milligrams per gram seed weight on a dry weight basis. Values for total isoflavones were obtained by summing genistein, daidzein, and glycitein. Values for protein and oil concentration were obtained on a 13% moisture basis and converted to grams per kilogram on a dry weight basis.

DNA Extraction and Molecular Analysis

Samples of DNA were collected from crushed leaves of RIL and parents using the Qiagen Plant DNeasy Extraction Kit (Qiagen). The RIL and parents were screened at USDA-ARS, Beltsville, MD, with 1536 single nucleotide polymorphism (SNP) markers from the Universal Soybean Linkage Panel (U.S.L.P. 1.0) (Hyten et al., 2010). The SNP markers were screened using the GoldenGate assay, which was performed as described by Fan et al. (2003) and Hyten et al. (2008). Polymorphisms were detected in 480 of the SNP markers.

Data Analysis

The data combined over replications and locations for genistein, daidzein, glycitein, and total isoflavones for each maturity test individually were tested for differences among RIL using the MIXED procedure in SAS 9.2 (SAS Institute, 2008). Random blocking factors in the model included environment (E), genotype (G), G × E, and replication. A similar model was fitted replacing the genotype term with a stem termination (ST) term with the exception of the ST × E term, as there was no difference in ST between locations. This model was fitted to test for isoflavone differences between indeterminate and determinate RIL. The CORR procedure was used to obtain phenotypic correlations between genistein, daidzein, glycitein, total isoflavones, and other important agronomic and seed quality traits for soybean, including maturity, lodging, height, yield, protein, oil, palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid. Heritability estimates for genistein, daidzein, glycitein, and total isoflavones were calculated on an entry means basis for three replications and three locations according to Nyquist (1991). Variance components were determined with MIXED in SAS 9.2 (SAS Institute Inc.) using REML estimation.

A genetic map was estimated from this population using R/QTL package (Broman et al., 2003) in the R language and environment for statistical computing (R Development Core Team, 2009). This map, along with the least squares means (LSMEAN) combined over locations for genistein, daidzein, glycitein, and total isoflavones was used for QTL identification with QTL Cartographer software v. 2.5 (Wang et al., 2011) with the CIM procedure. Analyses were conducted with the standard model Zmapqtl 6 in the CIM procedure with a 10 cM window and a 1 cM walk speed. Since this was an F_5–derived population, heterozygote marker loci were excluded from analyses as we were primarily interested in additive genetic effects. The empirical logarithm of odds (LOD) threshold was determined at the 5% level of probability with 1000 permutations for each trait in each maturity test (Churchill and Doerge, 1994). Any QTL whose LOD score exceeded the empirical LOD threshold was considered significant. For QTL that are confirmed or positional confirmations from previous studies, an additional LOD threshold was determined at the 1% level of probability using the same 1000 permutations (Churchill and Doerge, 1994). The LOD thresholds at the 5 and 1% probability levels for

Table 1. Quantitative trait loci (QTL) associated with the isoflavone genistein in three 'Essex' × 'Williams 82' subpopulations separated by maturity consisting of 91 or 92 recombinant inbred lines grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR) in 2009. The chromosome (Chr) and cM position for the QTL listed are from the linkage map created for this population. The closest single nucleotide polymorphism (SNP) marker to the QTL peak, logarithm of odds (LOD) score, LOD-1 interval, R^2 value, and additive effect are given for each QTL. The name and consensus 4.0 map position for the closest SNP marker to each QTL peak is also provided. The consensus map 4.0 positions of each BARC single nucleotide polymorphism marker are available at www.soybase.org/. GEN1-chr6 to GEN7-chr9 represent seven QTL for the isoflavone genistein.

Test	Chr	QTL	QTL position	LOD threshold	LOD threshold	LOD score	LOD-1 interval	Closest marker	Consensus map 4.0 marker position	R^2	Effect†
			cM	1%	5%				cM		
Early	6	GEN1-chr6‡	209.4	4.2	3.3	7.9	208.0–214.0	BARC-023517-05442	103.3	0.16	0.055
	9	GEN2-chr9	95.1			10.5	86.1–103.0	BARC-038909-07393	69.3	0.23	0.058
	13b	GEN3-chr13b	32.4			6.0	26.6–40.3	BARC-030359-06859	65.1	0.14	−0.046
Mid	5	GEN4-chr5	186.3	3.7	3.2	4.0	182.8–189.9	BARC-042853-08438	49.5	0.13	0.034
	17	GEN5-chr17	21.6			3.2	21.0–34.2	BARC-030909-06973	22.4	0.10	−0.029
	19	GEN6-chr19§	137.2			7.6	132.1–148.8	BARC-035235-07156	74.8	0.22	0.044
Late	6	GEN1-chr6‡	205.9	3.6	2.9	3.1	203.9–212.9	BARC-066175-19800	100.9	0.09	0.061
	9	GEN7-chr9	48.3			3.3	48.1–52.1	BARC-048623-10678	39.9	0.09	0.027
	19	GEN6-chr19§	138.8			3.3	137.3–140.9	BARC-016145-02292	75.6	0.09	0.029

†The additive effect with respect to the Essex allele expressed in milligrams genistein per gram of seed on a dry weight basis.

‡Closely associated with E1 locus.

§Closely associated with Dt1 locus.

Table 2. Quantitative trait loci (QTL) associated with the isoflavone daidzein in three 'Essex' × 'Williams 82' subpopulations separated by maturity consisting of 91 or 92 recombinant inbred lines grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR) in 2009. The chromosome (Chr) and cM position for the QTL listed are from the linkage map created for this population. The closest single nucleotide polymorphism (SNP) marker to the QTL peak, logarithm of odds (LOD) score, LOD-1 interval, R^2 value, and additive effect are given for each QTL. The name and consensus 4.0 map position for the closest SNP marker to each QTL peak is also provided. The consensus map 4.0 positions of each BARC SNP marker are available at www.soybase.org/. DAI1-chr6 to DAI5-chr5 represent five QTL for the isoflavone daidzein.

Test	Chr	QTL	QTL position	LOD threshold	LOD threshold	LOD score	LOD-1 interval	Closest marker	Consensus map 4.0 marker position	R^2	Effect†
			cM	1%	5%				cM		
Early	6	DAI1-chr6‡	209.4	4.1	3.3	5.8	208.0–214.1	BARC-023517-05442	103.3	0.13	0.033
	9	DAI2-chr9	93.4			7.1	84.4–99.6	BARC-038909-07393	69.3	0.16	0.032
	13b	DAI3-chr13b	28.7			4.3	24.1–38.8	BARC-030359-06859	65.1	0.11	−0.026
	19	DAI4-chr19§	141.1			4.8	126.5–141.9	BARC-024345-04854	76.5	0.10	0.029
Mid	5	DAI5-chr5	184.2	4.0	3.3	4.8	181.6–188.5	BARC-042853-08438	49.5	0.16	0.023
	19	DAI4-chr19§	144.5			8.0	135.6–149.5	BARC-026069-05243	76.8	0.30	0.032
Late	9	DAI2-chr9	98.0	4.0	3.0	4.1	82.9–120.5	BARC-038909-07393	69.3	0.13	0.019
	19	DAI4-chr19§	138.8			6.3	129.6–148.6	BARC-016145-02292	75.6	0.18	0.024

†The additive effect with respect to the Essex allele expressed in milligrams daidzein per gram of seed on a dry weight basis.

