Plant Genetics, Genomics, and Biotechnology 2 (1): 95–119, 2015
doi: 10.5147/pggb.2015.0145

Selective Genotyping for Marker Assisted Selection Strategies for Soybean Yield Improvement

Benjamin D. Fallen[1*], Fred L. Allen[2], Dean A. Kopsell[2], Arnold M. Saxton[2], Leah McHale[3], J. Grover Shannon[4], Stella K. Kantartzi[5], Andrea J. Cardinal[6], Perry B. Cregan[7], David L. Hyten[7,8], and Vincent R. Pantalone[2]

[1] Clemson University, Advanced Plant Technology Center, Clemson Pee Dee REC, 2200 Pocket Rd., Florence, SC 29506, USA; [2] University of Tennessee, Department of Plant Sciences, 2431 Joe Johnson Dr., Knoxville, TN 37996, USA; [3] Ohio State University, Department of Horticulture and Crop Science, 2021 Coffey Rd., Columbus, OH 43210, USA; [4] University of Missouri, Division of Plant Sciences, Delta Center, P.O. Box 160, Portageville, MO 63873, USA; [5] Southern Illinois University, Department of Plant, Soil Science and Agricultural Systems, 1205 Lincoln Drive, Carbondale, IL 62901, USA; [6] Syngenta Seeds, 3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA; [7] Soybean Genomics and Improvement Laboratory, Beltsville Agricultural Research Center – West, USDA, ARS, Beltsville, MD 20705, USA; [8] Current address: DuPont Pioneer, 8305 NW 62nd Ave., PO Box 7060, Johnston, IA 50131-7060, USA.

Received: October 6, 2014 / Accepted: January 6, 2015

Abstract

Using molecular markers in soybean [*Glycine max* (L.) Merr.] has lead to the identification of major loci controlling quantitative and qualitative traits that include: disease resistance, insect resistance and tolerance to abiotic stresses. Yield has been considered as one of the most important quantitative traits in soybean breeding. Unfortunately, yield is a very complex trait and most yield quantitative trait loci (QTL) that have been identified have had only limited success for marker assisted selection (MAS). The objective of this study was to identify QTL associated with soybean seed yield in preliminary yield trials grown in different environments and to evaluate their effective use for MAS using a yield prediction model (YPM), which included epistasis. To achieve this objective, 875 $F_{5:9}$ recombinant inbred lines (RIL) from a population developed from a cross between two prominent ancestors of the North American soybean (Essex and Williams 82) were used. The 875 RIL and check cultivars were divided into four groups based on maturity and each group was grown in Knoxville, TN and one other location that had an environment in which the maturity group (MG) was adapted to be grown. Each RIL was genotyped with >50,000 single nucleotide polymorphic markers (SNPs) of which 17,232 were polymorphic across the population. Yield QTL were detected using a single factor (SF) analysis of variance (ANOVA) and composite interval mapping (CIM). Based on CIM, 23 yield QTL were identified. Twenty-one additional QTL were detected using SF ANOVA. Individually, these QTL explained from 4.5% to 11.9% of the phenotypic variation for yield. QTL were identified on all 20 chromosomes and five of the 46 QTL have not been previously reported. This study provides new information concerning yield QTL in soybean and may offer important insights into MAS strategies for soybean.

Keywords: Genomic selection, epistasis, predictive breeding, QTL analysis.

Introduction

Cultivar improvements in yield have allowed the soybean [*Glycine max* (L.) Merr.] to become the most important source of vegetable protein and oil in the world and the second most important crop in the U.S. In 2012, the estimated seed yield of soybean in the U.S. was 82 million metric tons harvested from 31.2 million hectares of land (Soy Stats, 2013). However, the genetic gain is still only about 1% a year in soybean (Hao et al.,

* Corresponding author: bfallen@clemson.edu

2012; Rincker et al., 2014).

Sebastian (2010) and Hyten et al. (2006) showed that current selection procedures are not efficient in exploiting the available genetic diversity. Using MAS for yield could not only increase breeding efficiency, but also would improve our understanding of the genetic mechanisms of seed yield. Although there has been an increased interest in MAS, very few yield QTL in soybean have been validated across a wide range of environments and populations. Bernado (2008) concluded that because estimated QTL effects for traits such as grain yield are limited to the set of segregating progeny from a single cross, QTL mapping for such traits will likely have to be repeated for each breeding population. Sebastian et al (2010) used context-specific MAS (CSM) to detect yield QTL in elite soybean cultivars. Selected subline haplotypes were compared to their respective maternal lines in highly replicated yield trials across multiple locations and years. From the selected sublines, significant yield gains of up to 5.8% were confirmed and two of the improved sublines were released as improved cultivars.

However, one of the major problems when using MAS is building statistical models that can handle data sets consisting of a massive number of markers that well exceed the number of genotypes being evaluated. Traditionally, a subset of predictors in a regression model are obtained by forward selection, backward elimination or stepwise selection (Li et al., 2011), but these approaches are difficult to use when the number of predictors (SNPs) far exceed the number of observations. Long et al. (2011) conducted a study to evaluate two dimension reduction methods, supervised principal component regression (PCR) and sparse principal least-square regression (PLS), for predicting genomic breeding values (BV) of dairy bulls for milk yield using SNPs. PCR and PLS reduce model dimension and overcome multicollinearity problems by transforming the large number of original variables into a relatively small number of orthogonal latent components and then regress the response variable on those latent components. In their study supervised PCR was used to preselect SNPs based on strength of association of each SNP with the phenotype. Two types of supervised PCR were used: method I was based on single-SNP analyses and method II was based on multiple-SNP analyses. Then the Bayesian Lasso (a statistical method which uses maker-specific shrinkage of the effects) was used to estimate the regression coefficients of the principal components and these regression coefficients were used to rank and select SNPs. They concluded that PCR II was the best method for dimension reduction and variable selection for predicting genomic BVs. Li et al. (2011) also proposed a two stage procedure for multi-SNP modeling and analysis in genome wide association studies (GWASs), by first producing a 'preconditioned' response variable using a supervised principle component analysis and then formulating Bayesian Lasso to select a subset of significant SNPs. Using simulation data they demonstrated that when the number of markers greatly exceeds the number of observations 'preconditioned' or specialized PCA can successfully identify almost all SNPs with true genetic effects. Other studies have also used PCR and PLS for genome-assisted prediction of breeding values (Solberg et al., 2009; Macciotta et al., 2010). However, these methods are very challenging to use and require extensive computing technology and time.

The objectives of this study were to test whether: 1) MAS for haplotypes accumulating in the top 10% of loci positive for yield differ significantly than the population mean when grown in different environments and thus are considered favorable for selecting high yielding lines; 2) MAS for haplotypes can distinguish low yielding vs. high yielding lines; and 3) phenotypic selections for yield differ from genotypic SNP selections for yield.

Materials and Methods

Population Development

Essex originated from the cross 'Lee' × S5-7075 at the Virginia Agricultural Experiment Station and was released in 1972 (Smith and Camper, 1973). Essex is characterized as having purple flowers, gray pubescence, a group V maturity, average protein, oil, height and yield and is susceptible to sudden death syndrome (SDS) caused from Fusarium solani f. sp. glycines. Williams 82 was developed by the USDA-ARS and the Illinois Agricultural Experiment Station by combining four individual BC¬6F3 plants selected after a series of backcrosses to 'Williams' to transfer the Rps1k ¬ allele from Kingwa (Bernard and Lindahl, 1972). The Rps1k¬ allele confers resistances to certain races of Phytophthora sojae which causes phytophthora root rot. Williams 82 is characterized as having white flowers, tawny pubescence, a group III maturity, average seed protein and oil content, resistance to phytophthora root rot and mild resistance to SDS. Williams 82 has contributed to the genetic background of many northern U.S. cultivars and Essex has contributed to the genetic background of many southern U. S. cultivars and elite breeding lines (Sneller, 1994; Gizlice et al., 1996). A population formed from these diverse parents should reflect a broad measure of the range of seed yield loci contributing to incremental gains in elite U.S. soybean cultivars. Therefore, QTL detected in this population are likely to be segregating in a wide range of North American breeding programs.

The initial crosses for the 'Essex' × 'Williams 82' population were made at the East Tennessee Research and Extension Center (ETREC) in Knoxville, TN in the summer of 2005. In the fall of 2005, the F1 seeds obtained from the Essex ×Williams 82 cross were harvested and grown in Isabela, PR at the USDA-ARS Tropical Agricultural Research Station (TARS). The population was advanced from the F2 to the F5 generation through single seed descent (Brim, 1966). The F2 generation was grown at ETREC in 2006 and the F3 generation was grown at ETREC in 2007. The F4 and F5 generations were grown at the TARS location in the winter of 2007/2008 and the spring of 2008, respectively. In the summer of 2009 in Beltsville, MD F5 plants were grown in a greenhouse and leaf tissue was collected from each plant individually. A total of 977 individually tagged F¬5 plants were harvested from the greenhouse and planted as F¬5:6 plant rows in Homestead, FL in the fall of 2009. The F5:6 rows were harvested individually and in 2010 the F5:7 recombinant inbred lines were planted at ETREC in Knoxville, TN.

Experimental Design

The lines were divided into four groups based on the maturity date recorded on a single plant in Beltsville, MD in 2009. In 2010, 973 recombinant inbred lines were planted in Knoxville, TN. Each line was planted in one rep as a two row plot 6 m in length, with 76 cm spacing between rows. In 2011, the four groups containing a total of 875 recombinant inbred lines and 12 checks for overall agronomic comparisons were planted in Knoxville, TN. The four groups were designated as: Group A, Group B, Group C and Group D. In Group A there were 218 RIL and three checks: 'IA3024', 'IA3023', and 'LD00-3309'. The maturity ranged from an early maturity group (MG) III to a late MG III. In Group B there were 221 RIL and three checks: 'IA4005', LD00-3309 and LD00-2817P. The maturity ranged from a late MG III to an early MG IV. In Group C there were 216 RIL and three checks: LD00-2817P, TN09-008 and '5002T'. The maturity ranged from an early MG IV to a late MG IV. Check LD00-2817P was not included in the final mean seed yield comparison in Groups B and C because of poor germination and plant stand. In Group D there were 220 RIL and three checks: 5002T, '5601T' and 'Osage'. The maturity ranged from an early MG V to a late MG V. A randomized complete block design was used and each line was planted in two reps of a two row plot 3.5 m in length, with 76 cm spacing between rows. In addition, Group A was planted in Wooster, OH in two reps of a two row plot 4.9 m in length, with 76 cm spacing between rows. Group B was planted in Belleville, IL in two reps of a two row plot 4.5 m in length, with 76 cm spacing between rows. Group C was planted in Portageville, MO in two reps of a two row plot 3.5 m in length, with 76 cm spacing between rows. Group D was planted in Plymouth, NC in two reps of a two row plot 5 m in length, with 76 cm spacing between rows. This allowed all groups to be planted in the same location (Knoxville, TN) and for each group to be planted in another environment where its maturity was expected to be well adapted.

Experimental Procedures

Phenotypic Data

After planting, all the plots were evaluated for agronomic traits. At maturity, plant height was measured as an estimation of the distance from the soil surface to the tip of the main stem. Lodging was scored on a scale from 1-5; with 1 being all the plants in a plot were erect and 5 being all the plants in a plot were prostrate. Maturity was recorded as the date, according to the Julian calendar, when 95 % of the pods achieved their mature color. In Knoxville, TN seed yield was estimated from two rows after the plots had been end trimmed to 4.88 m in length. In Wooster, OH, Belleville, IL and Portageville, MO seed yield was estimated from harvesting two rows at 4.9 m, 4.5 m and 3.5 m length rows, respectively. In Plymouth, NC seed yield was estimated from harvesting two rows after the plots had been trimmed to 3.5 m in length. All yields were adjusted to 13% moisture.

Genotypic Data

DNA was extracted from each F5 greenhouse plant grown at the Soybean Genomics and Improvement Laboratory at the USDA Beltsville Agricultural Research Center (USDA-ARS) in Beltsville, MD. Each DNA sample was processed to contain 50 µl of DNA at a 200 ng/µl concentration. The samples were then assayed using >50,000 SNP markers using the Infinium® assay and analyzed on the Illumina BeadStation 500G (Illumina, San Diego, CA) (Song et al., 2013). A total of 17,232 polymorphic SNP markers were found in the population.

Statistical Analysis

Marker order, position and composite interval mapping (CIM) were conducted using CIM (Broman and Sen, 2009). A total of 1,000 permutations were performed for all chromosomes to establish an empirical LOD threshold at the 5% probability level. Of the 17, 232 polymorphic SNP markers 15, 448 were assigned to 20 chromosomes; the remaining 1,784 markers were unlinked. The estimated map length was 2072 cM with an average distance between markers of 0.2 cM.

A single factor (SF) analysis of variance (ANOVA) was also used for QTL analysis ($P<0.01$) using SAS (PROC MIXED, SAS ver. 9.1.s, Cary, NC). Each marker was considered a factor with two levels: "A" designating the Essex allele type and "B" designating the Williams82 allele type and the phenotype (yield) as the dependent variable. Heterozygotes were not included for QTL analysis using CIM or SF ANOVA.

An additive effect for each QTL was determined using the method in which the QTL was detected (CIM or SF ANOVA). Additive effects were determined separately for each environment and across environments within each group. Prediction models for yield in each group were made based on 2010 QTL data; from QTL data for each 2011 environment; and using QTL data combined over 2010 and 2011 environments. Yield was predicted using the following: (a) the overall mean yield of each genotype, (b) the additive effect of the QTL identified using SF ANOVA in SAS or CIM in R/qtl and (c) the additive and additive by additive epistatic QTL effects limited to those found to be highly significant ($P<0.01$) in epistasis with the detected additive QTL. Additive by additive epistatic effects were determined separately for each group for each environment and across environments at $P<0.01$ using the Epistacy macro, version 2.0 in SAS (Holland, 1998).

To determine yield in Knoxville, TN in 2010 the plot weight from one rep of each line was used to calculate yield in kg ha-1. Analysis of variance was conducted in SAS using PROC MIXED (SAS ver. 9.1.3, Cary, NC) to test for significant genotype differences among RIL for yield in each location grown in 2011 and combined across locations and years. Location, replication and year were considered as random blocking factors in the model and genotypes were considered fixed effects. Since each group had approximately 220 RILs the top 22 yielding lines were considered the top yielding 10% and the top 11 yielding lines were considered the top yielding 5% in each group. Likewise, the bottom 22 yielding lines were considered the bottom yielding

10% and the bottom 11 yielding lines were considered the bottom yielding 5% in each group.