‡Closely associated with E1 locus.

§Closely associated with Dt1 locus.

each maturity test are listed for genistein (Table 1), daidzein (Table 2), glycitein (Table 3), and total isoflavones (Table 4).

RESULTS AND DISCUSSION

Phenotypic Traits

There were significant differences among RIL, locations, reps, and location × RIL interactions for genistein, daidzein, glycitein, and total isoflavones in the early, mid, and late tests ($P < 0.001$). In each test, Essex had a numerically higher value for genistein, daidzein, and total isoflavones, whereas Williams 82 had a higher value for glycitein (Table 5). For each isoflavone in the early test and mid test, and for glycitein in the late test, transgressive segregation was observed, in which a large number of individual RIL values were higher or lower than either parent. This suggests that both parents carry alleles governing isoflavones heritable in progeny. In every test, daidzein was the most abundant isoflavone, followed by genistein and glycitein, which represents a rank order for isoflavones similar to some previous studies (Kassem et al., 2006; Zeng et al., 2009). This differs

Table 3. Quantitative trait loci (QTL) associated with the isoflavone glycitein in three 'Essex' × 'Williams 82' subpopulations separated by maturity consisting of 91 or 92 recombinant inbred lines grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR) in 2009. The chromosome (Chr) and cM position for the QTL listed are from the linkage map created for this population. The closest single nucleotide polymorphism (SNP) marker to the QTL peak, logarithm of odds (LOD) score, LOD-1 interval, R^2 value, and additive effect are given for each QTL. The name and consensus 4.0 map position for the closest SNP marker to each QTL peak is also provided. The consensus map 4.0 positions of each BARC SNP marker are available at www.soybase.org/. GLY1-chr6 to GLY3-chr20 represent three QTL for the isoflavone glycitein.

Test	Chr	QTL	QTL position	LOD threshold	LOD threshold	LOD score	LOD-1 interval	Closest Marker	Consensus map 4.0 marker position	R^2	Effect†
			cM	1%	5%				cM		
Early	6	GLY1-chr6‡	200.0	3.9	3.3	4.2	187.9–206.4	BARC-031337-07051	97.1	0.15	−0.003
Mid	9	GLY2-chr9	77.6	4.0	3.1	4.6	65.3–94.5	BARC-014813-01678	47.4	0.20	0.002
	20	GLY3-chr20	105.0			3.5	90.0–113.5	BARC-053725-11957	70.9	0.17	−0.002
Late		NONE		3.9	3.2						

†The additive effect with respect to the Essex allele expressed in milligrams glycitein per gram of seed on a dry weight basis.

‡Closely associated with E1 locus.

Table 4. Quantitative trait loci (QTL) associated with total isoflavones in three 'Essex' × 'Williams 82' subpopulations separated by maturity consisting of 91 or 92 recombinant inbred lines grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR) in 2009. The chromosome (Chr) and cM position for the QTL listed are from the linkage map created for this population. The closest single nucleotide polymorphism (SNP) marker to the QTL peak, logarithm of odds (LOD) score, LOD-1 interval, R^2 value, and additive effect are given for each QTL. The name and consensus 4.0 map position for the closest SNP marker to each QTL peak is also provided. The consensus map 4.0 positions of each BARC SNP marker are available at www.soybase.org/. ISO1-chr6 to ISO6-chr5 represent six QTL for total isoflavones.

Test	Chr	QTL	QTL Position	LOD Threshold	LOD Threshold	LOD score	LOD-1 Interval	Closest Marker	Consensus map 4.0 marker position	R^2	Effect†
			cM	1%	5%				cM		
Early	6	ISO1-chr6‡	209.4	4.0	3.3	6.4	208.2–213.6	BARC-023517-05442	103.3	0.13	0.085
	9	ISO2-chr9	94.7			9.3	86.3–102.5	BARC-038909-07393	69.3	0.21	0.092
	13b	ISO3-chr13b	31.4			4.8	25.0–41.2	BARC-030359-06859	65.1	0.11	−0.068
	19	ISO4-chr19§	141.1			4.5	123.7–142.0	BARC-024345-04854	76.5	0.09	0.068
Mid	5	ISO5-chr5	187.2	3.9	3.3	4.3	182.6–191.7	BARC-059081-15595	55.2	0.09	0.045
	19	ISO4-chr19§	137.3			6.1	130.5–148.5	BARC-035235-07156	74.8	0.18	0.062
Late	5	ISO6-chr5	26.6	4.0	3.1	3.3	14.7–50.6	BARC-019415-03923	17.6	0.09	0.041
	19	ISO4-chr19§	138.8			3.9	130.1–148.4	BARC-016145-02292	75.6	0.11	0.049

†The additive effect with respect to the Essex allele expressed in milligrams total isoflavones per gram of seed on a dry weight basis.

‡Closely associated with E1 locus.

§Closely associated with Dt1 locus.

from many previous studies, in which genistein is the most abundant isoflavone (Eldridge and Kwolek, 1983; Wang and Murphy, 1994; Brouns, 2002; Primomo et al., 2005; Gutierrez-Gonzalez et al., 2009, 2010b, 2011). Heritability estimates were highest for the early test, with values of 0.81, 0.81, 0.54, and 0.81 for genistein, daidzein, glycitein, and total isoflavones, respectively (Table 5), which were similar to the values obtained by Gutierrez-Gonzalez et al. (2009, 2010b, 2011). The mid test had similar but slightly lower heritability than the early test for genistein, daidzein, glycitein, and total isoflavones, with values of 0.68, 0.72, 0.36, and 0.71, respectively (Table 5). The late test had heritability values of 0.46, 0.39, 0.48, and 0.43 for genistein, daidzein, glycitein, and total isoflavones, respectively (Table 5), which were similar to the values obtained by Primomo et al. (2005) and Zeng et al. (2009). The large differences in heritability estimates among the maturity tests illustrate that while much of the variation may be genetic, as evidenced by the higher heritability estimates in the early and the mid tests (Table 5), a large degree of variation is still affected by external environmental factors, as notable in the late test (Table 5). A possible explanation for the differences in heritability between maturity tests could be length of exposure to environmental conditions. As each test was harvested at maturity, the early test was removed from the fields first and, therefore, received less interaction with environmental factors affecting seed isoflavone accumulation than the mid and late tests. Similarly, the late test had more time to interact with environmental factors such as moisture and temperature fluctuation, which may have caused greater environmental variability, resulting in lower heritability for each trait.