Results

Group A: Agronomic Traits

In Group A, Wooster, OH had an average yield (3339 kg ha-1) that was significantly ($p<0.01$) higher than the average yield in Knoxville, TN in 2010 (1740 kg ha^{-1}) and 2011 (1486 kg ha^{-1}). The higher yields in Wooster, OH in 2011 may be due to the highly adapted maturity of Group A for that environment (Sleper, 2006). The maturity ranged from an early MG III to a late MG III in Group A, which is more adapted to the latitude of Wooster, OH than Knoxville, TN (Sleper, 2006). Average lodging and height were not significantly different across locations. Average maturity was significantly different across locations. The average maturity date was 260 for Knoxville, TN in 2010, 250 for Knoxville, TN in 2011 and 270 for Wooster, OH in 2011.

Group A: MAS Using Only Additive Effects

Using QTL Discovered in Knoxville, TN in 2010 to Predict High Yielding Lines Across Multiple Environments in 2011

In 2010 in Knoxville, TN three QTL were identified for yield using CIM (Table 1). Using MAS to select lines with the favorable allele for these QTL five lines in the top yielding 10% of RIL combined over three environments (Knoxville, TN in 2010, 2011 and

Table 1. Quantitative trait loci identified using CIM or SF ANOVA located on various molecular linkage groups associated with yield in 875 RIL derived from a cross between Essex 86-15-1 x Williams 82-11-43-1. The lines were divided into four groups based on maturity and were grown in two environments. Data is presented from the two individual environments and combined across the two locations.

ENVIRONMENT	MARKERS	CHR	MLG	LOC (cM)	LOD	R^2 (%)	ADD. EFFECT†	FAV. ALLELE	P-VALUE	PROGRAM	GROUP
Knoxville, TN 2010	Gm01_1241762_A_C	1	D1a	4.60	.	8.50	2.24	W	0.0003	SAS	B
Wooster, OH 2011	Gm01_1494600_C_T	1	D1a	5.52	.	4.73	2.44	E	0.009	SAS	A
Knoxville, TN 2010	Gm01_1045893_G_A	1	D1a	5.88	2.63	5.45	1.18	E	.	R/qtl	C
Knoxville, TN 2010	Gm01_2747136_A_C	1	D1a	11.28	.	7.32	1.30	W	0.0008	SAS	C
Belleville, IL 2011	Gm01_29787876_G_A	1	D1a	59.29	.	10.02	0.92	E	<.0001	SAS	B
Knoxville, TN 2010-11 &											
Belleville, IL 2011	Gm01_29787876_G_A	1	D1a	59.29	.	8.08	1.00	E	<.0001	SAS	B
Knoxville, TN 2010-11 &											
Plymouth, NC 2011	Gm01_47115450_G_T	1	D1a	70.15	.	5.61	0.24	E	0.0008	SAS	D
Knoxville, TN 2010	Gm01_54171147_G_T	1	D1a	118.27	.	4.91	1.81	E	0.0082	SAS	D
Knoxville, TN 2010	Gm02_707483_A_G	2	D1b	5.25	3.07	6.7	2.48	E	.	R/qtl	A
Knoxville, TN 2010-11 &											
Portageville, MO 2011	Gm02_6820177_A_C	2	D1b	38.07	3.25	4.31	1.80	W	.	R/qtl	C
Knoxville, TN 2010	Gm02_6821311_A_C	2	D1b	38.24	2.35	4.35	1.18	E	.	R/qtl	C
Knoxville, TN 2010	Gm02_12770553_A_G	2	D1b	46.15	.	6.29	1.69	W	0.0022	SAS	B
Knoxville, TN 2010-11 &											
Belleville, IL 2011	Gm02_42469280_A_C	2	D1b	105.17	2.65	4.07	1.16	W	.	R/qtl	B
Knoxville, TN 2010	Gm02_44803277_C_T	2	D1b	107.06	.	6.11	0.51	W	0.0026	SAS	C
Belleville, IL 2011	Gm02_44803277_C_T	2	D1b	114.09	2.83	4.66	2.10	W	.	R/qtl	B
Knoxville, TN 2010	Gm02_47790307_C_T	2	D1b	121.66	.	6.04	3.39	E	0.0028	SAS	A
Wooster, OH 2011	Gm02_49126947_T_C	2	D1b	127.25	.	5.31	3.44	E	0.0051	SAS	A
Knoxville, TN 2010-11 &											
Wooster, OH 2011	Gm02_49126947_T_C	2	D1b	127.25	.	5.07	5.82	E	0.0071	SAS	A
Knoxville, TN 2010-11 &											
Portageville, MO 2011	Gm02_49746270_A_G	2	D1b	146.54	.	5.40	1.19	W	0.0046	SAS	C
Knoxville, TN 2010-11 &											
Wooster, OH 2011	Gm02_47790307_C_T	2	D1b	150.38	2.56	5.7	3.26	E	.	R/qtl	A
Portageville, MO 2011	Gm03_838582_T_C	3	N	4.68	.	4.82	2.34	W	0.0089	SAS	C
Wooster, OH 2011	Gm03_2151432_A_G	3	N	14.00	3.21	8.3	4.33	E	.	R/qtl	A
Belleville, IL 2011	Gm03_5264953_A_G	3	N	19.43	.	5.58	0.36	E	0.001	SAS	B
Knoxville, TN 2010-11 &											
Portageville, MO 2011	Gm03_21003884_A_G	3	N	44.15	.	6.76	0.37	E	0.0012	SAS	C
Plymouth, NC 2011	Gm03_39552601_T_C	3	N	87.68	.	5.54	3.81	E	0.0045	SAS	D
Plymouth, NC 2011	Gm03_39559139_G_A	3	N	93.64	2.78	7.38	3.09	E	.	R/qtl	D
Knoxville, TN 2010-11 &											
Wooster, OH 2011	Gm03_47386481_A_C	3	N	120.71	.	5.67	5.81	E	0.004	SAS	A
Knoxville, TN 2010-11 &											
Plymouth, NC 2011	Gm04_8845668_G_T	4	C1	63.93	.	4.84	0.28	E	0.0081	SAS	D
Knoxville, TN 2010	Gm04_8247949_C_T	4	C1	65.87	.	6.79	0.97	W	0.0014	SAS	D
Knoxville, TN 2010	Gm04_48782140_G_T	4	C1	152.98	2.48	6.4	2.13	E	.	R/qtl	A
Wooster, OH 2011	Gm04_48993297_T_G	4	C1	154.16	2.78	5.2	3.18	E	.	R/qtl	A
Knoxville, TN 2010	Gm05_1128604_A_G	5	A1	3.24	.	4.95	0.52	W	0.0024	SAS	C
Belleville, IL 2011	Gm05_3485480_T_C	5	A1	19.73	2.66	5.86	1.61	W	.	R/qtl	B
Knoxville, TN 2010-11 &											
Wooster, OH 2011	Gm05_33176582_G_A	5	A1	33.77	3.44	7.8	2.56	W	.	R/qtl	A
Knoxville, TN 2010-11 &											
Belleville, IL 2011	Gm05_30953466_G_T	5	A1	39.76	.	7.68	1.60	W	0.0005	SAS	B
Knoxville, TN 2010-11 &											
Plymouth, NC 2011	Gm05_31399360_G_A	5	A1	41.55	.	5.71	0.99	W	0.0007	SAS	D
Knoxville, TN 2010-11 &											
Portageville, MO 2011	Gm05_34850619_C_T	5	A1	72.38	.	5.71	0.27	W	0.0007	SAS	C

Table 1. Continued.

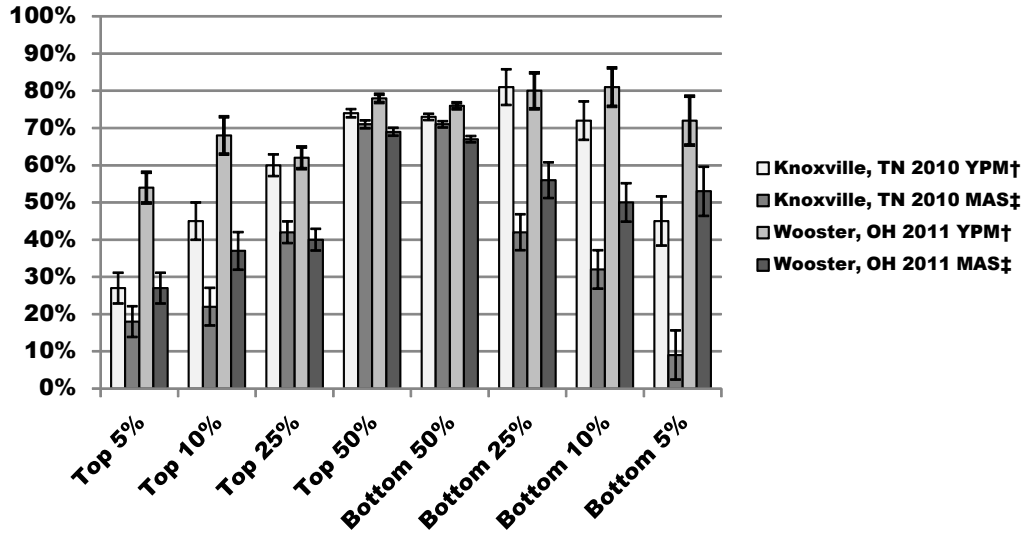
ENVIRONMENT	MARKERS	CHR	MLG	LOC (cM)	LOD	R^2 (%)	ADD. EFFECT†	FAV. ALLELE	P-VALUE	PROGRAM	GROUP
Portageville, MO 2011 Knoxville, TN 2010-11&	Gm06_10864751_A_G	6	C2	24.86	.	5.61	2.83	W	0.0042	SAS	C
Portageville, MO 2011	Gm06_16723946_G_A	6	C2	32.46	3.72	5.57	2.64	W	.	R/qtl	C
Knoxville, TN 2010	Gm06_17617727_G_T	6	C2	55.04	2.82	3.42	3.70	W	.	R/qtl	B
Knoxville, TN 2010 Knoxville, TN 2010-11&	Gm06_20996124_T_C	6	C2	58.54	.	9.03	7.90	W	0.0002	SAS	B
Belleville, IL 2011	Gm06_20996124_T_C	6	C2	58.54	.	10.63	4.03	W	<.0001	SAS	B
Belleville, IL 2011 Knoxville, TN 2010-11&	Gm06_20996124_T_C	6	C2	60.21	5.56	10.48	5.26	W	.	R/qtl	B
Belleville, IL 2011	Gm06_20996124_T_C	6	C2	62.03	3.92	6.23	3.22	W	.	R/qtl	B
Belleville, IL 2011	Gm06_27540819_T_G	6	C2	66.24	.	10.29	4.48	W	<.0001	SAS	B
Plymouth, NC 2011 Knoxville, TN 2010-11&	Gm07_149664_T_C	7	M	1.34	.	11.29	5.43	W	<.0001	SAS	D
Plymouth, NC 2011 Knoxville, TN 2010-11&	Gm07_4008483_C_T	7	M	5.19	2.92	8.64	1.86	W	.	R/qtl	D
Portageville, MO 2011	Gm07_4837493_A_G	7	M	11.06	.	5.71	2.04	E	0.0007	SAS	C
Knoxville, TN 2010 Knoxville, TN 2010-11&	Gm07_16814628_C_T	7	M	38.47	.	5.41	0.83	W	0.0051	SAS	C
Belleville, IL 2011	Gm07_17460956_C_A	7	M	39.95	.	14.85	1.90	W	<.0001	SAS	B
Knoxville, TN 2010	Gm07_18539902_T_G	7	M	42.42	.	5.69	3.04	W	0.0039	SAS	D
Knoxville, TN 2010 Knoxville, TN 2010-11&	Gm07_16144523_C_A	7	M	51.90	3.65	6.67	1.87	W	.	R/qtl	B
Belleville, IL 2011	Gm07_17362808_A_G	7	M	55.95	5.31	8.20	2.04	W	.	R/qtl	B
Knoxville, TN 2010 Knoxville, TN 2010-11&	Gm07_18539902_T_G	7	M	61.37	3.52	8.83	2.67	W	.	R/qtl	D
Belleville, IL 2011	Gm08_15866777_G_A	8	A2	22.31	.	7.09	0.35	E	0.0001	SAS	B
Plymouth, NC 2011	Gm09_457853_A_G	9	K	5.23	.	6.06	4.10	E	0.0027	SAS	D
Plymouth, NC 2011 Knoxville, TN 2010-11&	Gm09_2634593_G_A	9	K	5.62	3.02	7.87	3.09	E	.	R/qtl	D
Plymouth, NC 2011	Gm09_3394608_G_A	9	K	7.76	.	4.53	1.20	E	0.0037	SAS	D
Knoxville, TN 2010	Gm09_6967374_C_T	9	K	15.94	.	4.64	0.88	E	0.0106	SAS	A
Portageville, MO 2011	Gm09_18969901_T_C	9	K	28.52	2.32	3.81	2.77	W	.	R/qtl	C
Belleville, IL 2011 Knoxville, TN 2010-11&	Gm09_12463468_C_T	9	K	31.76	.	9.79	0.02	W	<.0001	SAS	B
Belleville, IL 2011	Gm09_12463468_C_T	9	K	31.76	.	7.11	0.45	W	<.0001	SAS	B
Portageville, MO 2011 Knoxville, TN 2010-11&	Gm09_34191288_T_C	9	K	78.24	.	6.88	3.47	W	0.0013	SAS	C
Belleville, IL 2011	Gm10_571698_A_G	10	O	1.30	.	6.48	0.14	E	0.0016	SAS	B
Wooster, OH 2011 Knoxville, TN 2010-11&	Gm10_47585270_T_G	10	O	108.89	.	5.35	2.27	E	0.0049	SAS	A
Plymouth, NC 2011	Gm10_48428720_T_C	10	O	110.82	.	5.46	0.11	E	0.001	SAS	D
Knoxville, TN 2010 Knoxville, TN 2010-11&	Gm11_4453218_T_C	11	B1	16.23	.	5.66	2.88	E	0.004	SAS	D
Wooster, OH 2011 Knoxville, TN 2010-11&	Gm11_5773052_G_A	11	B1	20.42	.	6.53	3.80	E	0.0018	SAS	A
Belleville, IL 2011 Knoxville, TN 2010-11&	Gm11_7323949_A_G	11	B1	26.24	.	6.83	0.28	E	0.0001	SAS	B
Portageville, MO 2011 Knoxville, TN 2010-11&	Gm11_7445495_G_A	11	B1	26.72	.	5.97	0.67	E	0.0026	SAS	C
Plymouth, NC 2011	Gm11_36807939_C_A	11	B1	84.22	.	5.95	1.25	E	0.0027	SAS	D
Knoxville, TN 2010	Gm12_1594873_A_G	12	H	3.64	.	5.34	0.62	W	0.0055	SAS	C
Belleville, IL 2011 Knoxville, TN 2010-11&	Gm12_7135310_A_G	12	H	36.25	3.71	6.22	2.28	W	.	R/qtl	B
Portageville, MO 2011 Knoxville, TN 2010-11&	Gm12_39962521_A_G	12	H	91.44	.	6.07	1.54	E	0.0004	SAS	C
Plymouth, NC 2011 Knoxville, TN 2010-11&	Gm13_11355266_T_C	13	F	35.49	.	6.73	1.34	E	0.0002	SAS	D
Wooster, OH 2011	Gm13_27348409_A_G	13	F	150.28	.	6.07	4.13	E	0.0006	SAS	A
Knoxville, TN 2010	Gm13_27092408_C_T	13	F	150.77	2.75	6.18	2.21	E	.	R/qtl	D
Plymouth, NC 2011	Gm13_29895148_C_T	13	F	154.76	.	4.73	2.54	W	0.0098	SAS	D

†ADD. EFFECT = Additive effect refers to the quantitative change in yield that is associated with either (E) Essex 15-86-1 or (W) Williams 82-11-43-1; Group A, B, C, D represent maturity subpopulations tested; LOD = likelihood of odds ratio; CHR= chromosome; MLG = molecular linkage group.