Table 5. Descriptive statistics and heritability values for three subpopulations separated by maturity consisting of 91 or 92 recombinant inbred lines derived from 'Essex' and 'Williams 82', grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR) with three replications in 2009. Parental least squares means (LSMEAN) and determinate (Det) and indeterminate (Indet) stem termination LSMEAN also included.

Test	Trait	Min†	Mean†	Max†	Std. Dev.	Essex†	Williams 82†	Det†	Indet†	*h*2‡
Early§	Genistein	0.44	0.69	0.96	0.12	0.77	0.74	0.81	0.65	0.81
	Daidzein	0.62	0.8	0.96	0.08	0.86	0.84	0.88	0.77	0.81
	Glycitein	0.15	0.17	0.19	0.01	0.16	0.18	0.17	0.17	0.54
	Total Isoflavones	1.21	1.66	2.05	0.2	1.8	1.75	1.85	1.59	0.81
Mid¶	Genistein	0.51	0.75	0.93	0.09	0.84	0.62	0.78	0.72	0.68
	Daidzein	0.69	0.83	0.93	0.06	0.84	0.79	0.85	0.80	0.72
	Glycitein	0.15	0.16	0.18	0.01	0.16	0.17	0.16	0.16	0.36
	Total Isoflavones	1.39	1.74	2.02	0.15	1.84	1.58	1.80	1.69	0.71
Late#	Genistein	0.4	0.68	0.89	0.08	0.7	0.42	0.74	0.66	0.46
	Daidzein	0.61	0.78	0.93	0.05	0.78	0.64	0.82	0.77	0.39
	Glycitein	0.15	0.16	0.18	0.01	0.16	0.16	0.16	0.16	0.48
	Total Isoflavones	1.16	1.62	1.98	0.13	1.64	1.22	1.72	1.58	0.43

†LSMEAN value expressed in mg g^{-1} seed on a dry weight basis.

‡Heritability on an entry-mean basis.

§Least significant difference for mean separation of genistein, daidzein, glycitein, and total isoflavones in the early test were 0.150, 0.097, 0.016, and 0.243, respectively.

¶Least significant difference for mean separation of genistein, daidzein, glycitein, and total isoflavones in the mid test were 0.146, 0.089, 0.013, and 0.230, respectively.

#Least significant difference for mean separation of genistein, daidzein, glycitein, and total isoflavones in the late test were 0.170, 0.119, 0.013, and 0287, respectively.

Table 6. Phenotypic correlations between genistein, daidzein, glycitein, and total isoflavones with agronomic and seed quality traits of interest in soybean [*Glycine max* (L.) Merr.] from a population of 274 recombinant inbred lines derived from 'Essex' and 'Williams 82', grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR) with three replications in 2009.

	Daidzein	Glycitein	Total	Maturity	Lodging	Height	Yield	Protein	Oil	Palmitic Acid	Stearic Acid	Oleic Acid	Linoleic Acid	Linolenic Acid
Genistein	0.95	0.11	0.99	0.38	−0.17	−0.23	0.39	−0.74	−0.38	0.34	−0.42	−0.53	0.50	0.51
P value	***	NS	***	***	**	***	***	***	***	***	***	***	***	***
Daidzein		0.21	0.98	0.24	−0.25	−0.34	0.28	−0.64	−0.38	0.40	−0.37	−0.59	0.56	0.50
P value		***	***	***	***	***	***	***	***	***	***	***	***	***
Glycitein			0.19	−0.50	−0.21	−0.22	−0.28	0.04	0.30	0.17	0.31	0.12	−0.16	−0.18
P value			**	***	***	***	***	NS	***	**	***	*	**	**
Total				0.30	−0.21	−0.28	0.34	−0.70	−0.37	0.37	−0.39	−0.55	0.52	0.50
P value				***	***	***	***	***	***	***	***	***	***	***

*Significant at 0.05 probability level.

**Significant at 0.01 probability level.

***Significant at 0.001 probability level.

NS, not significant at 0.05 probability level.

For each maturity test, there were significant differences ($P < 0.05$) between determinate and indeterminate RIL for genistein, daidzein, glycitein, and total isoflavone, with the exception of glycitein in the mid test. Numerically, the trends for stem termination were quite similar to those for the parental lines, with determinate RIL displaying a higher value in each maturity test for genistein, daidzein, and total isoflavones (Table 5). This was in common with the determinate parent Essex displaying a higher value for these traits than the indeterminate parent Williams 82. Glycitein did not follow this trend as closely, as indeterminate RIL displayed a higher value in the early test and the mid test, while determinate RIL displayed a higher value in the late test (Table 5).

Phenotypic correlations were obtained using the LSMEAN for all 274 RIL from the population. Strong positive correlations ($R = 0.95$, $P < 0.0001$) were found between each of the soybean isoflavones with each other, with the exception of glycitein (Table 6). Genistein, daidzein, and total isoflavones displayed low to moderate positive correlations with maturity, yield, palmitic acid, linoleic acid, and linolenic acid ($R = 0.24$, $P < 0.0001$). Weak negative correlations were observed between genistein, daidzein, and total isoflavones with lodging and height ($R = -0.17$, $P < 0.01$), but moderate negative correlations were observed with protein, oil, stearic acid, and oleic acid ($R = -0.37$, $P < 0.0001$). The significant positive correlations with yield and significant negative correlations with seed oil for genistein, daidzein, and total isoflavones were consistent with the results from Primomo et al. (2005). Correlations between glycitein and other traits were either very weak ($R < 0.30$) or insignificant at the 5% level of probability, with

the exception of maturity, which was moderately negatively correlated ($R = -0.50$, $P < 0.0001$), and oil, which was weakly positively correlated ($R = 0.30$, $P < 0.0001$).

The correlations between genistein, daidzein, and total isoflavones with other important agronomic and seed quality traits were largely undesirable, with the exception of the positive correlation with yield and the negative correlation with lodging and stearic acid. For soybean, reduced seed protein content, oil content, and oleic acid content are undesirable, while earlier maturity and reduced palmitic acid and linolenic acid are often targeted goals of improvement. Researchers working to improve these significant quality traits may inadvertently decrease seed isoflavone content to obtain strategic objectives. Future studies should continue research to better understand the relationships among soybean isoflavones and other important agronomic and seed quality traits as breeders strive to optimize improvements.