Wooster, OH in 2011) were selected (Figure 1). Two of these lines were in the top yielding 5% of RIL combined over the three environments and ranked 1st and 5th in yield. Further credibility of these yield QTL was demonstrated when seven lines in the bottom yielding 10% of RIL combined over the three environments were selected by MAS with the unfavorable alleles for the three QTL identified in Knoxville, TN in 2010 (Figure 1). Two of these lines were in the bottom yielding 5% of RIL combined over the three environments and were the 3rd and 5th lowest yielding lines.

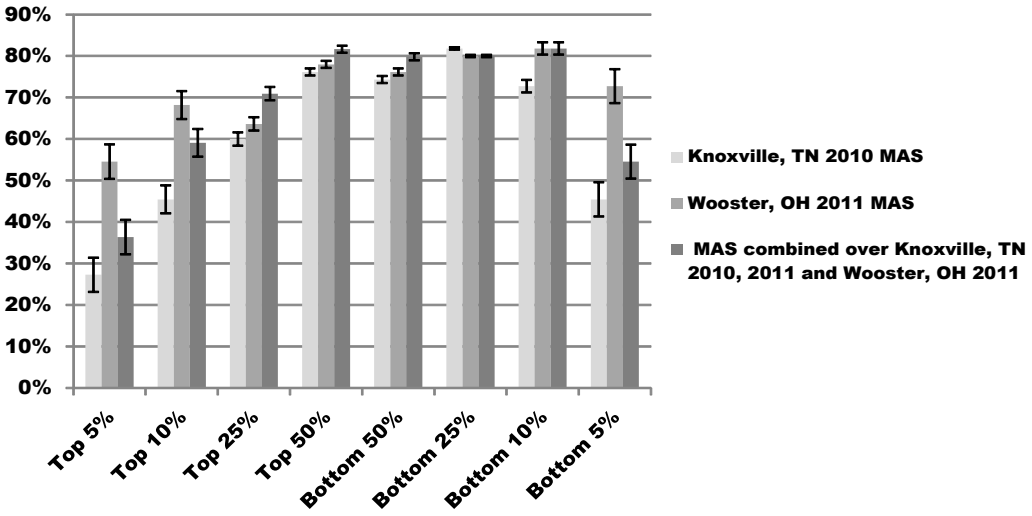
Using QTL Discovered in Wooster, OH in 2011 to Predict High Yielding Lines Across Multiple Environments in 2011

In 2011 in Wooster, OH three QTL were identified for yield using CIM (Table 1). Using MAS to select lines with the favorable allele for these QTL seven lines in the top yielding 10% of RIL combined over three environments (Knoxville, TN in 2010, 2011 and Wooster, OH in 2011) were selected (Figure 1). Three of these lines were in the top yielding 5% of RIL combined over the three environments and ranked 1st, 4th and 5th in yield. Eleven



†Indicates MAS made using the YPM, which included: mean yield, additive effects and additive by additive effects for the QTL detected in that environment; ‡Indicates MAS made using only additive effects for the QTL detected in that environment.

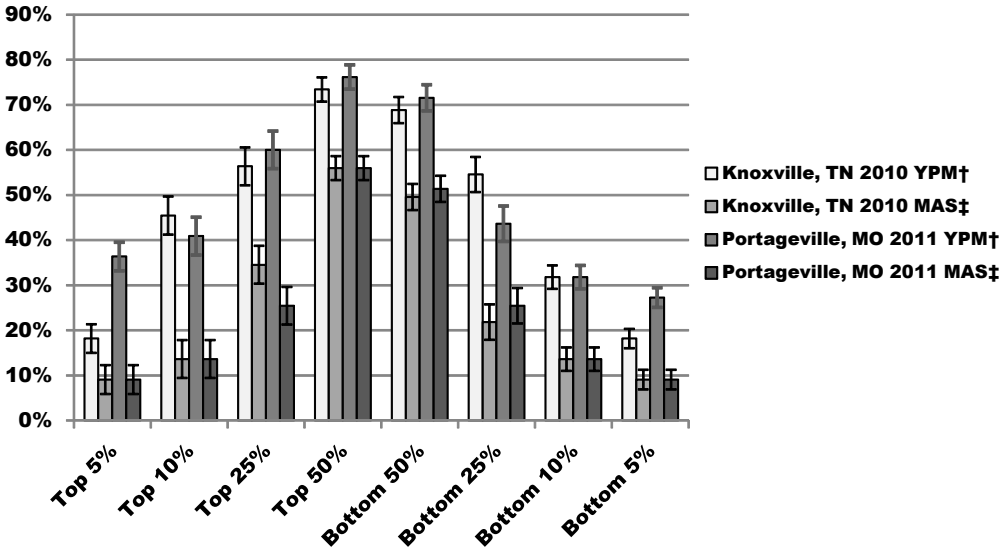
Figure 1. The percentage of marker assisted selections (MAS) made in each environment in Group A compared to phenotypic selections (PS) averaged over all environments (Knoxville, TN 2010, 2011 and Wooster, OH 2011) in Group A. Comparisons were made between the top and bottom % of MAS that were in the corresponding top and bottom % of PS. MAS were made using only additive effects and a yield prediction model (YPM) developed using QTL detected in each environment. PS were based on yield in kg ha^{-1}.



†YPM indicates what environment(s) the data for the model was collected: mean yield, additive effects and additive by additive effects.

Figure 2. The percentage of marker assisted selections (MAS) made in each environment(s) in Group A compared to phenotypic selections (PS) averaged over all environments in Group A. Comparisons were made between the top and bottom % of MAS that were in the corresponding top and bottom % of PS. MAS were made using a yield prediction model (YPM†) developed using QTL detected in each environment(s). PS were based on yield in kg ha^{-1}.

lines in the bottom yielding 10% of RIL combined over the three environments were selected by MAS with the unfavorable alleles for the three QTL identified in Wooster, OH in 2011 (Figure 1). In total six QTL were identified using CIM on five chromosomes (2, 3, 4, 5 and 19) and eleven QTL using SF ANOVA on eleven chromosomes (2, 3, 9, 10, 11, 13, 14, 15 and 19) (Table 1). In certain instances the same marker was associated with the same QTL using different programs or in different environments. A yield QTL was identified with marker Gm02_47790307_C_T from data averaged over Knoxville, TN in 2010, 2011 and Wooster, OH in 2011 using CIM (150.4 cM) and in Knoxville, TN in 2010 using SF ANOVA (121.7 cM) (Table 1). Gm19_44937486_T_C was associated with a yield QTL in Knoxville, TN in 2010 using SF ANOVA at 76.7 cM and CIM at 70.6 cM (Table 1).



†Indicates MAS made using the YPM, which included: mean yield, additive effects and additive by additive effects for the QTL detected in that environment; ‡Indicates MAS made using only additive effects for the QTL detected in that environment.

Figure 3. The percentage of marker assisted selections (MAS) made in each environment in Group C compared to phenotypic selections (PS) averaged over all environments (Knoxville, TN 2010, 2011 and Portageville, MO2011) in Group C. Comparisons were made between the top and bottom % of MAS that were in the corresponding top and bottom % of PS. MAS were made using only additive effects and a yield prediction model (YPM) developed using QTL detected in each environment. PS were based on yield in kg ha^{-1}.

Gm02_49126947_T_C (127.2 cM) was associated with a yield QTL in Wooster, OH in 2011 and from data averaged over Knoxville, TN in 2010, 2011 and Wooster, OH in 2011 using SF ANOVA (Table 1).

Although fewer QTL were identified using CIM than using SF ANOVA more top yielding and bottom yielding lines were selected in individual environments and averaged over all environments by using CIM MAS. And when using CIM more lines were selected among the top 5 yielding lines in individual environments and averaged over all environments. These results suggest that MAS is better when additive QTL were detected using CIM in an early to late MG III soybean.

Group A: YPM Including Mean Yield, Additive and Additive by Additive Effects

To further improve upon the results we found using only additive effects, we then developed an yield prediction model (YPM) which included mean yield, additive and additive by additive QTL effects. In 2010 in Knoxville, TN five QTL were shown to have a significant interaction with two of the QTL identified for yield using CIM (Table 2). This information was used to develop an YPM to select by MAS high yielding lines in subsequent years. Eleven lines that were in the top yielding 10% of RIL grown in Knoxville, TN in 2011 were selected by MAS using the YPM and of those selected lines, three lines were in the top yielding 5% of RIL grown in Knoxville, TN in 2011, including the highest yielding line (Table 3). This information was also used to develop an YPM to select by MAS high yielding lines across environments. Nine lines that were in the top yielding 10% of RIL from the combined analysis of three environments (Knoxville, TN in 2010 and 2011 and Wooster, OH in 2011) were selected by MAS using the YPM and four of those lines were in the top yielding 5% of RIL from the combined analysis of the three environments, including the top two yielding lines 481 and 833 (Table 3).

In 2011 in Wooster, OH seven QTL were shown to have a significant interaction with two of the QTL identified for yield using CIM (Table 2). This information was used to develop an YPM to select by MAS high yielding lines across environments. Fifteen lines in the top yielding 10% of RIL combined over three environments (Knoxville, TN in 2010 and 2011 and Wooster, OH in 2011) were selected by MAS using the YPM, including the top seven yielding lines (Figure 1).

From data averaged across Knoxville, TN in 2010 and 2011 and Wooster, OH in 2011 eleven QTL were shown to have a significant interaction with three of the QTL identified for yield using CIM (Table 2). This information was used to develop an YPM to select by MAS high yielding lines across environments. Thirteen lines in the top yielding 10% of RIL grown over the combined environments of Knoxville, TN in 2010 and 2011 and Wooster, OH in 2011 were selected by MAS using the YPM, including the top three yielding lines (Figure 2).

Using the YPM more lines were selected than using only additive QTL MAS in Group A. Moreover, more of the top yielding lines were selected using QTL identified by CIM than by SF ANOVA using the YPM. This trend was observed through Groups B, C and D. Therefore, detailed results from using SF ANOVA are not discussed in this paper. Additional information on Groups B and D can be found in the Supplementary Data section (Figures 1–4; Tables 1–4). In addition, Wooster, OH had

Table 2. Significant (P<0.01) epistatic interactions between loci for yield in 218 RIL in Group A derived from a cross between Essex 86-15-1 x Williams 82-11-43-1. Locus 1 indicates the markers where yield QTL were detected using CIM and locus 2 indicates the markers where QTL(s) were detected using Epistacy in SAS that were interacting with the yield QTL at locus 1.

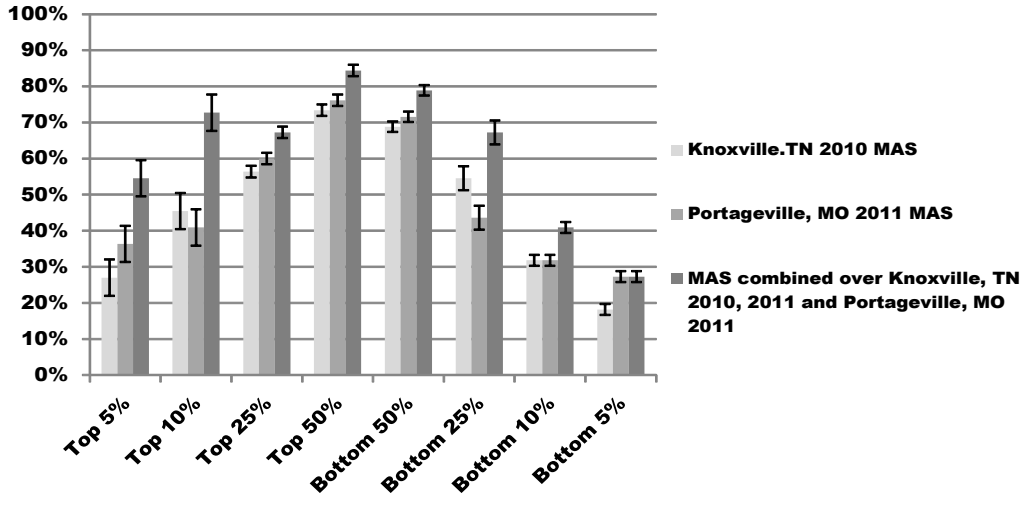
											ADDITIVE X ADDITIVE EFFECT†	
ENVIRONMENT	LOCUS 1	CHR	MLG	LOC (cM)	FAVORABLE ALLELE	LOCUS 2	CHR	MLG	LOC (cM)	R^2 (%)	E	W
Knoxville, TN 2010	Gm19_44937486_T_C	19	L	70.65	W	GM15_10059948_T_C	15	E	15.82	3.12	5.80	3.01
						GM15_50338705_T_C	15	E	79.15	2.77	5.83	3.31
						GM20_41180602_G_A	20	I	64.75	3.01	5.72	3.10
Knoxville, TN 2010	Gm04_48782140_G_T	4	C1	152.98	E	GM06_45433980_G_A	6	C2	71.44	4.22	-0.46	3.09
						GM11_37065128_T_C	11	B1	58.28	4.20	-1.43	1.59
Wooster, OH 2011	Gm19_45198812_C_A	19	L	72.00	W	GM04_11182315_A_G	4	C1	17.58	3.54	0.19	5.91
						GM05_32908802_T_C	5	A1	51.74	5.14	-1.30	5.46
						GM13_28429921_T_C	13	F	44.70	3.68	-0.14	5.81
						GM20_12318232_A_G	20	I	19.37	3.52	5.18	-0.49
Wooster, OH 2011	Gm04_48993297_T_G	4	C1	154.16	E	GM06_49103970_C_T	6	C2	77.21	4.65	-0.65	5.77
						GM10_37618173_A_G	10	O	59.15	5.92	-2.44	4.68
						GM19_44478931_A_G	19	L	69.94	2.67	0.90	6.10
Knoxville, TN 2010-11&												
Wooster, OH 2011	Gm19_44937486_T_C	19	L	70.75	W	GM05_39611177_C_T	5	A1	62.28	1.94	4.83	7.09
						GM11_38762112_G_T	11	B1	60.95	1.78	4.65	6.70
						GM15_49657706_C_T	15	E	78.08	3.70	7.32	4.30
						GM19_42189531_T_C	19	L	66.34	1.66	9.48	5.19
Knoxville, TN 2010-11&												
Wooster, OH 2011	Gm05_33176582_G_A	5	A1	33.77	W	GM02_32518097_T_C	2	D1b	51.13	3.69	0.95	-1.62
						GM16_28901653_G_A	16	J	45.44	3.66	1.27	-1.24
						GM20_34223656_G_A	20	I	53.81	3.89	1.40	-1.32
Knoxville, TN 2010-11&												
Wooster, OH 2011	Gm02_47790307_C_T	2	D1b	150.38	E	GM02_46778366_G_A	2	D1b	73.55	4.42	-1.89	2.85
						GM04_29535808_A_G	4	C1	46.44	3.64	0.04	2.73
						GM18_48533018_G_A	18	D2	76.31	4.13	-0.03	2.88
						GM19_50486916_C_T	19	L	79.38	4.14	0.29	3.13

†Additive by additive effect refers to the quantitative change in yield that is associated with the epistatic combination of the additive genetic effect of locus 1 having the favorable allele with the additive genetic effect of the homozygous state of locus 2 from (E) Essex 15-86-1 or (W) Williams 82-11-43-1.