Genetic Mapping and QTL Detection

The 480 polymorphic SNP markers used to create a genetic map for this population spanned 21 linkage groups (Figs. S1, S2, S3, and S4). The map covered a total distance of 3035.4 cM, with an average distance of 6.32 cM between markers. While the overall map distance was greater than the Consensus Map 4.0 (Hyten et al., 2010), the SNP order and linkage group assignments were similar. The exceptions to this similarity were that for this study, chromosome 13 was split into two linkage groups (designated 13a and 13b), and one marker, BARC-015435-01966, was mapped to chromosome 13 rather than chromosome 6.

In total, seven QTL for genistein (Table 1) (Figs. S5, S6, and S7), five QTL for daidzein (Table 2) (Figs. S8, S9, and S10), three QTL for glycitein (Table 3) (Figs. S11 and S12), and six QTL for total isoflavones (Table 4) (Figs. S13, S14, and S15) were detected. The QTL detected in this study could have been affected by the E1 maturity locus (chromosome 6) and the Dt1 growth habit locus (chromosome 19), both of which segregated in this population. Also, the E3 maturity locus on chromosome 19 may have caused an effect on the QTL detected in this population, as it is closely linked with the Dt1 growth habit locus (Watanabe et al., 2009). However, previous studies have reported QTL for genistein, glycitein, daidzein, and total isoflavones on chromosome 6 (Primomo et al., 2005; Zeng et al., 2009; Gutierrez-Gonzalez et al., 2011) and chromosome 19 (Gutierrez-Gonzalez et al., 2009, 2010b) in nearby genetic regions to the E1 locus and the Dt1 locus, respectively. Furthermore, soybean isoflavones can be associated with maturity and plant height (Primomo et al., 2005). With these considerations, the QTL detected for genistein, daidzein, glycitein, and total isoflavones on chromosomes 6 and 19 will be considered as actual effects for isoflavones in this study. Yet it should be noted that nearby E1 and Dt1 genes on chromosomes 6 and 19, respectively, have a major effect on soybean growth and development. As previously noted, the stem termination values for genistein, daidzein, and total isoflavones followed the same trends as the parental values, with determinate RIL exceeding indeterminate RIL and Essex (determinate parent) exceeding Williams 82 (indeterminate parent) (Table 5). Isoflavone QTL on chromosomes 6 and 19 may be pleiotropic effects of the E1 and Dt1 loci, respectively.

The QTL for genistein, designated GEN1-chr6 through GEN7-chr9, were located on chromosomes 5, 6, 9, 13b, 17, and 19 (Table 1) (Figs. S5, S6, and S7). In general, QTL detected on the same chromosome in different maturity tests were considered to be the same if the additive effect was in the same direction and they were located within 10 cM of each other, corresponding to the window size for QTL detection. However, the QTL detected on chromosome 9 in the early test was considered to be different from the QTL on chromosome 9 in the late test because the distance between them exceeded 40 cM (Figs. S5 and S7). The R^2 values for QTL detected for genistein in this study ranged from 0.09 to 0.23 (Table 1), representing large genetic effects. Four of the QTL detected for genistein, located on chromosomes 5, 9, and 17, have not been previously reported. Of these, the QTL GEN2-chr9 on chromosome 9 detected in the early test had the largest R^2 value (0.23), which represents a putative new major QTL for genistein that could be useful for MAS. Examination of the Consensus Map 4.0 (Hyten et al., 2010) revealed that the QTL for genistein on chromosomes 6 (GEN1-chr6) (Primomo et al., 2005; Zeng et al., 2009; Gutierrez-Gonzalez et al., 2011), 13 (GEN3-chr13b) (Zeng et al., 2009; Gutierrez-Gonzalez et al., 2009, 2010b), and 19 (GEN6-chr19) (Gutierrez-Gonzalez et al., 2009, 2010b) have been reported previously and are supported by our findings. The studies conducted by Primomo et al. (2005) and Zeng et al. (2009) were conducted in extremely different environments using different parents than the current study. Thus, finding the same QTL under rather different conditions represents a positional confirmation. The studies by Gutierrez-Gonzalez et al. (2009, 2010b) as well as this study used Essex as a parent in population development, representing the confirmation of QTL from genetically similar populations grown in different sets of environments. However, in this study, Essex was the high isoflavone parent, whereas in Gutierrez-Gonzalez et al. (2009, 2010b), Essex was the low isoflavone parent. In Gutierrez-Gonzalez et al. (2009, 2010b), the presence of the Essex allele resulted in a decrease from the population mean for the genistein QTL on chromosome 13 and an increase over the population mean for the QTL on chromosome 19. The same effects for the Essex alleles were observed in this study for the genistein QTL detected on chromosomes 13b and 19, for which

we suggest the confirmed QTL symbols cqGenistein-001 and cqGenistein-002, respectively. This consistency in the effect of the Essex allele for the QTL on chromosomes 13 and 19 across studies could have implications for MAS. The reoccurrence of these QTL in different environments, but with a common parent, demonstrates a consistency that is important when considering heritable QTL for MAS.

On the basis of our results for a fully additive genetic model, MAS assembling QTL in the allelic form for higher genistein within each maturity test could potentially increase that isoflavone by 19, 9, and 7% above the mean value in the early, mid, and late tests, respectively. This would provide a targeted prediction of 0.82 mg g^{-1} and 0.73 mg g^{-1} genistein for the early and late tests, respectively, both of which exceed the values of the high parent (Essex) in their tests. The predicted value for MAS in the mid test at 0.82 mg g^{-1} was similar to the value for Essex (0.84 mg g^{-1}). For the early, mid, and late tests, the predicted improvement by MAS of genistein is below the maximum value, which indicates that further genetic variation for genistein could exist in this population that was not accounted for by the detected QTL.

The QTL detected for daidzein, designated as DAI1-chr6 through DAI5-chr5 in this study, were located on chromosomes 5, 6, 9, 13, and 19 (Table 2) (Figs. S8, S9, and S10). As with genistein, QTL detected on the same chromosome in different maturity tests were considered to be the same if the additive effect was in the same direction and they were located within 10 cM of each other. All the QTL detected for daidzein were major, with R^2 values ranging from 0.10 to 0.30. Both the smallest (0.10) and the largest (0.30) R^2 values detected for daidzein were for the same QTL, DAI4-chr19, detected in different maturity tests. This difference could represent the effect of maturity on the expression of daidzein. There were four QTL detected for daidzein in the early test, of which DAI4-chr19 had the lowest R^2 value (0.10), and two QTL detected for daidzein in the mid test, of which DAI4-chr19 had the highest R^2 value (0.30). Of the QTL detected for daidzein, only the one located on chromosome 5 had not been previously reported. Examination of the Consensus Map 4.0 (Hyten et al., 2010) revealed that QTL in similar genetic regions to those detected in this study on chromosomes 6 (Gutierrez-Gonzalez et al., 2011), 9 (Kassem et al., 2004), 13 (Primomo et al., 2005; Zeng et al., 2009), and 19 (Gutierrez-Gonzalez et al., 2009) had all been identified by previous research efforts and are positional confirmations in this study. Moreover, Gutierrez-Gonzalez et al. (2009) and Kassem et al. (2004) also used Essex as a parent in population development. As in this study, the Essex allele for the QTL detected on chromosomes 9 (Kassem et al., 2004) and 19 (Gutierrez-Gonzalez et al., 2009) contributed to an increase over the population mean value. We propose the confirmed QTL symbols cqDaidzein-001 and cqDaidzein-002 for these QTL on chromosomes 9 and 19, respectively. The confirmation of these QTL from previous studies, in particular those in which Essex was used as a common parent, are promising when considering MAS for daidzein, which is quite susceptible to environmental variation.

Using our results for a fully additive genetic model, an improvement of 13, 4, and 2% for daidzein over the mean value in the early, mid, and late tests, respectively, could be expected if MAS for daidzein QTL were employed. This would provide predicted values of 0.90, 0.87, and 0.80 mg g^{-1} daidzein in the early, mid, and late tests, respectively; each is greater than the high parent (Essex) value but less than the maximum value for daidzein in their respective tests, suggesting that further genetic variation for daidzein exists in this population that was not accounted for by the detected QTL.

The QTL detected for glycitein, designated as GLY1-chr6 through GLY3-chr20, were located on chromosomes 6, 9, and 20 (Table 3) (Figs. S11 and S12). The R^2 values for glycitein QTL ranged from 0.15 to 0.20, indicative of major QTL. Fewer QTL were detected for glycitein than for any of the other isoflavones, with none being detected in the late maturity test. This may be the result of the lower heritability values and lower accuracy of the NIRS prediction equation (R^2 = 0.65) for glycitein in relation to the other isoflavones. Examination of the Consensus Map 4.0 (Hyten et al., 2010) revealed that glycitein QTL in nearby genetic regions to those detected in this study on chromosome 6 (Gutierrez-Gonzalez et al., 2010b) and chromosome 9 (Gutierrez-Gonzalez et al., 2009) have been previously reported in studies containing Essex as a parent. Gutierrez-Gonzalez et al. (2010b) reported that the Essex allele for the glycitein QTL on chromosome 6 caused a decrease from the population mean, which was consistent with the results from this study, so we propose the confirmed QTL symbol cqGlycitein-001 for this QTL. However, in contrast to this study, Gutierrez-Gonzalez et al. (2009) found the Essex allele to cause a reduction in the population mean value for glycitein for the QTL on chromosome 9, thus representing a positional confirmation rather than a confirmation.

Using a fully additive genetic model to estimate MAS improvement for glycitein, an increase of 1% in both the early and the mid tests could be expected. These predicted improvements over the mean are still below the high parent (Williams 82) values in each test. No potential for increase would be expected using MAS in the late test, as no glycitein QTL were detected. The use of MAS for glycitein in this population does not appear to be an effective method. Rather than seeking to improve glycitein content alone, it may be more useful to seek glycitein improvement as a component of total isoflavone content.

Since each of the isoflavones has beneficial effects for humans, the identification of QTL for total isoflavone content is perhaps the most important information to obtain for use in genetic improvement. Because the two major isoflavones, genistein and daidzein, are strongly correlated with each other as well as total isoflavones (Table 6), improvement of total isoflavones would result in the increase of genistein and daidzein. The QTL detected for total isoflavone content in this study, designated as ISO1-chr6 through ISO6-chr5, were located on chromosomes 5, 6, 9, 13, and 19 (Table 4) (Figs. S13, S14, and S15). As with genistein and daidzein, QTL detected on the same chromosome in different maturity tests were considered to be the same if the additive effect was in the same direction and they were located within 10 cM of each other. Like daidzein, which was the largest contributor for total isoflavones, four QTL for total isoflavones were detected in the early test, while only two QTL were detected in the mid and late tests. The heritability values were once again much lower in the late test (0.43) than in the early (0.81) or mid (0.71) tests. For the QTL detected for total isoflavone content, the R^2 values ranged from 0.09 to 0.21. Both of the QTL on chromosome 5 and the QTL on chromosome 9 have not been previously reported. Of these, the QTL on chromosome 9 is notable, as it explained the greatest variation in total isoflavone content. Examination of the Consensus Map 4.0 (Hyten et al., 2010) showed that total isoflavone QTL in nearby genetic regions to those detected in this study on chromosomes 6 (Primomo et al., 2005; Gutierrez-Gonzalez et al., 2011), 13 (Zeng et al., 2009), and 19 (Gutierrez-Gonzalez et al., 2009, 2010b) had been detected previously, representing a positional confirmation. This study, as well as Gutierrez-Gonzalez et al. (2009, 2010b), used Essex as a parent. For the QTL detected on chromosome 19 by Gutierrez-Gonzalez et al. (2009, 2010b), the Essex allele contributed to an increase over the population mean, which was consistent with the results from this study. This QTL for total isoflavone content (ISO4-chr19) may be useful for MAS as it has been also reported in previous studies (Gutierrez-Gonzalez et al., 2009, 2010b), it was detected in each maturity test in this study, and a similar QTL was detected in approximately the same genetic region for both genistein (GEN6-chr19) and daidzein (DAI4-chr19). We propose the confirmed QTL symbol of cqIsoflv-001 for this total isoflavone QTL.

A fully additive genetic model estimating the effects of MAS for total isoflavones in this population would predict values exceeding the mean by 15, 4, and 2%, in the early, mid, and late tests, respectively. These improvements would result in predicted values of 1.91, 1.82, and 1.66 mg g^{-1} for the early, mid, and late tests, respectively, which are greater than the mean value in every instance. These potential improvements using MAS are greater than the high parent (Essex) value in the early and late tests but less than Essex in the mid test. The early test had a higher heritability value for total isoflavone content and more QTL detected than either the mid or the late test, so it is consistent with expectation that the subset of RIL genotypes in the early test would exhibit the greatest potential for improvement through MAS. However, in each test, the maximum value for total isoflavone content exceeded the predicted MAS value. Greater genetic variation for total isoflavone content may exist in this population than was accounted for by the detected QTL. Moreover, some of the differences between the maximum value observed and the highest value predicted by MAS can be explained by the environmental factors that contribute to isoflavone variation.