Table 3. Yield prediction model (YPM) developed using QTL detected in Knoxville, TN in 2010 by CIM to select by MAS the top yielding 10% of RIL in Group A grown in individual environments and averaged across multiple environments. These MAS lines are indicated in bold.

LINE	RANK	LINE	YIELD	LINE	YIELD	LINE	YIELD
†‡§**833**	01	†**668**	2415.5	814	5227.4	§**481**	3319.2
‡§**481**	02	978	2390.3	292	5166.9	§**833**	3110.9
†**155**	03	632	2380.2	689	5160.2	978	3003.4
†‡§**675**	04	754	2345.1	559	4998.9	689	2976.5
‡§**774**	05	†**155**	2341.6	978	4992.2	§**144**	2969.8
†§**668**	06	578	2301.1	896	4918.3	463	2956.4
104	07	†**130**	2197.1	‡**481**	4904.9	§**675**	2875.7
62	08	143	2197.1	463	4857.8	578	2869.0
†**90**	09	689	2163.5	‡**144**	4763.8	814	2828.7
‡§**951**	10	203	2141.7	‡**833**	4710.0	756	2815.3
854	11	559	2138.3	146	4669.7	502	2808.5
995	12	480	2133.3	751	4642.8	292	2801.8
†**734**	13	†**833**	2131.6	211	4636.1	896	2801.8
†**919**	14	†**865**	2126.6	754	4575.6	632	2795.1
799	15	†**675**	2106.4	148	4562.2	§**774**	2795.1
1004	16	743	2093.0	489	4562.2	637	2754.8
524	17	†**919**	2091.3	‡**951**	4562.2	§**951**	2748.1
†§**130**	18	†**144**	2077.9	767	4521.9	§**668**	2748.1
†**865**	19	†**734**	2074.5	‡**675**	4521.9	§**130**	2727.9
†‡§**144**	20	266	2039.2	‡**774**	4508.4	§**454**	2721.2
156	21	†**90**	2030.8	253	4508.4	146	2714.5
†§**454**	22	†**454**	2029.1	604	4501.7	751	2694.3

†Indicates lines in the top yielding 10% of RIL grown in Knoxville, TN in 2011 that were selected using the YPM developed using QTL detected in Knoxville, TN in 2010 by CIM; ‡Indicates lines in the top yielding 10% of RIL grown in Wooster, OH in 2011 that were selected using the YPM developed using QTL detected in Knoxville, TN in 2010 by CIM; §Indicates lines in the top yielding 10% of RIL averaged over Knoxville, TN in 2010, 2011 and Wooster, OH in 2011 that were selected using the YPM developed using QTL detected in Knoxville, TN in 2010 by CIM.



†YPM indicates what environment(s) the data for the model was collected: mean yield, additive effects and additive by additive effects.

Figure 4. The percentage of marker assisted selections (MAS) made in each environment(s) in Group C compared to phenotypic selections (PS) averaged over all environments in Group C. Comparisons were made between the top and bottom % of MAS that were in the corresponding top and bottom % of PS. MAS were made using a yield prediction model (YPM†) developed using QTL detected in each environment(s). PS were based on yield in kg ha^{-1}.

Table 4. Significant (P<0.01) epistatic interactions between loci for yield in 216 RIL in Group C derived from a cross between Essex 86-15-1 x Williams 82-11-43-1. Locus 1 indicates the markers where yield QTL were detected using CIM and locus 2 indicates the markers where QTL(s) were detected using Epistacy in SAS that were interacting with the yield QTL at locus 1.

											ADDITIVE X ADDITIVE EFFECT†	
ENVIRONMENT	LOCUS 1	CHR	MLG	LOC (cM)	FAVORABLE ALLELE	LOCUS 2	CHR	MLG	LOC (cM)	R^2 (%)	E	W
Knoxville, TN 2010	Gm19_46733772_T_C	19	L	84.11	W	GM02_44803277_C_T	2	D1b	99.56	1.53	-1.88	-0.43
						GM05_39686377_T_C	5	A1	88.19	2.21	-1.89	-0.11
						GM06_16450669_T_C	6	C2	36.56	2.74	1.82	-1.58
						GM08_39969061_C_T	8	A2	88.82	2.73	-2.39	-0.40
						GM09_45833394_G_A	9	K	101.85	2.98	-2.34	-0.29
						GM10_36871822_T_G	10	O	81.94	3.26	-0.07	-2.18
						GM13_20628643_G_T	13	F	45.84	3.30	-0.02	-2.15
						GM18_60221294_C_T	18	G	133.83	3.14	-0.15	-2.25
Portageville, MO	Gm13_34751493_C_A	13	F	165.33	W	GM02_46971562_G_A	2	D1b	104.38	5.84	3.21	-0.14
						GM12_34600990_C_T	12	H	76.89	5.43	3.37	0.17
Portageville, MO	Gm09_18969901_T_C	9	K	28.52	W	GM05_32329300_T_G	5	A	71.84	5.42	0.38	-2.91
						GM13_25895304_C_T	13	F	57.55	5.39	0.31	-2.84
						GM17_13589025_G_A	17	D2	30.20	3.89	-2.78	-0.12
Knoxville, TN 2010-11 & Portageville, MO 2011	Gm06_16723946_G_A	6	C2	32.46	W	GM04_46940182_G_T	4	C1	104.31	3.96	-2.75	0.47
						GM06_47833095_T_G	6	C2	106.30	4.72	-3.73	0.19

†Additive by additive effect refers to the quantitative change in yield that is associated with the epistatic combination of the additive genetic effect of locus 1 having the favorable allele with the additive genetic effect of the homozygous state of locus 2 from (E) Essex 15-86-1 or (W) Williams 82-11-43-1.

Table 5. Yield prediction model (YPM) developed using QTL detected in Knoxville, TN in 2010 by CIM to select by MAS the top yielding 10 % of RIL in Group C grown in individual environments and averaged across multiple environments. These MAS lines are indicated in bold.

YPM		YIELD (kg ha^{-1})					
KNOXVILLE, TN 2010		KNOXVILLE, TN 2011		PORTAGEVILLE, MO 2011		KNOXVILLE, TN 2010-11 PORTAGEVILLE, MO 2011	
LINE	RANK	LINE	YIELD	LINE	YIELD	LINE	YIELD
671	01	†**199**	2608.7	‡**213**	5301.3	§**213**	3332.6
†§**932**	02	†**938**	2583.5	352	4911.6	§**450**	3258.7
265	03	†**378**	2561.6	263	4763.8	263	3245.3
†‡§**378**	04	†**448**	2548.2	607	4710.0	§**378**	3178.1
†§**78**	05	†**450**	2539.8	‡**450**	4696.6	§**938**	3157.9
‡§**760**	06	849	2536.4	680	4649.5	§**867**	3124.3
†§**426**	07	†**426**	2529.7	36	4602.5	183	3097.5
†**198**	08	†**63**	2521.3	966	4602.5	908	3090.7
†**523**	09	263	2491.1	908	4595.8	505	3090.7
†**448**	10	183	2470.9	505	4589.1	§**426**	3084.0
†**382**	11	†**78**	2460.8	141	4582.4	607	3063.9
†**620**	12	460	2460.8	‡**760**	4555.5	612	3057.1
†§**938**	13	764	2450.8	165	4508.4	§**760**	3057.1
‡§**213**	14	†**867**	2447.4	320	4481.6	§**78**	3057.1
‡§**378**	15	†**932**	2430.6	‡**1006**	4474.9	165	3043.7
§**553**	16	†**523**	2430.6	‡**867**	4468.1	§**199**	3043.7
†‡§**867**	17	†**198**	2425.6	311	4461.4	§**932**	2996.7
†**63**	18	612	2423.9	572	4461.4	§**553**	2990.0
898	19	359	2418.8	596	4441.3	§**1006**	2983.2
†‡§**450**	20	†**620**	2410.4	‡**378**	4421.1	368	2969.8
‡§**1006**	21	430	2407.1	963	4407.7	803	2963.1
†§**199**	22	†**382**	2395.3	270	4387.5	485	2963.1

†Indicates lines in the top yielding 10% of RIL grown in Knoxville, TN in 2011 that were selected using the YPM developed using QTL detected in Knoxville, TN in 2010 by CIM; ‡Indicates lines in the top yielding 10% of RIL grown in Portageville, MO in 2011 that were selected using the YPM developed using QTL detected in Knoxville, TN in 2010 by CIM; §Indicates lines in the top yielding 10% of RIL averaged over Knoxville, TN in 2010, 2011 and Portageville, MO in 2011 that were selected using the YPM developed using QTL detected in Knoxville, TN in 2010 by CIM.

higher yields than Knoxville, TN and more lines selected using MAS (using only additive effects and the YPM) for the favorable and unfavorable alleles found in Wooster, OH were in the top and bottom yielding lines combined over three environments, respectively. From this we concluded that the more adapted the maturity group of the soybean to the environment the better the MAS were in that environment and across environments. Similar results were seen in Group B. Belleville, IL had higher yields than Knoxville, TN and MAS were considerably better using the Belleville, IL environment.

Group C: Agronomic Traits

In Group C, Portageville, MO had an average yield (3810 kg ha-1) that was significantly ($p<0.01$) higher than the average yield in Knoxville, TN in 2010 (2188 kg ha-1) and 2011 (1915 kg ha-1) The maturity of Group C ranged from an early MG IV to a late MG IV, which are well adapted to Portageville, MO and Knoxville, TN (Sleper, 2006). However, Portageville, MO has growing conditions similar to Milan, TN and in the 2011 Tennessee State Variety Test (TSVT) Milan, TN had higher yields than those in Knoxville, TN in 2011 (Allen, 2011) which supports our observation of higher yield in Portageville, MO then Knoxville, TN. Average maturity was significantly different between Portageville, MO in 2011 (281) and Knoxville, TN in 2011 (271), but no significant difference was seen between Knoxville, TN in 2010 (274) and Knoxville, TN in 2011 (271) or between Knoxville, TN in 2010 (274) and Portageville, MO in 2011 (281).

Group C: MAS Using Only Additive Effects

Using QTL Discovered in Knoxville, TN in 2010 to Predict High Yielding Lines Across Multiple Environments in 2011

In 2010 in Knoxville, TN three QTL were identified for yield using CIM (Table 1). Using MAS to select lines with the favorable allele for these QTL two lines in the top yielding 10% of RIL combined over three environments (Knoxville, TN in 2010 and 2011 and Portageville, MO in 2011) were selected (Figure 3). Further credibility of these yield QTL was demonstrated when three lines in the bottom yielding 10% of RIL combined over the three environments were selected with the unfavorable alleles

for the three QTL identified in Knoxville, TN in 2010 (Figure 3).

Using QTL Discovered in Portageville, MO in 2011 to Predict High Yielding Lines Across Multiple Environments in 2011.

In 2011 in Portageville, MO three QTL were identified for yield using CIM (Table 1). Using MAS to select lines with the favorable alleles for these QTL three lines in the top yielding 10% of RIL combined over three environments (Knoxville, TN in 2010 and 2011 and Portageville, MO in 2011) were selected (Figure 3). Two of these lines selected were among the top yielding 5% of RIL combined over the three environments and ranked 3rd and 4th in yield. Three lines in the bottom yielding 10% of RIL combined over the three environments were selected by MAS with the unfavorable allele for the same three QTL identified in Portageville, MO in 2011 (Figure 3).

Seventeen QTL were detected on chromosomes 1, 2, 3, 5, 6, 7, 9, 11, 12, 13, 16, 18 and 20 using SF ANOVA (Table 1). Using CIM seven QTL were detected on chromosomes 1, 2, 6, 9, 13, 16 and 19 (Table 1). Although, the yields were higher in Portageville, MO in 2011 than in Knoxville, TN in 2010 similar selections were made by MAS across environments (Figure 3). This may be because Knoxville, TN and Portageville, MO are in the same maturity zone for growing soybeans and are similarly adapted for the maturity of Group C. Again, a similar number of top yielding lines were selected by MAS for the favorable allele of the QTL identified using SF ANOVA as MAS for the favorable allele of the QTL identified using CIM in certain instances. However, like in Groups A and B more top yielding lines averaged overall were selected by MAS for the favorable allele of the QTL identified using CIM. In addition, these results agree with the results from Groups A and B that suggest MAS produces better results when using an environment that is adaptable for the maturity group of the soybean.

Group C: YPM Including Mean Yield, Additive and Additive by Additive Effects

In 2010 in Knoxville, TN eight QTL were shown to have a significant interaction with one of the QTL identified for yield using CIM (Table 4). This information was used to develop an YPM to select by MAS high yielding lines in subsequent years. Fourteen lines that were in the top yielding 10% of RIL grown in Knoxville, TN in 2011 were selected by MAS using the YPM and of those selected, eight lines were in the top yielding 5% of RIL grown in Knoxville, TN in 2011, including the top 5 lines (Table 5). This information was also used to develop an YPM to select by MAS high yielding lines across environments. Twelve lines that were in the top yielding 10% of RIL combined over three environments (Knoxville, TN in 2010 and 2011 and Portageville, MO in 2011) were selected by MAS using the YPM and of those selected, six lines were in the top yielding 5% of RIL combined over the three environments, including the top two yielding lines (Table 5). Previously when using only additive effects identified using CIM in Knoxville, TN in 2010 for MAS (Figure 3) without using additive by additive effects in an YPM; only three lines were selected in the top yielding 10% and only one of those lines was in the top yielding 5% of RIL grown over Knoxville, TN in 2010 and 2011 and Portageville, MO in 2011 (Figure 3).

In 2011 in Portageville, MO five QTL were shown to have a significant interaction with two of the QTL identified for yield using CIM (Table 4). This information was used to develop an YPM to select by MAS high yielding lines across environments. Nine lines in the top yielding 10% of RIL combined over three environments (Knoxville, TN in 2010 and 2011 and Portageville, MO in 2011) were selected by MAS using the YPM, including the top yielding line (Figure 3). Previously when using only additive effects identified using CIM in Portageville, MO without using additive effects in the YPM; three lines were selected in the top yielding 10% of RIL combined over Knoxville, TN in 2010 and 2011 and Portageville, MO in 2011 (Figure 3).