Many of the QTL detected in this study for different isoflavones existed in nearby genetic regions. For example, five of the QTL detected for genistein (GEN1-chr6, GEN2-chr9, GEN3-chr13b, GEN4-chr5, and GEN6-chr19) were in nearby genetic regions to QTL detected for daidzein (DAI1-chr6, DAI2-chr9, DAI3-chr13b, DAI5-chr5, and DAI4-chr19) and for total isoflavones (ISO1-chr6, ISO2-chr9, ISO3-chr13b, ISO5-chr5, and ISO4-chr19). As genistein and daidzein are the primary components of total isoflavones in soybean, it seems evident that the QTL for genistein or daidzein would be the same as those detected for total isoflavones. For comparison of the QTL detected for genistein, daidzein, and glycitein, examination of the phenylpropanoid pathway from which they are synthesized helps to shed some light on the results (Fig. 1). Synthesis of each of the isoflavones have many steps in common, so QTL detected in nearby genetic regions for different isoflavones could possibly represent genes controlling enzyme expression for common steps in the phenylpropanoid pathway.

Identifying markers that are associated with specific enzymes in the phenylpropanoid pathway will be an important step in the genetic improvement of soybean isoflavones. Among the more influential enzymes to be targeted are phenylalanine ammonia lyase, which is the first committed step in the phenylpropanoid pathway, and isoflavone synthase, which is a key step in the biosynthesis of isoflavones (Du et al., 2010). Other important enzymes that would be expected to influence isoflavone concentration include cinnamate-4-hydroxylase, 4-coumarate CoA ligase, chalcone synthase, chalcone reductase, and chalcone isomerase (Fig. 1). The isoflavone synthase 1 gene and the isoflavone synthase 2 gene have been mapped to chromosomes 7 and 13, respectively (Cheng et al., 2010). While none of the QTL detected in this study were located on chromosome 7, there were QTL detected for genistein (Table 1), daidzein (Table 2), and total isoflavones (Table 4) on chromosome 13, which could possibly be associated with the isoflavone synthase 2 gene.

A more in-depth look into the location of candidate genes in relation to SNP markers closely associated with

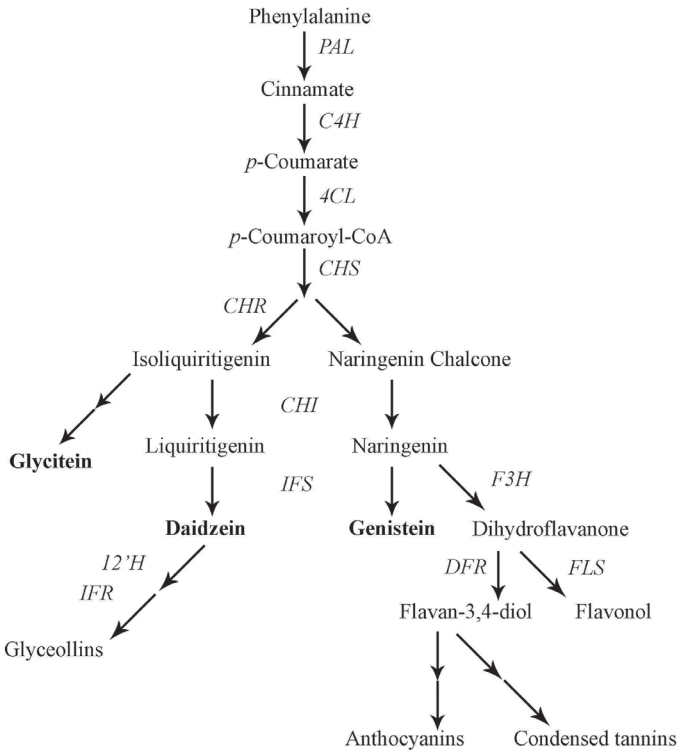

Fig. 1. Diagram of the phenylpropanoid pathway adapted from Gutierrez-Gonzalez et al. (2010a). Enzymes used in pathway: phenylalanine ammonia lyase (PAL), cinnamic acid 4-hydroxylase (C4H), 4-coumarate:CoA ligase (4CL), chalcone synthase (CHS), chalcone reductase (CHR), chalcone isomerase (CHI), isoflavone synthase (IFS), flavanone 3-hydroxylase (F3H), dihydrofl avonol reductase (DFR), isoflavone reductase (IFR), flavonol synthase (FLS), and isoflavone hydroxylase (12'H).

QTL detected in this study reveals some interesting insight. One QTL each for genistein (GEN4-chr5), daidzein (DAI5-chr5), and total isoflavones (ISO5-chr5) is located less than 1 mega base pair (Mbp) from a gene with possible influence on soybean isoflavones. Using the SNP marker genomic sequence position (SoyBase and the Soybean Breeder's Toolbox, 2013) for the QTL for genistein (Table 1), daidzein (Table 2), and total isoflavones (Table 4), it was determined that GEN4-chr5, DAI5-chr5, and ISO5-chr5 all are located less than 0.7 Mbp from the chalcone synthase 2 gene (National Center for Biotechnology Information, 2013). This relative closeness in proximity could indicate that these QTL represent a genetic difference for the chalcone synthase 2 gene that partially explains the differences for genistein, daidzein, and total isoflavones in this population. Additionally, comparison of the SNP marker genomic sequence position (SoyBase and the Soybean Breeder's Toolbox, 2013) for GEN3-chr13b, DAI3-chr13b, and ISO3-chr13b for the QTL for genistein (Table 1), daidzein (Table 2), and total isoflavones (Table 4) revealed that these QTL are located less than 4.7 Mpb from the isoflavone synthase 2 gene (National Center for Biotechnology Information, 2013). While this represents a greater genetic distance than the QTL associated with the chalcone synthase 2 gene, these QTL could still be associated with the isoflavone synthase 2 gene, which would help to explain the variation for these isoflavones in this population.

CONCLUSION

Overall, 21 QTL for genistein, daidzein, glycitein, or total isoflavones were detected by this study, of which 12 have been previously identified. The QTL detected in this study are important for several reasons. Selections for another trait, such as yield, could be made while simultaneously selecting QTL for isoflavone improvement. However, increased isoflavone content may occur coincidentally when selecting for yield because of the positive correlation between these two traits. Perhaps more important would be using isoflavone QTL to make simultaneous selections with a negatively correlated trait such as protein, so that improvements could be made, or at least losses could be minimized in both traits. Additionally, QTL could be used for selection at offsite nurseries, making the breeding process more efficient.

The parents chosen for this study both have significance, as Essex has been used for several previous isoflavone detection studies (Meksem et al., 2001; Kassem et al., 2004; Kassem et al., 2006; Gutierrez-Gonzalez et al., 2009, 2010b), and Williams 82 was the genotype that was sequenced to obtain the soybean whole genome sequence (Schmutz et al., 2010). In addition, Essex is a prominent ancestor of modern southern U.S. cultivars, and 'Williams', from which Williams 82 was derived, is a prominent ancestor of modern northern U.S. cultivars (Hyten et al., 2004). The detection of 12 previously reported QTL for genistein, daidzein, glycitein, and total isoflavones grown in different environments, using different as well as similar parents is noteworthy in identifying consistent genetic regions controlling isoflavone content. As environmental variation for isoflavones is a significant hurdle for making genetic improvements, the confirmation and positional confirmation of QTL in diverse environments is particularly useful for capturing genetic improvements. The importance of confirmation and positional confirmation for these 12 QTL is further illustrated in that most previous studies used HPLC analysis for isoflavone detection, while this study used the much faster and lower cost method of NIRS. The QTL detected by this study could be beneficial for MAS, resulting in the genetic improvement of soybean isoflavones.