From data averaged across Knoxville, TN in 2010 and 2011 and Portageville, MO in 2011, two QTL were shown to have a significant interaction with one of the QTL identified for yield using CIM (Table 1). This information was used to develop an YPM to select by MAS high yielding lines across environments. Sixteen lines in the top yielding 10% of RIL combined over three environments (Knoxville, TN in 2010 and 2011 and Portageville, MO in 2011) were selected by MAS using the YPM (Figure 4). Like in Groups A, B and D, in Group C more top yielding lines were selected using the YPM than using only additive effects for MAS. In Group C similar top and bottom yielding selections were made in Knoxville, TN and Portageville, MO even though the yields were significantly different. Similar selections were also made between Knoxville, TN and Plymouth, NC in Group D and yields between both environments were statistically similar. So, like in Groups A and B the more adaptable the environment to the maturity group of the soybean the better the MAS were in that environment. Also, in Groups C and D when using data collected in one individual environment in the YPM, very few top yielding lines were selected in another individual environment even though the environments were similar in latitude.

In Group C, when using the Knoxville, TN 2010 data to develop an YPM more than 60% of top yielding lines in Knoxville, TN in 2011 were selected by MAS. Using the YPM 14 out of the top 22 yielding lines and 8 out of the top 11 yielding lines grown in Knoxville, TN in 2011 were selected by MAS using QTL identified by CIM from data collected in Knoxville, TN in 2010 (Table 5). This is important to note because when using an YPM it is important for the performance of selections made in one year to carry forth into subsequent years. While this YPM does not predict 100% of the top yielding lines from one year to the next it does indicate that yield predictions using genotypic data warrants further study.

Discussion

In this study predictions made for an individual environment with data collected in that environment were better than predictions made with data averaged from across environments, if one environment was more adaptable to the soybean maturity group. If the environments were similar for adapted maturity, a multi-environment YPM was better for predicting top yielding

lines in multiple individual environments. Bernardo et al. (2008) proposed that if the early generation test environments used for MAS are not representative of the environments in which the lines will be grown, then the results seen in early generation testing might not predict the genotypes that are favorable across a broader sample of environments encountered in subsequent replicated trials. Sebastian et al. (2010) suggested environments with high error variance or environments suspected to be unrepresentative of the targeted environment should be excluded from QTL analysis so that more valid QTL estimates can be obtained to construct the favorable genotype. This agrees with the results found in this study where the environment most adaptable to the maturity group made the best predictions. A comparison of previously reported yield QTL that coincide to the yield QTL reported in this study is available in the supplementary material section.

Yield is a very difficult trait to predict because it can be influenced by many different factors, including genetic and environmental factors and their interactions (Hao et al, 2012; Palomeque et al., 2009, 2012; Sebastian et al., 2010). In other words, it is difficult to use QTL selected from one population evaluated in a few environments to another population evaluated in different environments. There are few reports of validated seed yield QTL in different environments and even fewer validating the reported QTL across diverse genetic backgrounds (Palomeque et al. 2009; Fasoula et al. 2004; Reyna and Sneller 2001). Palomeque et al. (2009) conducted a study to identify yield QTL in two locations with a RIL population derived from a cross of high yielding adapted and high-yielding exotic soybean lines. A cross between Canadian cultivar 'OAC Millennium' and Chinese cultivar, 'Heinong 38.' The population was evaluated in China and Canada in multiple environments from 2004 to 2006. Seven yield QTL were identified of which five were found in at least two year-location environments. Three of the QTL were detected using multiple QTL mapping (MQM) and four were detected using SF ANOVA. To validate these seven QTL Palomeque et al. (2010) evaluated a cross between Canadian cultivar 'Pioneer 9071' and Chinese cultivar '8902' in two locations in China and five locations in Canada in 2005 and 2006. No association between seed yield and the previously identified QTL was observed. However, one of the seven QTL evaluated by Palomeque et al. (2010) was previously reported as being associated with seed yield in diverse genetic backgrounds and environments by other researchers (Guzman et al. 2007; Orf et al. 1999; Smalley et al. 2004; Specht et al. 2001).

Hao et al. (2012) evaluated a population of 191 soybean landraces in five environments to detect molecular markers associated with soybean yield and its components using 1,536 SNPs. Using genome-wide association, they identified 19 SNPs associated with yield. Most SNPs were detected only in a specific environment and only a small number of SNPs were identified in three or more environments.

Maturity has also been shown to affect the verification or validation of yield QTL in soybean. Kabelka et al. (2004) reported that only two out of fifteen yield QTL were detected across three maturity groups (MG II, MG III and MG IV). In this study most QTL were detected in at least two groups, but some were only found in one group. In addition, some QTL detected by Kabelka et al. (2004) in only one maturity group were found in multiple maturity groups in this study. This indicates that while some yield QTL may not be specific to particular maturity groups other yield QTL may be specific to maturity groups within certain genetic backgrounds. Although some of the genomic regions explained a small portion of genotypic variation, or were identified only in a specific environment, they could be important to understanding the genetic control of soybean seed yield. Evaluation of these QTL in distinct environments and in different genetic backgrounds along with demonstrated effectiveness of MAS will be the true test of the concept of molecular breeding for seed yield.

The environment and genetic background both play an essential part in determining the success of using MAS. QTL for a specific trait are not always stable across environments and/or genetic backgrounds, therefore, their breeding value depends on the strength and stability of trait associations. When yield QTL are evaluated in diverse genetic backgrounds a number of different results can be produced. Epistatic effects could be regarded as one of the main reasons for the limited success in validating QTL across different populations and environments. Another possibility could be that the variability between the parental lines used to derive these populations is limited, i.e. the parents of the validation population or the current mapping population have less genetic variation than the parents used to form the population for QTL detection. Potentially, with the genetic diversity of the parents in this study and the diverse ancestry of each parent, the yield QTL found in this study might be found in different populations. In this study yield prediction models including epistatic effects were used to predict top yielding lines.

When using the YPM to make predictions the data collected from the environment that was more adaptable to a particular maturity group made the best selections in that environment and across environments. This was prominent in Groups A and B where the maturity groups were more adapted to the locations in OH and IL which are more northern in latitude than Knoxville, TN. In Groups C and D the multi-environment YPM predicted more top yielding lines in each individual environment and across environment compared to each individual environment being able to predict top yielding lines in other environments and across environments. Further research is needed to determine the best overall YPM to use to predict top yielding lines.

When making selections using only the marker information and using the maker information combined with additive effects and additive by additive effects, MAS performed with significant markers identified using CIM as carried out with R/qtl. However, MAS performed with significant markers identified by SF ANOVA carried out with SAS sometimes made similar predictions and in a few instances better predicted the top yielding lines. While using the program Epistacy (Holland, 1998) to determine the additive by additive effects of significant markers that were pre- determined using SF ANOVA and CIM, it was determined that Epistacy could be used to scan all pairwise interactions to detect significant interactions. This would greatly decrease the time needed to test pairwise combinations of >1000 SNPs (results not reported in this study). In addition, more additive by

additive effects (epistatic effects) could be used in the YPM. These interactions where neither marker identifies a significant effect, but where the two markers together create a significant epistatic effect could be very valuable in predicting quantitative traits. Thus Epistacy could help eliminate the need to test multiple statistical programs for MAS and simplify the process of using epistatic interactions in genomic selection.

Previous research has suggested that including MAS for yield QTL in a breeding program can increase the genetic gain for yield. Sebastian et al. (2010) conducted a study in which F7:8 lines derived from elite cultivars were grown as plant-row yield trials in three environments. The objective of that study was to select for an improved genotype. Analysis was done using a mixed linear model and at statistically significant loci, the allele associated with the highest yield mean was considered the favorable allele for the purpose of selecting higher-yielding lines. The yield potential of the selected lines was then compared to their respective parents across multiple environments and years. The seed yields of the reselected lines were greater than the original five elite cultivars by an average of 3.1% and yield gains of up to 5.8% were confirmed in some of the selected lines. Two of the improved lines were released as improved cultivars.

There are only a few reported studies on using MAS for improving quantitative traits where the QTL were confirmed across different populations. Most studies refer mainly to computer simulations using various data sets. Campos et al. (2009) adapted the Bayesian LASSO to arrive at a regression model where markers, pedigrees and covariates other than markers are considered jointly. The model was fitted to two data sets from wheat and mouse populations. Results showed that models using molecular markers had better prediction accuracy of grain yield in wheat than those based on pedigree. Crossa et al. (2010) conducted a MAS study using a wheat data set containing various traits, including yield and a maize data set with two disease traits. Separate models were fitted to each trait and environment. Results indicated models including marker information led to improved predictive ability, but estimates of marker effects were different across environments. It was speculated that multiple environment prediction would allow information to be borrowed between correlated environments and could yield similar or even better predictions for individual environments. Using only 80 markers and 126 soybean RIL Hu et al. (2011) used MAS to predict the genomic value of somatic embryo number for each line. The correlation coefficient between the observed and predicted embryo numbers was 0.33 when only the additive effects were used in prediction. When the epistatic effects were also included in the model, the correlation coefficient increased to 0.78. Data analysis was conducted using PROC QTL in SAS. However, when marker density is high, the Bayesian method in that QTL procedure (as used in their study) may be limited for handling all pair-wise interactions.

Quantitative traits are controlled by multiple QTL. The contribution of each locus may be small or large, but the collective contribution of all loci is often significant. Including epistatic effects to predict the genomic values of plants can achieve enhanced gains for soybean improvement. The results from this study suggest using an YPM with additive and additive by additive effects detected from environments that are similar in latitude may lead to the best YPM for predicting seed yield in multiple individual environments. However, more top yielding lines in an individual environment can be predicted using an YPM with additive and additive by additive effects detected from the environment in which the selections will be made.

Conclusion

This study suggests that environment specific data continues to be valuable and that while MAS can successfully predict high yielding lines, it might miss some of the very top yielding lines unless the prediction equation includes data from the environment in which the yield trial is conducted. This begs the question of resource management and effectiveness in identifying the most superior individuals in a population for a targeted trait of low heritability, like yield. Nevertheless, this study proves MAS from one year can successfully identify some of the top yielding lines in subsequent years and distant environments. This leads to the credibility of continuing further research to enhance the YPM approach for improved efficiency. With the knowledge of the QTL segregating in our Essex x Williams 82 population along with QTL discovered from other mapping populations, plant breeder and other genetic researchers should have a more complete picture of which QTL are available to utilize as tools for soybean yield improvement by MAS.

Acknowledgements

The authors would like to thank the United Soybean Board, the Tennessee Soybean Promotion Board, the Tennessee Agricultural Experiment Station, the University of Missouri, North Carolina State University, Southern Illinois University, Ohio State University and the Soybean Genomics Lab, USDA-ARS, Beltsville, MD for making this research possible. Their support is greatly appreciated.

Conflict of Interest: The authors declare that they have no conflict on interest.

References

Allen FA, R Johnson, RC Williams Jr., AT McClure, M Newman, and P Donald (2011) Soybean variety performance test in Tennessee. University of Tennessee, Knoxville, TN.

Bernard RL, and DA Lindahl (1972) Registration of 'Williams' soybean. Crop Sci 12:716.

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

Broman KW, and S Sen (2009) A guide to QTL mapping with R/qtl. Springer, New York.

Campos G, H Naya, D Gianola, J Crossa, A Legarra, E Manfredi, K Weigel, and JM Cotes

(2009) Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genet 182:375-385.

Crossa J, G Campos G, P Perez, D Gianola, J Burgueno, JL Araus, D Makumbi, RP Singh,

S Dreisigacker, J Yan, V Arief, M Banziger, and HJ Braun (2010) Pre-

diction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. Genet 186:713-724.

Csanadi G, J Vollman, G Stift, and T Lelley (2001) Seed quality QTL identified in a molecular

map of early maturing soybean. Theor Appl Genet 103:912–919.

Du W, M Wang, S Fu, and D Yu (2009) Mapping QTL for seed yield and drought susceptibility index in soybean (*Glycine max* L.) across different environments. J Genet Genomics 36:721-731.

Fasoula VA, DK Harris, and HR Boerma (2004) Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. Crop Sci 44:1218–1225.

Gizlice Z, TE Carter Jr., TM Gerig, and JW Burton. 1996. Genetic diversity patterns in North American public soybean cultivars based on coefficient of parentage. Crop Sci 36:753–765.

Guzman PS, BW Diers, DJ Neece, SK Martin, AR Leroy, CR Grau, TJ Hughes, and RL Nelson (2007) QTL associated with yield in three backcross-derived populations of soybean. Crop Sci 47:111–122.

Hao D, H Chang, Z Yin, S Cui, D Zhang, H Wang, and D Yu (2012) Identification of single nucleotide polymorphisms and haplotypes associated with yield and yield components in soybean (*Glycine max*) landraces across multiple environments. Theor Appl Genet 124: 447-458.

Heatherly LG, and JR Smith (2004) Effect of soybean stem growth habit on height and node number after beginning bloom in the midsouthern USA. Crop Sci 44:1855–1858.

Holland JB (1998) Epistacy: A SAS program for detecting two-locus epistatic interactions using genetic marker information. J Her 89: 364-375.

Hyten DL, Q Song, Y Zhu, IY Choi, RL Nelson, JM Costa, JE Specht, RC Shoemaker, and PB Cregan (2006) Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA 103:16666–16671.

Hyten DL, Q Song, Y Zhu, IY Choi, MS Yoon, JE Specht, LK Matukumalli, RL Nelson, RC Shoemaker, ND Young, and PB Cregan (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet 116: 945-952.

Hu Z, Y Li, X Song, Y Han, X Cai, S Xu, and W Li (2011) Genomic value prediction for quantitative traits under the epistatic model. BMC Genet 12:15.

Kabelka EA, BW Diers, WR Fehr, AR LeRoy, IC Baianu, T You, DJ Neece, and RL Nelson (2004) Putative alleles for increased yield from soybean plant introductions. Crop Sci 44:784–791.

Li J, K Das, G Fu, R Li, and R Wu (2011) The Bayesian lasso for genome-wide association studies. Bio. 27:516-523.

Long N, D Gianola, GJM Rosa, and KA Weigel (2011) Dimension reduction and variable selection for genomic selection: application to predicting milk yield in Holsteins. J Anim Breed Genet 128:247-257.

Macciotta, N., G Gaspa, R Steri, E Nicolazzi, C Dimauro, C Pieramati, and A Cappio-Borlino (2010) Using eigenvalues as variance priors in the prediction of genomic breeding values by principal component analysis. J Dairy Sci 93:2765-2774.

Orf JH, K Chase, T Jarvik, LM Mansur, PB Cregan, FR Adler, and KG Lark (1999) Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. Crop Sci. 39:1642-1651.

Palomeque LC, LL Jun, W Li, BR Hedges, ER Cober, I Rajcan (2009) QTL in mega-environments: I. Universal and specific seed yield QTL detected in a population derived from a cross of high yielding adapted x high-yielding exotic soybean lines. Theor Appl Genet 119:417–427.

Palomeque LC, LJ Liu, W Li, BR Hedges, ER Cober, MP Smid, L Lukens, and I Rajcan (2010) Validation of mega-environment universal and specific seed yield QTL associated with seed yield and agronomics traits in soybeans. Theor. Appl. Genet. 120:997–1003.

Reinprecht Y, VW Poysa, K Yu, I Rajcan, GR Ablett, and KP Pauls (2006) Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.) Merrill) germplasm. Genome 49(12):1510-1527.

Reyna N, and CH Sneller (2001) Evaluation of marker-assisted introgression of yield QTL alleles into adapted soybean. Crop Sci 41:1317–1321.

Rincker K, R Nelson, J Specht, D Sleper, T Cary, SR Cianzio, S Casteel, S Conley, P Chen, V Davis, C Fox, G Graef, C Godsey, D Holshouser, GL Jiang, SK Kantartzi, W Kenworthy, C Lee, R Mian, L McHale, S Naeve, J Orf, V Poysa, W Schapaugh, G Shannon, R Uniatowski, D Wang, and B Diers (2014) Genetic improvement of U.S. soybean in maturity groups II, III and IV. Crop Sci 54:1-14.

SAS Institute (2007) SAS/STAT 9.1.3. SAS Inst., Cary NC.

Sebastian SA, LG Streit, PA Stephens, JA Thompson, BR Hedges, MA Fabrizius, JF Soper, DH Schmidit, RL Kallem, MA Hinds, L Feng, and JA Hoeck (2010) Context-specific marker-assisted selection for improved grain yield in elite soybean populations. Crop Sci 50:1196-1206.

Sleper DA, and JM Poehlam (2006) Breeding Field Crops. Blackwell Publishing Ames, IA, pp. 130-133.

Smalley MD, WR Fehr, SR Cianzio, F Han, SA Sebastian, LG Streit (2004) Quantitative trait loci for soybean seed yield in elite and plant introduction germplasm. Crop Sci 44:436–442.

Sneller CH (1994) Pedigree analysis of elite soybean lines. Crop Sci 34:1515–1522.

Smith TJ, and HM Camper (1973) Registration of 'Essex' soybean. Crop Sci 13:495.

Solberg T, A Sonesson, J Woolliams, T Meuwissen (2009) Reducing dimensionality for prediction of genome-wide breeding values. Genet Sel Evol 41:29.

Song QJ, LF Marek, RC Shoemaker, KG Lark, VC Concibido, X Delannay, JE Specht, and PB Cregan (2004) A new integrated genetic linkage map of the soybean. Ibid. 109:122–128.

Song QJ, DL Hyten, G Jia, CV Quigley, EW FicFickus, RL Nelson, and PB Cregan (2013) Development and evaluation of SoySNP50K, a high density genotyping array for soybean. PLoS One 8(1):e54985

Soy Stats (2013) A reference guide to important soybean facts and figures. (www.soystats.com/ (verified 03/18/14).

Wang D, GL Graef, AM Procopiuk, and BW Diers (2004) Identification of putative QTL that underlie yield in interspecific soybean backcross populations. Theor Appl Genet 108:458–467.

Plant Genetics, Genomics, and Biotechnology - ISSN 2332-2012. Published By Atlas Publishing, LP (www.atlas-publishing.org)

A Comparison of Previously Reported Yield QTL That Coincide With the Yield QTL Reported in This Study

Chromosome 1

A yield QTL was identified on chromosome 1 associated with marker Gm01_1494600_C_T (5.52 cM) using SF ANOVA and marker Gm01_1045893_G_A (5.88 cM) using CIM (Table 1). Also, markers Gm01_1241762_A_C (4.6 cM) and Gm01_2747136_A_C (11.28 cM) were identified using SF ANOVA and associated with the same yield QTL. Two other yield QTL were identified using SF ANOVA further down the chromosome near markers Gm01_29787876_G_A (59.29 cM) and Gm01_47115450_G_T (70.15 cM) and Gm01_54171147_G_T (118.27 cM) (Table 1). Kabelka et al. (2004) conducted a QTL study with three maturity groups (MG II, MG III and MG IV) and in MG IV they detected a QTL for seed yield on chromosome 1 (position not reported). Smalley et al. (2004) reported three yield QTL on chromosome 1 in regions similar to the ones reported in this study. The objective of their study was to identify QTL for yield in elite and PI germplasm using three populations that differed in their percent of PI parentage. They reported three yield QTL significantly associated with markers Satt184 (8.3 cM), Satt368 (41.1 cM) and Satt436 (89.3 cM), respectively.

Chromosome 2

In Group A a yield QTL on chromosome 2 was identified in each individual environment and across all environments using SF ANOVA. This yield QTL was linked to markers Gm02_47790307_C_T (121.66 cM) and Gm02_49126947_T_C (127.25 cM) in Group A. The same QTL was also associated with markers Gm02_44803277_C_T (107.06 cM) using SF ANOVA in Group C. CIM linked it to marker Gm02_44803277_C_T (114.09 cM) and Gm02_42469280_A_C (105.17 cM) (Table 1). A yield QTL was also identified on chromosome 2 near marker Gm02_49746270_A_G (146.54 cM) using SF ANOVA and Gm02_47790307_C_T (150.38 cM) using CIM. Another yield QTL on chromosome 2 was linked to marker Gm02_12770553_A_G (46.15 cM) using SF ANOVA and markers Gm02_6821311_A_C (38.24 cM) and Gm02_6820177_A_C (38.07 cM) using CIM. Smalley et al. (2004) reported a yield QTL on chromosome 2 linked to marker Satt 141 (52.8 cM) and Du et al (2009) reported a yield QTL near marker Satt546 (110 cM) on chromosome 2 in a RIL population from a cross between Kefeng1 and Nannong 1138-2.

Chromosome 3

On chromosome 3 only two QTL were identified with both SF ANOVA and CIM. Using SF ANOVA QTL were identified near markers Gm03_5264953_A_G (19.43 cM) and Gm03_39552601_T_C (87.68 cM) (Table 1). CIM identified these QTL near markers Gm03_2151432_A_G (14 cM) and Gm03_39559139_G_A (93.64 cM). Smalley et al (2004) detected two yield QTL linked to markers Satt152 (16.3 cM) and Satt_091 (95.5 cM). In our study SF ANOVA also identified three yield QTL associated with markers Gm03_47386481_A_C (120.71 cM), Gm03_838582_T_C (4.68 cM) and Gm03_21003884_A_G (44.15 cM). Smalley et al. (2004) also reported a yield QTL linked to marker Satt584 (35.4 cM), but no studies have reported any yield QTL in the region around the other two markers we identified using SF ANOVA.

Chromosome 4

A yield QTL on chromosome 4 was identified in Knoxville, TN in 2010 and Wooster, OH in 2011 in Group A using CIM near markers Gm04_48782140_G_T (152.98 cM) and Gm04_48993297_T_G (154.16 cM), respectively. Another yield QTL on chromosome 4 was identified in both in Knoxville, TN in 2010 and across Knoxville, TN in 2010, 2011 and Plymouth, NC in 2011 in Group D using SF ANOVA near markers Gm04_8247949_C_T (65.87 cM) and Gm04_8845668_G_T (63.93 cM), respectively (Table 1). Guzman et al. (2007) identified a yield QTL on chromosome 4 associated with marker Satt399 (76.2 cM), which is the same region where Yuan et al. (2002) mapped a QTL in an Essex x Forrest cross. Yuan et al. (2002) reported that the yield QTL was only detected in one of four environments, while Guzman et al. reported the yield QTL was detected across four environments in 2004 and averaged across 2003 and 2004. Three yield QTL on chromosome 4 were also identified by Smalley et al. (2004) near markers Satt578 (74 cM), Satt294 (105 cM) and Satt338 (173 cM). The location of these markers and the one reported in this study indicates that there may be a large region on chromosome 4 responsible for yield QTL.

Chromosome 5

Markers Gm05_31399360_G_A (41.55 cM), Gm05_30953466_G_T (39.76 cM) using SF ANOVA and Gm05_33176582_G_A (33.77 cM) using CIM were linked to a yield QTL on chromosome 5 (Table 1). The yield QTL on chromosome 5 by Guzman et al. (2007) was near marker Satt300 (30.9 cM) in 2003, 2004 and across years. Using SF ANOVA a yield QTL was identified on chromosome 5 linked to marker Gm05_1128604_A_G (3.24 cM) and a yield QTL linked to marker Gm05_34850619_C_T (72.38 cM). CIM identified one additional QTL associated with marker Gm05_3485480_T_C (19.73 cM). A yield QTL linked to Satt276 (5.1 cM) and another yield QTL linked to markers Satt385 (69.9 cM) and Satt545

Table 1. Significant (P<0.01) epistatic interactions between loci for yield in 221 RIL in Group B derived from a cross between Essex 86-15-1 x Williams 82-11-43-1. Locus 1 indicates the markers where yield QTL were detected using R/qtl and locus 2 indicates the markers where QTL(s) were detected using Epistacy in SAS that were interacting with the yield QTL at locus 1.

											ADDITIVE X ADDITIVE EFFECT†	
ENVIRONMENT	LOCUS 1	CHR	MLG	LOC (cM)	FAVORABLE ALLELE	LOCUS 2	CHR	MLG	LOC (cM)	R^2 (%)	E	W
Knoxville, TN 2010	Gm07_16144523_C_A	7	M	51.90	W	GM01_46579445_G_A	1	D1a	109.34	3.84	-3.14	-0.50
						GM05_39673657_T_G	5	A1	93.13	5.84	0.07	-3.09
						GM06_19653985_A_G	6	C2	46.14	3.28	4.47	-2.21
						GM11_17113172_G_A	11	B1	40.17	6.25	-3.25	-0.06
Knoxville, TN 2010	Gm06_17617727_G_T	6	C2	55.04	W	GM07_42111727_C_T	7	M	98.85	4.61	0.16	-4.33
						GM14_12556387_T_C	14	B2	29.48	3.95	-4.82	-0.71
						GM20_44554028_G_A	20	I	104.59	4.05	-4.47	-0.56
Belleville, IL 2011	Gm06_20996124_T_C	6	C2	60.21	W	GM01_29990637_T_C	1	D1a	70.40	11.69	-12.86	-1.60
						GM02_13771227_A_G	2	D1b	32.33	5.09	-5.87	-0.45
						GM03_37376203_C_T	3	N	87.74	6.20	0.16	-5.94
						GM04_48819142_A_C	4	C1	114.60	8.85	-9.47	-1.41
						GM07_35091912_G_T	7	M	82.38	7.20	-1.02	-7.42
						GM08_12693852_G_A	8	A2	29.80	9.62	-12.82	-2.23
						GM10_47833380_A_G	10	O	112.28	9.09	-2.18	-12.64
						GM11_36811720_C_A	11	B1	86.41	3.41	-5.78	-1.44
						GM12_33656706_G_A	12	H	79.01	11.36	-1.32	-13.05
						GM13_26705499_C_T	13	F	62.69	11.37	-1.30	-12.39
						GM14_19103544_T_C	14	B2	44.84	12.69	-1.20	-12.80
						GM15_49375283_T_C	15	E	115.90	11.04	-1.40	-12.77
						GM16_29081010_A_G	16	J	68.27	5.32	-0.20	-6.09
						GM17_36966551_A_C	17	D2	86.78	10.09	-12.87	-2.11
						GM18_52455765_C_A	18	G	123.14	4.69	-5.38	-0.42
						GM19_33586981_A_G	19	L	78.84	3.95	-1.80	-7.12
Belleville, IL 2011	Gm02_44803277_C_T	2	D1b	114.09	W	GM01_51416475_G_A	1	D1a	120.70	3.57	-2.14	-0.19
						GM02_11182262_C_T	2	D1b	26.25	4.65	-2.50	-0.23
						GM06_41416032_T_C	6	C2	97.22	3.55	-3.85	-0.92
						GM07_15590266_C_T	7	M	36.60	5.25	0.09	-2.28
						GM17_15834164_T_C	17	D2	37.17	3.56	-2.04	-0.10

Table 1. Continued.

ENVIRONMENT	LOCUS 1	CHR	MLG	LOC (cM)	FAVORABLE ALLELE	LOCUS 2	CHR	MLG	LOC (cM)	R^2 (%)	ADDITIVE X ADDITIVE EFFECT† E	ADDITIVE X ADDITIVE EFFECT† W
Knoxville, TN 2010-11												
Belleville, IL 2011	Gm07_17362808_A_G	7	M	55.95	W	GM_01-5021663_A_G	1	D1a	11.79	3.98	-0.27	-1.68
						GM06_20835584_T_C	6	C2	48.91	4.90	3.04	-1.35
						GM06_20996124_T_C	6	C2	62.03	4.82	2.43	-1.40
						GM13_26707540_C_T	13	F	62.69	4.42	-1.58	-0.08
						GM15_11274131_A_G	15	E	26.47	3.60	-1.56	-0.18
						GM18_58266066_T_C	18	G	136.77	5.41	-1.63	-0.03
Knoxville, TN 2010-11												
Belleville, IL 2011	Gm06_20996124_T_C	6	C2	62.03	W	GM01_29990637_T_C	1	D1a	70.40	4.93	31.75	36.68
						GM07_17460956_C_A	7	M	40.99	4.79	36.21	35.23
						GM08_12693852_G_A	8	A2	29.80	4.71	31.89	36.46
						GM10_47858822_C_T	10	O	112.34	4.03	36.51	31.63
						GM12_33657269_G_T	12	H	79.01	4.83	36.55	31.77
						GM13_26707540_C_T	13	F	62.69	3.72	36.55	31.42
						GM15_49375283_T_C	15	E	115.90	5.43	36.55	31.18
						GM19_45082401_G_A	19	L	105.83	4.78	36.56	31.72
Knoxville, TN 2010-11												
Belleville, IL 2011	Gm02_42469280_A_C	2	D1b	105.17	W	GM12_34378311_T_C	12	H	80.70	5.31	0.17	-1.58
						GM16_29150479_A_G	16	J	68.43	3.65	0.15	-1.28

† Additive by additive effect refers to the quantitative change in yield that is associated with the epistatic combination of the additive genetic effect of locus 1 having the favorable allele with the additive genetic effect of the homozygous state of locus 2 from (E) Essex 15-86-1 or (W) Williams 82-11-43-1

(75.3 cM) were reported by Smalley et al. (2004).