Acknowledgments

Funding from the Tennessee Soybean Promotion Board is greatly appreciated for the support of this project. We thank the United Soybean Board for their support in the development of the genetic mapping population that made this work possible. Support from several research stations was paramount in the completion of this research. We would like to thank the Illinois Soybean Center and the Arkansas Agricultural Experiment Station for their valuable contributions. The Tennessee Agricultural Experiment Station is commended for maintaining excellent

field research facilities for soybean breeding and genetics. We appreciate the dedication of Dr. Bobby Simpson, Director of the East Tennessee Research and Education Center, who encourages our soybean genetics research every year. Contributions made by Deborah Landau-Ellis are gratefully acknowledged.

References

Barion, G., M. Hewidy, G. Mosca, and T. Vamerali. 2010. Intraspecific variability for soybean cotyledon isoflavones in different cropping and soil conditions. Eur. J. Agron. 33:63–73. doi:10.1016/j.eja.2010.03.002

Bennett, J.O., O. Yu, L.G. Heatherly, and H.B. Krishnan. 2004. Accumulation of genistein and daidzein, soybean isoflavones implicated in promoting human health, is significantly elevated by irrigation. J. Agric. Food Chem. 52:7574–7579. doi:10.1021/jf049133k

Bernard, R.L., and C.R. Cremeens. 1988. Registration of 'Williams 82' soybean. Crop Sci. 28:1027–1028.

Bernardo, R. 2008. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. Crop Sci. 48:1649–1664. doi:10.2135/cropsci2008.03.0131

Bhathena, S.J., and M.T. Velasquez. 2002. Beneficial role of dietary phytoestrogens in obesity and diabetes. Am. J. Clin. Nutr. 76:1191–1201.

Birt, D.F., S. Hendrich, and W.Q. Wang. 2001. Dietary agents in cancer prevention: Flavonoids and isoflavonoids. Pharmacol. Ther. 90:157–177. doi:10.1016/S0163-7258(01)00137-1

Brim, C.A. 1966. A modified pedigree method of selection in soybeans. Crop Sci. 6:220. doi:10.2135/cropsci1966.0011183X000600020041x

Broman, K.W., H. Wu, Ś. Sen, and G.A. Churchill. 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890. doi:10.1093/bioinformatics/btg112

Brouns, F. 2002. Soya isoflavones: A new and promising ingredient for the health foods sector. Food Res. Int. 35:187–193. doi:10.1016/S0963-9969(01)00182-X

Charron, C.S., F.L. Allen, R.D. Johnson, V.R. Pantalone, and C.E. Sams. 2005. Correlations of oil and protein with isoflavone concentration in soybean [*Glycine max* (L.) Merr.]. J. Agric. Food Chem. 53:7128–7135. doi:10.1021/jf050610o

Cheng, H., H. Yang, D. Zhang, J. Gai, and D. Yu. 2010. Polymorphisms of soybean isoflavone synthase and flavanone 3-hydroxylase genes are associated with soybean mosaic virus resistance. Mol. Breed. 25:13–24. doi:10.1007/s11032-009-9305-8

Churchill, G.A., and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963–971.

Du, H., Y. Huang, and Y. Tang. 2010. Genetic and metabolic engineering of isoflavanoid biosynthesis. Appl. Microbiol. Biotechnol. 86:1293–1312. doi:10.1007/s00253-010-2512-8

Eldridge, A.C., and W.F. Kwolek. 1983. Soybean isoflavones—effect of environment and variety on composition. J. Agric. Food Chem. 31:394–396. doi:10.1021/jf00116a052

Fan, J.B., A. Oliphant, R. Shen, B.G. Kermani, F. Garcia, K.L. Gunderson, M. Hansen, F. Steemers, S.L. Butler, P. Deloukas, et al. 2003. Highly parallel SNP genotyping. Cold Spring Harb. Sym. 68:69–78. doi:10.1101/sqb.2003.68.69

Gutierrez-Gonzalez, J.J., S.K. Guttikonda, L.S.P. Tran, D.L. Aldrich, R. Zhong, O. Yu, H.T. Nguyen, and D.A. Sleper. 2010a. Differential expression of isoflavone biosynthetic genes in soybean during water deficits. Plant Cell Physiol. 51:936–948. doi:10.1093/pcp/pcq065

Gutierrez-Gonzalez, J.J., T.D. Vuong, R. Zhong, O. Yu, J.-D. Lee, G. Shannon, M. Ellersieck, H.T. Nguyen, and D.A. Sleper. 2011. Major locus and other novel additive and epistatic loci involved in modulation of isoflavone concentration in soybean seeds. Theor. Appl. Genet. 123:1375–1385. doi:10.1007/s00122-011-1673-x

Gutierrez-Gonzalez, J.J., X.L. Wu, J.D. Gillman, J.D. Lee, R. Zhong, O. Yu, G. Shannon, M. Ellersieck, H.T. Nguyen, and D.A. Sleper. 2010b. Intricate environment-modulated genetic networks control isoflavone accumulation in soybean seeds. BMC Plant Biol. 10:105. doi:10.1186/1471-2229-10-105

Gutierrez-Gonzalez, J.J., X.L. Wu, J. Zhang, J.D. Lee, M. Ellersieck, J.G. Shannon, O. Yu, H.T. Nguyen, and D.A. Sleper. 2009. Genetic control of soybean seed isoflavone content: Importance of statistical model and epistasis in complex traits. Theor. Appl. Genet. 119:1069–1083. doi:10.1007/s00122-009-1109-z

Hyten, D.L., I.Y. Choi, Q.J. Song, J.E. Specht, T.E. Carter, R.C. Shoemaker, E.Y. Hwang, L.K. Matukumalli, and P.B. Cregan. 2010. A high-density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping. Crop Sci. 50:960–968. doi:10.2135/cropsci2009.06.0360

Hyten, D.L., V.R. Pantalone, C.E. Sams, A.M. Saxton, D. Landau-Ellis, T.R. Stefaniak, and M.E. Schmidt. 2004. Seed quality QTL in a prominent soybean population. Theor. Appl. Genet. 109:552–561. doi:10.1007/s00122-004-1661-5

Hyten, D., Q. Song, I.Y. Choi, M.S. Yoon, J. Specht, L. Matukumalli, R. Nelson, R. Shoemaker, N. Young, and P. Cregan. 2008. High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor. Appl. Genet. 116:945–952. doi:10.1007/s00122-008-0726-2