Chromosome 6

Satt557 (112.5 cM) was detected in 2003, 2004 and across years by Guzman et al. (2007) to be linked to a yield QTL on chromosome 6. However, they only reported marker Satt640 (30.5 cM) was linked to yield QTL on chromosome 6 in 2003. Specht et al. (2001) reported a yield QTL linked to marker Satt281 (43.6 cM) on chromosome 6, which was 10 cM from Satt640 (30.5 cM) reported by Guzman et al. in 2007. Smalley et al (2004) reported a yield QTL linked to Sat_062 (29.2 cM). These finding agree with the yield QTL linked to marker Gm06_10864751_A_G (24.86 cM) found in Portageville, MO in 2011 using SF ANOVA and marker Gm06_16723946_G_A (32.46 cM) found across environments using CIM in Group C in our study. Another yield QTL was found in Group B in both individual environments and across environments using both SF ANOVA and CIM associated with markers Gm06_17617727_G_T (55.04 cM), Gm06_20996124_T_C (60.21 cM) and Gm06_20996124_T_C (62.03 cM) identified using CIM and Gm06_20996124_T_C (58.54 cM) and Gm06_27540819_T_G (66.24 cM) identified using SF ANOVA. Kabelka et al. (2004) only reported one yield QTL on chromosome 6 and it was detected across three maturity groups (MG II, MG III and MG IV) and averaged over twelve environments.

Chromosome 7

Two yield QTL on chromosome 7 have been reported by Specht et al. (2001) near markers Satt150 (17.6 cM) and Satt567 (36.2 cM) and Smalley et al. (2004) reported two yield QTL near markers Satt 590 (12.4 cM) and Satt567 (45.5 cM). Orf et al. (1999) also reported a yield QTL near Satt150 (16.1 cM). In this study one yield QTL was identified using SF ANOVA linked to marker Gm07_4837493_A_G (11.06 cM) and marker Gm07_149664_T_C (1.34 cM) and marker Gm07_4008483_C_T (5.19 cM) using CIM (Table 1). Another yield QTL was linked to makers Gm07_17460956_C_A (39.95 cM), Gm07_16814628_C_T (38.47 cM) and Gm07_18539902_T_G (42.42 cM) using SF ANOVA and Gm07_16144523_C_A (51.90 cM), Gm07_17362808_A_G (55.95 cM) and Gm07_18539902_T_G (61.37 cM) using CIM.

Chromosome 8

Only one yield QTL was identified on chromosome 8 using SF ANOVA and it was linked to Gm08_15866777_G_A (22.31 cM) (Table 1). No QTL were found using CIM. Smalley et al. (2004) linked Satt493 (23.3 cM) to a yield QTL on chromosome 8, but no other studies were found that reported a yield QTL on chromosome 8.

Chromosome 9

Yuan et al. (2002), Kabelka et al. (2004) and Smalley et al. (2004) reported yield QTL near marker Satt119 (20.3 cM) on chromosome 9. In this study a yield QTL was mapped near markers Gm09_18969901_T_C (28.52 cM) detected using CIM and Gm09_12463468_C_T (31.76 cM) detected using SF ANOVA. Guzamn et al. (2007) reported a yield QTL across 2003 and 2004 linked to Satt046 (45.6 cM) on chromosome 9. Smalley et al. (2004) and Yuan et al. (2002) also reported a yield QTL near markers Satt087 (7.3 cM) and Satt539 (4.03 cM), respectively on chromosome 9. Yield QTL were also reported linked to markers Satt544 (72.8 cM) and Satt 273 (120 cM) by Smalley et al. (2004). In this study a yield QTL was associated with Gm09_6967374_C_T (15.94 cM), Gm09_3394608_G_A (7.76 cM) and Gm09_457853_A_G (5.23 cM) detected using SF ANOVA and Gm09_2634593_G_A (5.62 cM) using CIM. Also, a yield QTL was identified near marker Gm09_34191288_T_C (78.24 cM) using SF ANOVA.

Chromosome 10 & 11

No QTL were reported on chromosome 10 or 11 using CIM. Using SF ANOVA two yield QTL were detected on chromosome 10 associated with Gm10_47585270_T_G (108.89 cM)/ Gm10_48428720_T_C (110.82 cM) and Gm10_571698_A_G (1.3 cM) (Table 1). Kalbelka et al. (2004) and Smalley et al. (2004) reported a QTL for yield associated with Satt358 (2.4 cM) on chromosome 10. Satt358 was detected across three maturity groups (MG II, MG III and MG IV) averaged across twelve environments by Kalbelka et al. (2004). Csanadi et al. (2001) also detected an association between seed weight and Satt358. An additional yield QTL was reported by Smalley et al. (2004) associated with Satt477 (103.8 cM), Satt592 (120.5 cM) and Satt331 (127.9 cM). Two yield QTL were also detected on chromosome 11 associated with Gm11_5773052_G_A (20.42 cM)/ Gm11_7323949_A_G (26.24 cM)/ Gm11_7445495_G_A (26.72 cM)/ Gm11_4453218_T_C (16.23 cM) and Gm11_36807939_C_A (84.22 cM) using SF ANOVA. Only one study has reported yield QTL within 10 cM of marker Gm11_36807939_C_A (84.22 cM). Smalley et al. (2004) reported a yield QTL linked to markers Satt444 (76.4 cM) and Satt359 (92.1 cM), respectively. In addition, they reported Satt509 (26.7 cM) was associated with a yield QTL on chromosome 11. Du et al (2009) reported a yield QTL near markers at 36.4 cM and 9.61 cM on chromosome 11.

Chromosome 12

Three yield QTL were detected on chromosome 12 using SF ANOVA and CIM. Using SF ANOVA markers Gm12_1594873_A_G (3.64 cM) and Gm12_39962521_A_G (91.44 cM) were linked to two different yield QTL. Du et al. (2009) and Kalbelka et al. (2004) reported a yield QTL near markers at 86 cM on chromosome 12. No studies were found that reported a yield QTL near a marker at 3 cM on chromosome 12, but Du et al. (2009) did report a yield QTL associated with marker Satt317 (11.71 cM). For our study using CIM only one QTL was detected and it was associated with marker Gm12_7135310_A_G (36.25 cM). Only one study was found that reported a yield QTL in the same region linked to marker

Satt192 (41.1 cM) (Smalley et al., 2004).

Chromosome 13

One yield QTL was identified on chromosome 13 linked to markers Gm13_27348409_A_G (150.28 cM), Gm13_32183364_A_C (162.13 cM), and Gm13_29895148_C_T (154.76 cM) using SF ANOVA and Gm13_34751493_C_A (165.33 cM) and Gm13_27092408_C_T (150.77 cM) using CIM. Another yield QTL was identified using SF ANOVA linked to Gm13_34946643_T_C (180.68 cM). In 2001, Specht et al. reported Satt074 (143.40 cM) was linked to a yield QTL in a Minsoy x Noir 1 population of 236 RIL genotyped at 665 loci. In 2004, Smalley et al. reported Sat_074 (181.8 cM) to be linked to a yield QTL in two different populations with 184 SSR markers spaced 15 cM apart. The proximity of these markers and the span in which they stretch may indicate that the same yield QTL may have been detected in all studies.

Chromosome 14

Only one QTL was associated with yield on chromosome 14 (linked to Gm14_49107190_G_A) using SF ANOVA and no QTL were detected using CIM. Concibido et al. (2003), Smalley et al. (2004) and Kabelka et al. (2004) reported a yield QTL detected by Satt168 (94 cM) on chromosome 14, which is 8 cM below Gm14_49107190_G_A (102.52 cM). Orf et al. (1999) and Smalley et al. (2004) reported yield QTL linked to Satt066 (97.3 cM), which is 5 cM from Gm14_49107190_G_A (102.52 cM).

Chromosome 15

In this study only one yield QTL was mapped on chromosome 15 using SF ANOVA and CIM associated with markers Gm15_48028533_G_A, Gm15_43797502_G_T and Gm15_49231503_C_T at 72.40 cM, 72.68 cM, and 89.13 cM, respectively. A yield QTL was reported by Wang et al (2004) on chromosome 15 linked to marker Satt575 (2.3 cM).

Chromosome 16

The yield QTL on chromosome 16 linked to Gm16_6262227_C_T (10.66 cM), Gm16_5735654_A_G (8.95 cM), Gm16_6233586_A_G (14.23 cM), Gm16_6496577_A_C (14.86 cM) and Gm16_1339719_T_C (6.55 cM) is in the same region as the yield QTL mapped by Orf et al. (1999) and Guzman et al. (2007). Both studies mapped the QTL to markers near 11.7 cM on chromosome 16. In the population in the Guzman et al. (2007) study another yield QTL was mapped to chromosome 16 associated with Satt215 (44.8 cM) only in 2004. In the same population a yield QTL associated with Satt547 (67.7 cM) was detected in 2003, 2004 and across years. In a different population Satt414 (37.8 cM) and Satt622 (42.4 cM) were linked to a yield QTL in 2004 and across years, respectively.

Chromosome 17

A yield QTL identified by single factor ANOVA associated with Gm17_13240263_C_T (30.29 cM) was in the same region as the yield QTL identified by CIM associated with Gm17_32687336_C_T (49.59 cM) and Gm17_12822621_A_G (35.12 cM). Reinprecht et al. (2006) and Orf et al (1999) identified a yield QTL associated with Satt002 (46.73 cM) and Smalley et al. (2004) identified a yield QTL associated with Satt135 (34.7 cM) and Satt458 (34.7 cM). The proximity of these markers also indicates that the same yield QTL may have been detected in all studies, providing evidence for the credibility of MAS for yield utilizing this locus.

Chromosome 18

On chromosome 18 three yield QTL were detected using SF ANOVA. One yield QTL was associated with markers Gm18_8772679_T_C (33.67 cM), Gm18_23913313_A_G (54.72 cM) and Gm18_15660496_T_G (44.64 cM). The second QTL was associated with Gm18_265662_T_C (1.19 cM) and the third QTL was associated with Gm18_58055444_T_C (112.85 cM). Smalley et al. (2004) also identified three yield QTL on chromosome 18 associated with Satt309 (1.9 cM), Satt324 (25.9 cM) and Satt517 (103.2 cM), respectively. Satt324 has also been associated with a yield QTL on chromosome 18 at 37.47 cM (Reinprecht et al., 2006) and on chromosome 18 at 42.38 cM (Kabelka et al., 2004). CIM detected a yield QTL on chromosome 18 linked to Gm18_57988264_A_G (78.75 cM). In 2009, Du et al. reported a yield QTL associated with Satt223 (76.81 cM) and Satt288 (88.01 cM). These makers and the two reported in this study using CIM are 25 cM from Satt517, which indicates that they are independent QTL. However, Satt517 and Gm18_58055444_T_C are less than 10 cM apart and may be identifying the same QTL.

Chromosome 19

In Group A one yield QTL on chromosome 19 was identified in each individual environment and across environments using both SF ANOVA and CIM associated with Gm19_44937486_T_C (70.75 cM), Gm19_45198812_C_A (72.00), Gm19_44955912_T_G (76.84 cM), and Gm19_44964042_C_T (76.91 cM). Also, in one individual environment in Group B and Group D markers Gm19_45062248_T_C (77.05 cM) and Gm19_39246602_T_C (73.68 cM) were associated with the same QTL using SF ANOVA. The same QTL was identified in Group C associated with marker Gm19_46733772_T_C (84.11 cM) using CIM. The large effect of this interval on chromosome 19 could be due to the gene for growth habit (Dt1) which is located in the same interval at 89.1 cM. The locus for growth habit segregates in the Essex (determinate) by Williams (indeterminate) cross. Heatherly et al. (2004) found growth habit and increased yield are not independent and indeterminate growth habit can produce higher yields in early maturing soybean lines. This would agree with our discovery of a minor QTL from the Williams cultivar for increasing yield. Another yield QTL was identified using SF ANOVA as-

sociated with Gm19_2404683_A_G (25.12 cM). The marker Satt313 (32.3 cM) was found to be associated with seed weight on chromosome 19 in a cross between the cultivars Ma Belle x Proto (Csanadi et al., 2001). Guzman et al. (2007) reported a yield QTL with the same marker on chromosome 19 at 34.5 cM. Smalley et al. (2004) reported a yield QTL in the same region associated with Satt143 (31.8 cM), which are all less than 10 cM from the QTL reported in this study.

Chromosome 20

Gm20_43890641_G_T (54.79 cM), Gm20_46574547_T_C (65.04 cM) and Gm20_41827386_T_C (43.53 cM) were associated with a yield QTL on chromosome 20 using SF ANOVA. Satt354 (45.22 cM) reported by Reinprecht et al. (2006) and Satt270 (57.9 cM) reported by Smalley et al. (2004) were also associated with a yield QTL on chromosome 20. Another yield QTL was linked to Gm20_800671_A_G (1.83 cM) using SF ANOVA. Smalley et al. (2004) reported a yield QTL linked to Satt127 (15.5 cM) in three populations, however no other studies were found that reported QTL in that region of chromosome 20. No yield QTL were detected using R/qtl on chromosome 20.

Table 2. Yield prediction model (YPM) developed using QTL detected in Knoxville, TN in 2010 by R/qtl to select by MAS the top yielding 10% of RIL in Group B grown in individual environments and averaged across multiple environments. These MAS lines are indicated in bold.

YPM		YIELD (kg ha^{-1})					
KNOXVILLE, TN 2010		KNOXVILLE, TN 2011		BELLEVILLE, IL 2011		KNOXVILLE, TN 2010-11 BELLEVILLE, IL 2011	
LINE	RANK	LINE	YIELD	LINE	YIELD	LINE	YIELD
†§**197**	01	694	2699.4	65	4266.6	§**550**	3137.8
†**413**	02	†**681**	2670.8	172	4206.1	676	3124.3
‡§**383**	03	†**550**	2662.4	676	4132.2	172	3043.7
†**431**	04	676	2620.4	‡**550**	4085.2	§**722**	3016.8
267	05	†**1013**	2604.8	826	4078.4	§**681**	3010.1
783	06	518	2591.9	439	4078.4	702	2916.0
†§**1013**	07	†**722**	2591.9	‡**881**	4024.7	332	2909.3
†§**681**	08	†**413**	2583.5	‡**383**	4011.2	888	2902.6
597	09	†**197**	2580.1	570	4004.5	§**1013**	2902.6
653	10	478	2578.4	533	3984.4	665	2902.6
7	11	665	2559.9	11	3970.9	330	2889.2
‡§**881**	12	702	2553.2	‡**329**	3957.5	§**197**	2875.7
886	13	672	2538.1	437	3944.1	694	2875.7
691	14	332	2514.6	619	3917.2	970	2869.0
422	15	330	2502.8	793	3910.5	346	2862.3
†‡§**550**	16	184	2467.6	362	3903.7	§**383**	2862.3
42	17	172	2465.9	998	3903.7	1008	2862.3
230	18	259	2450.8	342	3903.7	362	2855.6
‡**329**	19	346	2428.9	888	3897.0	826	2855.6
411	20	397	2428.9	375	3883.6	§**881**	2848.9
275	21	†**431**	2425.6	625	3876.9	65	2842.1
†‡§**722**	22	1008	2423.9	‡**722**	3863.4	922	2842.1

†Indicates lines in the top yielding 10% of RIL grown in Knoxville, TN in 2011 that were selected using the YPM developed using QTL detected in Knoxville, TN in 2010 by R/qtl; ‡Indicates lines in the top yielding 10% of RIL grown in Belleville, IL in 2011 that were selected using the YPM developed using QTL detected in Knoxville, TN in 2010 by R/qtl; §Indicates lines in the top yielding 10% of RIL averaged over Knoxville, TN in 2010, 2011 and Belleville, IL in 2011 that were selected using the YPM developed using QTL detected in Knoxville, TN in 2010 by R/qtl.