Kassem, M.A., K. Meksem, M.J. Iqbal, V.N. Njiti, W.J. Banz, T.A. Winters, A. Wood, and D.A. Lightfoot. 2004. Definition of soybean genomic regions that control seed phytoestrogen amounts. J. Biomed. Biotechnol. 1:52–60. doi:10.1155/S1110724304304018

Kassem, M.A., J. Shultz, K. Meksem, Y. Cho, A.J. Wood, M.J. Iqbal, and D.A. Lightfoot. 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. doi:10.1007/s00122-006-0361-8

Lozovaya, V.V., A.V. Lygin, A.V. Ulanov, R.L. Nelson, J. Dayde, and J.M. Widhohn. 2005. Effect of temperature and soil moisture status during seed development on soybean seed isoflavone concentration and composition. Crop Sci. 45:1934–1940. doi:10.2135/cropsci2004.0567

Meksem, K., V.N. Njiti, W.J. Banz, M.J. Iqbal, M.M. Kassem, D.L. Hyten, J. Yuang, T.A. Winters, and D.A. Lightfoot. 2001. Genomic regions that underlie soybean seed isoflavone content. J. Biomed. Biotechnol. 1:38–44. doi:10.1155/S1110724301000110

Meng, F., Y. Han, W. Teng, Y. Li, and W. Li. 2011. QTL underlying the resistance to soybean aphid (*Aphis glycines* Matsumura) through isoflavone-mediated antibiosis in soybean cultivar 'Zhongdou 27'. Theor. Appl. Genet. 123:1459–1465. doi:10.1007/s00122-011-1680-y

Morrison, M.J., E.K. Cober, M.F. Saleem, N.B. McLaughlin, J. Fregeau-Reid, B.L. Ma, W. Yan, and L. Woodrow. 2008. Changes in isoflavone concentration with 58 years of genetic improvement of short-season soybean cultivars in Canada. Crop Sci. 48:2201–2208. doi:10.2135/cropsci2008.01.0023

Murphy, S.E., E.A. Lee, L. Woodrow, P. Seguin, J. Kumar, I.

Rajcan, and G.R. Ablett. 2009. Genotype X environment interaction and stability for isoflavone content in soybean. Crop Sci. 49:1313–1321. doi:10.2135/cropsci2008.09.0533

National Center for Biotechnology Information. 2013. NCBI homepage. http://www.ncbi.nlm.nih.gov (accessed 1 Sept. 2013). U.S. National Library of Medicine, Bethesda, MD.

Nyquist, W.E. 1991. Estimation of heritability and prediction of selection response in plant populations. Crit. Rev. Plant Sci. 10:235–322. doi:10.1080/07352689109382313

Ogbuewu, I.P., M.C. Uchegbu, O.O. Emenalom, I.C. Okoli, and M.U. Iloeje. 2010. Overview of the chemistry of soy isoflavones, potential threats and potential therapeutic benefits. EJEAFChe. 9:682–695.

Panthee, D.R., V.R. Pantalone, C.E. Sams, A.M. Saxton, D.R. West, J.H. Orf, and A.S. Killam. 2006. Quantitative trait loci controlling sulfur containing amino acids, methionine and cysteine, in soybean seeds. Theor. Appl. Genet. 112:546–553. doi:10.1007/s00122-005-0161-6

Primomo, V.S., V. Poysa, G.R. Ablett, C.J. Jackson, M. Gijzen, and I. Rajcan. 2005. Mapping QTL for individual and total isoflavone content in soybean seeds. Crop Sci. 45:2454–2464. doi:10.2135/cropsci2004.0672

R Core Development Team. 2009. R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria.

Sagara, M., T. Kanda, M.N. Jelekera, T. Teramoto, L. Armitage, N. Birt, C. Birt, and Y. Yamori. 2003. Effects of dietary intake of soy protein and isoflavones on cardiovascular disease risk factors in high risk, middle-aged men in Scotland. J. Am. Coll. Nutr. 23:85–91.

SAS Institute. 2008. The SAS system for Windows. Release 9.2. SAS Inst., Cary, NC.

Schmutz, J., S.B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D.L. Hyten, Q. Song, J.J. Thelen, J. Cheng, et al. 2010. Genome sequence of the palaeopolyploid soybean. Nature 463:178–183. doi:10.1038/nature08670

Setchell, K.D.R. 1998. Phytoestrogens: The biochemistry, physiology, and implications for human health of soy isoflavones. Am. J. Clin. Nutr. 68:1333S–1346S.

Smith, T.J., and H.M. Camper. 1973. Registration of Essex Soybean1 (Reg. No. 97). Crop Sci. 13:495. doi:10.2135/cropsci1973.0011183X001300040033x

SoyBase and the Soybean Breeder's Toolbox. 2013. Homepage. http://www.soybase.org (accessed 1 Sept. 2013).

SoyBase and the Soybean Breeder's Toolbox. 2007. QTL nomenclature. http://www.soybase.org/resources/QTL.php (accessed 23 May 2013).

Spencer, M.M., D. Landau-Ellis, E.J. Meyer, and V.R. Pantalone. 2004. Molecular markers associated with linolenic acid content in soybean. J. Am. Oil Chem. Soc. 81:559–562. doi:10.1007/s11746-006-0941-4

Torres, A.M., C.M. Avila, N. Gutierrez, C. Palomino, M.T. Moreno, and J.I. Cubero. 2010. Marker-assisted selection in faba bean (*Vicia faba* L.). Field Crops Res. 115:243–252. doi:10.1016/j.fcr.2008.12.002

Wang, S., J. Basten, and Z.B. Zeng. 2011. Windows QTL cartographer 2.5.

Wang, H.J., and P.A. Murphy. 1994. Isoflavone content in commercial soybean foods. J. Agric. Food Chem. 42:1666–1673. doi:10.1021/jf00044a016

Watanabe, S., R. Hideshima, Z. Xia, Y. Tsubokura, S. Sato, Y. Nakamoto, N. Yamanaka, R. Takahashi, M. Ishimoto, T. Anai, S. Tabata, and K. Harada. 2009. Map-based cloning of the gene associated with the soybean maturity locus *E3*. Genetics 182:1251–1262. doi:10.1534/genetics.108.098772

Wilson, R.F. 2004. Seed composition. In: H.R. Boerma and J.E. Specht, editors, Soybeans: Improvement, production, and uses. 3rd ed. ASA, CSSA, and SSSA, Madison, WI. p. 621–678.

Zeng, G.L., D.M. Li, Y.P. Han, W.L. Teng, J. Wang, L.Q. Qiu, and W.B. Li. 2009. Identification of QTL underlying isoflavone contents in soybean seeds among multiple environments. Theor. Appl. Genet. 118:1455–1463. doi:10.1007/s00122-009-0994-5