Table 3. Significant (P<0.01) epistatic interactions between loci for yield in 220 RIL in Group D derived from a cross between Essex 86-15-1 x Williams 82-11-43-1. Locus 1 indicates the markers where yield QTL were detected using R/qtl and locus 2 indicates the markers where QTL were detected using Epistacy in SAS that were interacting with the yield QTL at locus 1.

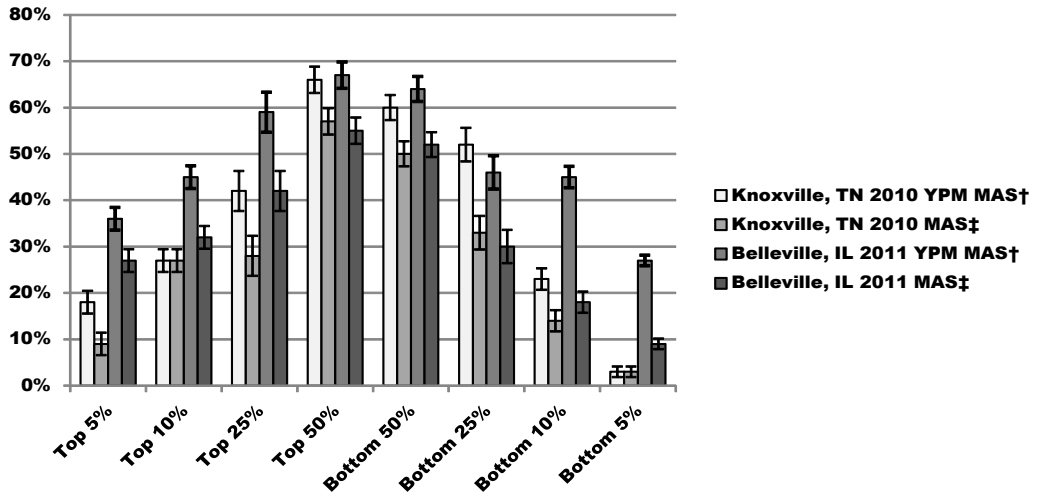
											ADDITIVE X ADDITIVE EFFECT†	
ENVIRONMENT	LOCUS 1	CHR	MLG	LOC (cM)	FAVORABLE ALLELE	LOCUS 2	CHR	MLG	LOC (cM)	R^2 (%)	E	W
Knoxville, TN 2010-11&												
Belleville, IL 2011	Gm07_17362808_A_G	7	M	55.95	W	GM_01-5021663_A_G	1	D1a	11.79	3.98	-0.27	-1.68
						GM06_20835584_T_C	6	C2	48.91	4.90	3.04	-1.35
						GM06_20996124_T_C	6	C2	62.03	4.82	2.43	-1.40
						GM13_26707540_C_T	13	F	62.69	4.42	-1.58	-0.08
						GM15_11274131_A_G	15	E	26.47	3.60	-1.56	-0.18
						GM18_58266066_T_C	18	G	136.77	5.41	-1.63	-0.03
Knoxville, TN 2010-11&												
Belleville, IL 2011	Gm06_20996124_T_C	6	C2	62.03	W	GM01_29990637_T_C	1	D1a	70.40	4.93	31.75	36.68
						GM07_17460956_C_A	7	M	40.99	4.79	36.21	35.23
						GM08_12693852_G_A	8	A2	29.80	4.71	31.89	36.46
						GM10_47858822_C_T	10	O	112.34	4.03	36.51	31.63
						GM12_33657269_G_T	12	H	79.01	4.83	36.55	31.77
						GM13_26707540_C_T	13	F	62.69	3.72	36.55	31.42
						GM15_49375283_T_C	15	E	115.90	5.43	36.55	31.18
						GM19_45082401_G_A	19	L	105.83	4.78	36.56	31.72
Knoxville, TN 2010-11&												
Belleville, IL 2011	Gm02_42469280_A_C	2	D1b	105.17	W	GM12_34378311_T_C	12	H	80.70	5.31	0.17	-1.58
						GM16_29150479_A_G	16	J	68.43	3.65	0.15	-1.28

† Additive by additive effect refers to the quantitative change in yield that is associated with the epistatic combination of the additive genetic effect of locus 1 having the favorable allele with the additive genetic effect of the homozygous state of locus 2 from (E) Essex 15-86-1 or (W) Williams 82-11-43-1.

Table 4. Yield prediction model (YPM) developed using QTL detected in Knoxville, TN in 2010 by CIM to select by MAS the top yielding 10% of RIL in Group B grown in individual environments and averaged across multiple environments. These MAS lines are indicated in bold.

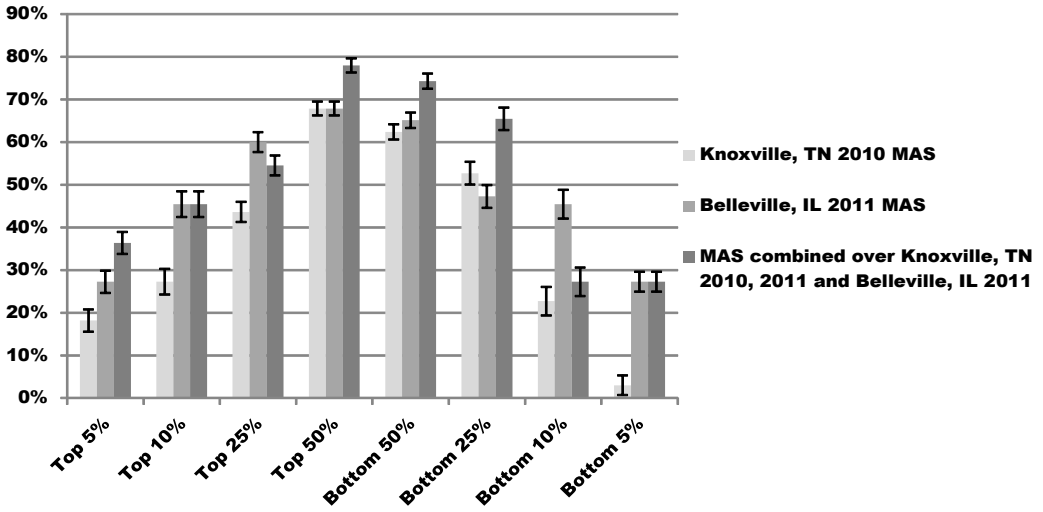
YPM		YIELD (kg ha^{-1})					
KNOXVILLE, TN 2010		KNOXVILLE, TN 2011		PLYMOUTH, NC 2011		KNOXVILLE, TN 2010-11 PLYMOUTH, NC 2011	
LINE	RANK	LINE	YIELD	LINE	YIELD	LINE	YIELD
†**461**	01	†**461**	2851.5	772	3467.0	864	2734.6
†§**94**	02	†**706**	2809.2	216	3426.7	§**81**	2647.3
†§**81**	03	†**94**	2743.7	645	3299.0	686	2640.6
†**381**	04	†**81**	2587.8	122	3265.4	530	2600.3
†**459**	05	201	2571.0	686	3205.0	§**918**	2600.3
334	06	530	2563.6	682	3178.1	122	2580.1
161	07	262	2555.6	984	3151.2	605	2546.5
‡§**846**	08	†**741**	2498.1	980	3117.6	984	2539.8
†§**706**	09	864	2485.7	434	3090.7	§**491**	2533.1
226	10	†**918**	2468.6	‡**846**	3037.0	§**706**	2526.3
†**766**	11	531	2465.9	236	3037.0	847	2519.6
140	12	522	2456.5	57	3023.6	531	2492.7
281	13	†**491**	2453.1	700	3016.8	220	2486.0
328	14	647	2448.7	910	2990.0	§**846**	2479.3
†§**918**	15	75	2446.4	678	2990.0	688	2459.2
†**12**	16	†**381**	2435.6	623	2976.5	917	2459.2
548	17	†**766**	2427.9	855	2976.5	647	2452.4
†**228**	18	†**228**	2427.2	456	2929.5	1010	2439.0
†§**491**	19	†**459**	2421.9	516	2929.5	§**94**	2439.0
†**741**	20	†**12**	2406.1	1021	2922.8	23	2439.0
35	21	475	2402.7	902	2909.3	682	2432.3
508	22	605	2399.4	777	2882.5	118	2418.8

†Indicates lines in the top yielding 10% of RIL grown in Knoxville, TN in 2011 that were selected using the YPM developed using QTL detected in Knoxville, TN in 2010 by CIM; ‡Indicates lines in the top yielding 10% of RIL grown in Plymouth, NC in 2011 that were selected using the YPM developed using QTL detected in Knoxville, TN in 2010 by CIM; §Indicates lines in the top yielding 10% of RIL averaged over Knoxville, TN in 2010, 2011 and Plymouth, NC in 2011 that were selected using the YPM developed using QTL detected in Knoxville, TN in 2010 by CIM.



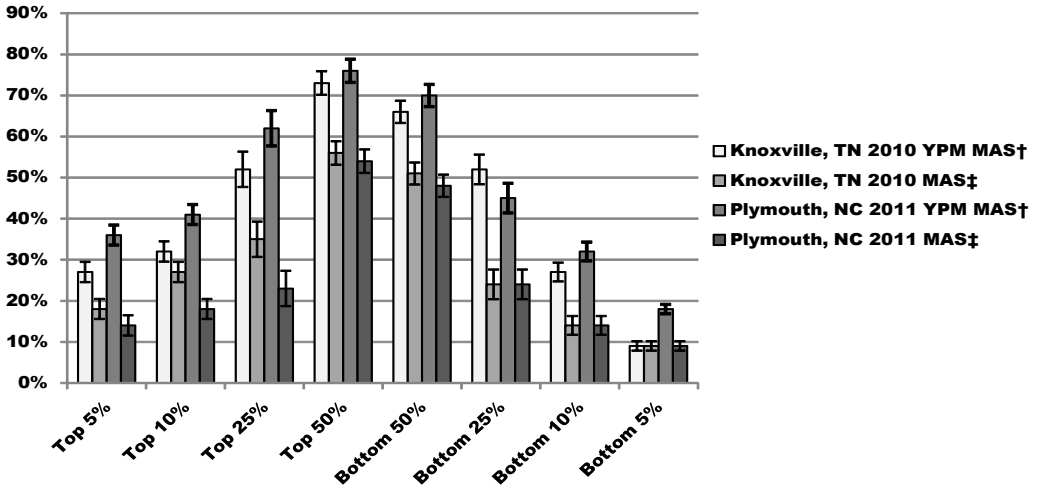
†Indicates MAS made using the YPM, which included: mean yield, additive effects and additive by additive effects for the QTL detected in that environment; ‡Indicates MAS made using only additive effects for the QTL detected in that environment.

Figure 1. The percentage of marker assisted selections (MAS) made in each environment in Group B compared to phenotypic selections (PS) averaged over all environments (Knoxville, TN 2010, 2011 and Bellville, IL 2011) in Group B. Comparisons were made between the top and bottom % of MAS that were in the corresponding top and bottom % of PS. MAS were made using only additive effects and a yield prediction model (YPM) developed using QTL detected in each environment. PS were based on yield in kg ha^{-1}.



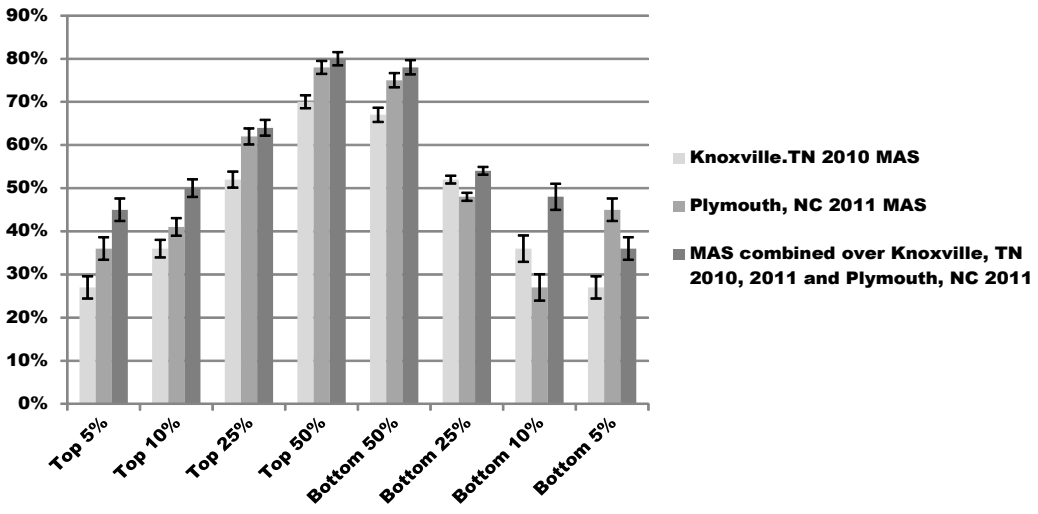
†YPM indicates what environment(s) the data for the model was collected: mean yield, additive effects and additive by additive effects.

Figure 2. The percentage of marker assisted selections (MAS) made in each environment(s) in Group B compared to phenotypic selections (PS) averaged over all environments in Group B. Comparisons were made between the top and bottom % of MAS that were in the corresponding top and bottom % of PS. MAS were made using a yield prediction model (YPM†) developed using QTL detected in each environment(s). PS were based on yield in kg ha^{-1}.



†Indicates MAS made using the YPM, which included: mean yield, additive effects and additive by additive effects for the QTL detected in that environment; ‡Indicates MAS made using only additive effects for the QTL detected in that environment.

Figure 3. The percentage of marker assisted selections (MAS) made in each environment in Group D compared to phenotypic selections (PS) averaged over all environments (Knoxville, TN 2010, 2011 and Portageville, MO2011) in Group D. Comparisons were made between the top and bottom % of MAS that were in the corresponding top and bottom % of PS. MAS were made using only additive effects and a yield prediction model (YPM) developed using QTL detected in each environment. PS were based on yield in kg ha^{-1}.



†YPM indicates what environment(s) the data for the model was collected: mean yield, additive effects and additive by additive effects.

Figure 4. The percentage of marker assisted selections (MAS) made in each environment(s) in Group D compared to phenotypic selections (PS) averaged over all environments in Group D. Comparisons were made between the top and bottom % of MAS that were in the corresponding top and bottom % of PS. MAS were made using a yield prediction model (YPM†) developed using QTL detected in each environment(s). PS were based on yield in kg ha^{-1}.